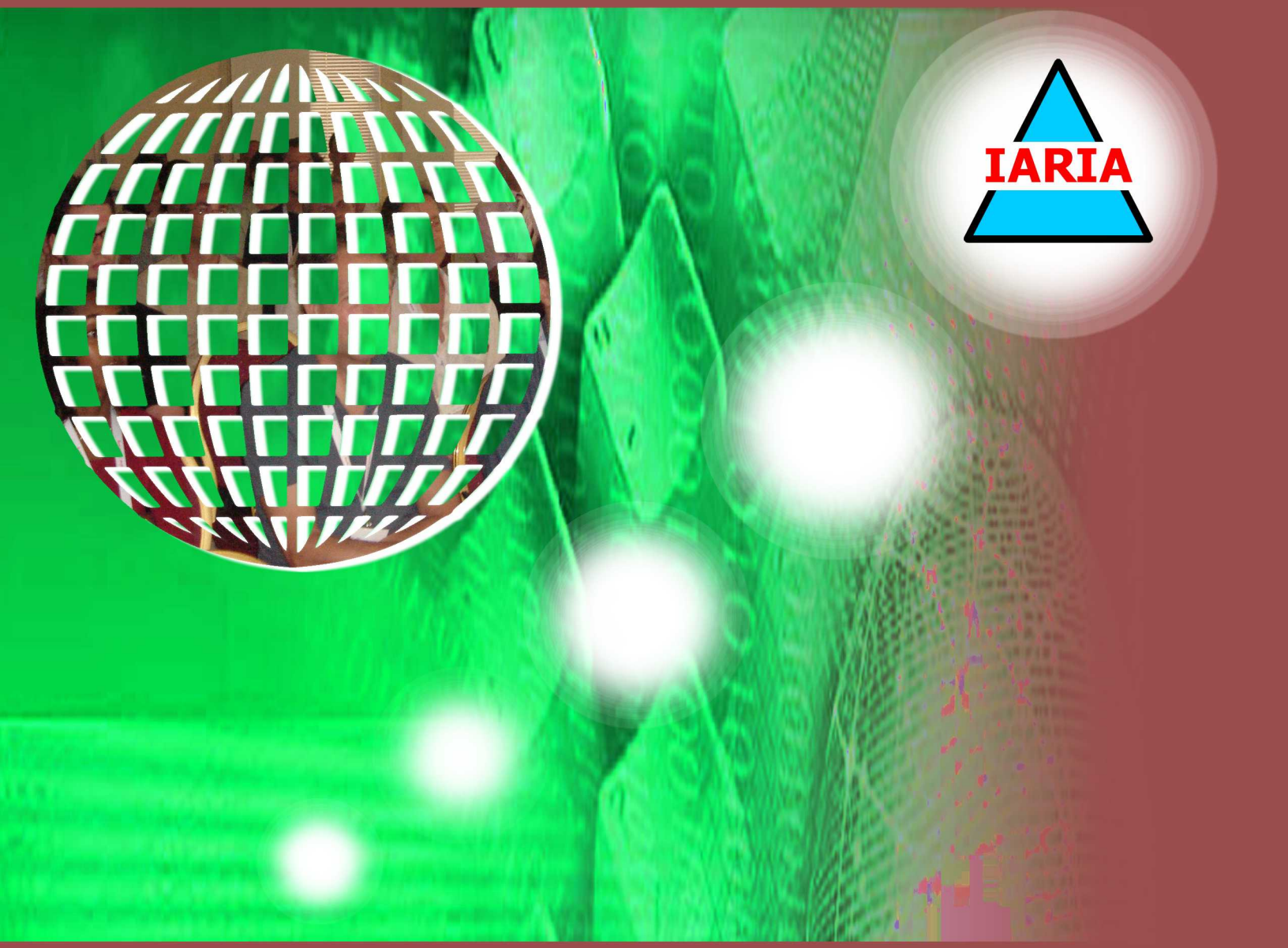
International Journal on Advances in Life Sciences



2013 vol. 5 nr. 3&4

The *International Journal on Advances in Life Sciences* is published by IARIA.
ISSN: 1942-2660
journals site: http://www.iariajournals.org
contact: petre@iaria.org

Sponsored by IARIA
www.iaria.org

International Journal on Advances in Life Sciences
Volume 5, Number 3 & 4, 2013

International Journal on Advances in Life Sciences
Volume 5, Numbers 3 & 4, 2013

CONTENTS

Claude Weber, Air Liquide, France
Daniel Navajas, University of Barcelona-IBEC-CIBERES, Faculty of Medicine, Unit of Biophysics and Bioengineering, Spain
Ramon Farré, University of Barcelona-IDIBAPS-CIBERES, Faculty of Medicine, Unit of Biophysics and Bioengineering, Spain

Reflections on UK Local Government Challenges in the use of Twitter as a Communications Channel

Darren Mundy and Qasim Umer
University of Hull
Scarborough, GB
d.mundy@hull.ac.uk; qasimumer667@hotmail.com

***Abstract* — Over the previous five years local government agencies have begun to use social media networks (such as Twitter) as mechanisms to promote engagement with local citizens. However, as identified in previous research there can be substantial challenges in relation to the use of these spaces to encourage bi-directional conversation and engagement. This paper provides a critical perspective on the challenges presented in the results of a focused research project on governmental communication through Twitter over a three month period from October to December 2011 and a one month period in August 2013. The research presented in this paper, contributes to the growing number of research papers related to the effective use of social media platforms in governmental, organisational and other community spaces. It is clear that as service provision develops, a growing maturity of usage is enabling councils to further develop their understanding of what is good practice in communicating through social media channels.**

Keywords - social media networks; social network analysis; eGovernment; communication channels.

I. INTRODUCTION

This paper provides an extension of work presented at the Second International Conference on Social Eco-Informatics in Venice in October 2012 [1]. The work centres on an analysis of local government communication using the social media platform of Twitter. This extended work uses further analysis of communication over the month of August in 2013 to reflect on the initial findings and posit further challenges to social media communication for local government organisations.

In March 2012, Facebook reported a worldwide audience of almost 836 million users [2]. In the same month, Twitter reported 140 million active users with 340 million tweets per day [3]. These statistics demonstrate the global presence of networks that have only been in existence for less than ten years. These global audiences may be dispersed, and not always balanced across local communities. With such large user bases companies are exploring how they can best engage users on these platforms, and leverage a return on investment, from time committed to social media spaces.

Social media platforms such as Facebook, Twitter, Pinterest, YouTube, Google+ etc. provide opportunities for individuals to collectively discuss, share, participate, produce, and respond to specific materials dependant on platform (e.g., videos on YouTube and short text messages on Twitter). Many of the social media platforms provide an opportunity for individuals to provide a picture of themselves (often through a profile), and opportunities to obtain and engage with information, often in real time. In addition, such platforms provide opportunities for individuals to discover new mechanisms to engage with existing organisations or to discuss organisations in a public setting.

One such organisation that individuals around the world are using social media platforms to engage in discussing is their national and local government. This paper focuses specifically on engagement by the public with local government services. UK government is managed through a mixture of national government departments and local government authorities. A primary requirement of the UK democratic system is a need to engage with, and listen to, constituents through a multitude of different access channels.

Over the past three years with the austerity measures imposed across the whole of Europe, and other parts of the world, the UK has had to look again at the delivery of public services. These austerity measures have impacted on local government funding, meaning in places, cuts in front line services and reductions in service provision.

Organisations (including local government authorities) recognise the benefits that shifting customer enquiries from physical face to face customer contact centres, to online provision, can bring. These benefits include: reductions in the cost per transaction of customer enquiries; and where front end services are directly connected to back end services the potential for disintermediation. Social media platforms have been identified as platforms through, which to engage in providing front end service information, and in answering customer enquiries in cost effective processes. UK government recognises how multi-channel solutions can be developed, however, more recently there is concentration on matching the most effective channel with the most appropriate service. It is also recognized that this service may not necessarily be the lowest cost of contact, as this may not necessarily translate through to the lowest cost of transaction, e.g., because of error percentages or the need for multiple contacts.

Local government agencies in the UK consist of county, district, borough, city, and unitary authorities. These agencies provide a mixture of local management functions including those focused on education, transport, the environment, recycling and waste. Such councils are present across a range of social media spaces with many individual councils engaging across multiple spaces. Structures of staff engagement vary with some councils creating cross departmental structures to facilitate wider council engagement, whilst others limiting engagement to singular

teams. In addition to staff within local authorities engaging within social media spaces, many other publically funded organisations are also using these networks as mechanisms to engage the local taxpaying public. This network of public sector engagement provides the citizen with direct access to queries about public sector services, and mechanisms for obtaining current localised information.

This paper builds on the research results presented in [1][4]. The initial paper of this study presented the findings from research focused on the engagement of ten local UK councils within the Twitter network. It also outlined a range of quantitative statistics regarding the study and indicated a number of issues discovered within the textual analysis of the tweets. This was followed by a paper [1] exploring the deeper challenges presented from the data, determined from detailed textual analysis of the tweets collected for the study. This study extends [1] and [4] providing updated quantitative results for 2013 and a reflection of the meaning of these results on the deeper challenges. In addition, a further couple of challenges are presented coming out of the 2013 data.

Whilst the study has focused on Twitter throughout each of the delivered papers, many of the challenges presented within this paper, can be easily applied to other social media networks. The paper contributes to a growing number of research papers related to the effective use of social media platforms in governmental, organisational, and other community spaces.

This introduction has provided an overview of the importance of social media networks, and the potential for social media networks to impact on the government: citizen relationship. Section II gives an overview of literature in the area of social media communication, particularly highlighting the effective use of social media in government spaces. This is followed by a detailed review of the methodology used to analyse local government conversations in the ‘ Twitterverse’. These conversations are then used to identify clear challenges in the use of Twitter, and other social media platforms to engage the citizen. Finally, a series of recommendations linked to the challenges are provided and the paper closes with a conclusion and the identification of further work.

II. SOCIAL MEDIA AND UK GOVERNMENT

The growth of social media is impacting the ways in which communities work, communicate, and socialize. According to [5] social media platforms can help to fulfill the needs of rapid communication, to engage individuals with multimedia artefacts and problematically to blur what is private and what is public. Similar to the demands imposed by users of the short messaging service (SMS) on mobile phones, social media networks demand prompt response, knowledge of and continued engagement with the platform. Twitter, Facebook and YouTube are examples of three multilingual social networking websites launched since 2004 that respectfully provide forums for social networking, micro blogging, and multimedia content sharing (text, video and photos). They give users the opportunity of being real time in a virtual world [6], and enable users to create their own accounts, content channels and interest group sites. Different governments, channels and groups also have their own individual pages and accounts on social media networking sites.

This paper focuses on the use of and challenges in the use of social media, in particular, Twitter by local government in the UK. Research into the use of social media networking sites by national governments, and organisations across the world continues to grow, as we continue to look for how transformative communication can be provided through such channels. Recently Stephen Goldsmith used the term ‘government by network’ [7] to describe how online communication channels were being used at different levels of government, to engender a greater sense of participative relationship with citizens. In addition, there is evidence that social media networks when employed effectively have the power to create change in relation to political discourse, encouraging individuals to re-engage with democratic systems [8] and create a greater sense of the citizen voice [9].

Research in the area of government use of social media networks, can be broken down into two primary areas of interest:

- Analysis at national level of how parties and political leaders utilise these communication channels to engage citizens with national and international political issues.
- Engagement with how local government organisations are using such platforms to engage citizens with local services and issues.

Research related to local government usage of Twitter is not as well established as national level research, as the national issues often generate more substantial interest. In this national space, the most interesting study to date has focused on how Barack Obama (current president of the USA), utilised the web and social media networks in his historic election victory in 2008 [8]. Other researchers have focused on an analysis of the use of social media networks in relation to encouraging political dialogue [10], the use of social media for political public relations [11] and analysis of social media channels as political communication channels [12]. Researchers have found from a national perspective that broadcast information over these channels is broadly favoured, and that individual channels are not in themselves ‘game changers’, but merely an additional communications channels for contact with citizen groups [10]. From a national perspective it is useful to note [12], which suggests that these network channels can be used to provide a prediction on the results of national elections. This in itself is not unique (given perhaps we could also use analysis of party prevalence in other forms of media or surveys), it does suggest that political discourse at the national level is frequent, and the size of it is nationally relevant on social media channels.

Whilst national issues are of interest this paper focuses primarily on investigating local government discourse. One of the key issues for local government is engendering citizen interaction in positive, and progressive ways. Social media platforms can offer opportunities for individuals within local communities to provide their view on a local issue, report a broken street light, or to simply interact with a local

councillor. Researchers such as [13][14][7] have indicated that these channels can enable a transformation in the way in which local government, and citizens communicate leading to transformative relationships developing. In particular, [14] argues that these networks can enable "an atmosphere of co-operation" as citizens work with local government in developing better physical communities.

A barrier to the co-operative atmosphere highlighted in the above paragraph, arises in the ways in which local government utilise social media spaces. Reference [15] suggests that the ways in which many local government organizations, manage their information technology infrastructure, and the ways in which local government service structures are established, can limit the ways that such organisations are able to utilise social media systems. They state within these circumstances the use of social media channels *"can only practically be used to broadcast announcements rather than to interact with people"*. As a practical example one could posit the traditional ways in which customer service interaction, and public relations are managed in local government organizations, can often lead to uncertainty with regards to the translation of such services on to social media platforms.

Noting the growth in usage by local government of social media, recently research has been published to try to provide such organisations with a framework of good practice [16]. However, this area continues to evolve and further work is needed to understand where local government is now with its engagement in social media spaces, what the challenges are with regards to this engagement, and how such engagement can continue to evolve. This paper represents work towards establishing a response to the above, highlighting in particular, the challenges to local authorities effective use of social media platforms such as Twitter, Facebook and YouTube.

III. METHODOLOGY

Over the periods October – December 2011 and August 2013 tweets related to ten local councils in the UK were analysed. The results of this analysis for 2011 have been provided in [4]. The research presented below updates the earlier analysis looking at how councils have changed their practice over the past two years and focuses on a critical analysis of the challenges presented.

The councils selected for analysis were drawn from a group of UK councils with active accounts within Twitter. Therefore, the analysis focused only on those councils who were experimenting with this social media network in 2011, and were already engaging in some way, with their local populace in the ' Twitterverse'. It is notable that all the analysed who were using Twitter in 2011 have continued to use Twitter and continued to grow their communities in the updated research in 2013.

Tweets were collected from within defined council spaces and from outside of defined council spaces in 2011. For the updated research in 2013, the focus stayed within defined council spaces. This effectively translates through to collection of tweets that each council had made (2011, 2013), collection of citizen tweets to the council (2011, 2013), and tweets made, which made reference to the council or the local community (2011). It should also be noted that no collection was possible for those communications happening within Twitter through private communication channels (e.g., through Twitter's direct messaging system).

The aggregated data collected provided quantitative metrics covering aspects such as:

- the number of tweets;
- the number of followers;
- the number of re-tweets etc.

In addition, the text within each tweet was analysed in relation to:

- categorising the focus of the content;
- understanding the direction of the communication (e.g., whether it was just broadcast or a response to another Twitter user);
- a detailed content analysis of the messages contained in the tweet content.

In relation to the detailed analysis of the messages in the tweet content, each tweet was read, and broader issues were identified within the tweet content from 2011 and 2013 tweet samples. This paper provides an updated critical analysis of these broader issues from earlier papers [1], and identifies the challenges presented by these.

There were a number of limitations presented in the research data gathered within this project. The main issues were linked to the choice of councils, the time of year studied, and the maturity of various councils' use of Twitter as a communication channel. However, these limitations have limited impact on the broader challenges presented through this analysis work.

In addition, to this direct identification of challenges from analysis of the research data a reverse approach has also been used. Conversation with individuals responsible for social media management in local authority spaces has helped to inform and identify other challenges present within the research dataset.

In total 1,565 tweets for 2011 and 1,264 tweets for 2013, were analysed from within council defined spaces in Twitter. As an addition to this a total number of 593 re-tweets from other organisations included in the council communication channels were also analysed. Finally, a substantial number of tweets were also reviewed in spaces outside of council control.

IV. ANALYSIS OF RESULTS

Table I provides detail about council activity in Twitter (measured by the number of Tweets) and the number of individuals who have signed up to follow each channel within Twitter (measured by the number of followers). This detail is provided for October – December 2011 and September 2013. The data demonstrates that over the past two years all the analysed government organisations have continued to Tweet and expanded the number of individuals following their service, with all councils registering greater than 50% increases in their network followers.

The number of followers as a percentage of the population was also calculated in 2011 based on census data.

Whilst the percentage has substantially grown it still represents across all councils a small proportion of the total population. Over this time whilst the community has grown in number of followers it is likely that the number of followers from outside of council areas has also continued to grow.

TABLE I. GENERAL DATA ABOUT SELECTED COUNCIL SPACES

Council	No of Followers (Jan 12)	Total No of Tweets (Jan 12)	No of Followers (Sep 13)	Total No of Tweets (Sep 13)
Newcastle Upon Tyne City Council	8637	4533	19875	8021
City Of Lincoln Council	3215	2740	6617	4181
Tameside Metropolitan Borough Council	3357	2242	8131	7492
Salford City Council	4797	3585	10857	7237
Leeds City Council	6803	1631	14798	6407
Oxford City Council	4541	546	8970	1721
Southend-On-Sea Borough Council	2115	1241	6280	5213
Camden London Borough Council	4034	2665	8516	5984
Kirklees Council	4089	3685	9672	6903
Wakefield Council	3024	2090	8450	6593

TABLE II. INFORMATION ABOUT TWITTER DATA SET

Council	Total No of Sampled Tweets (Jan 12)	Total No of Re-Tweets (Jan 12)	Total No of Sampled Tweets (Sep 13)	Total No of Re-Tweets (Sep 13)
Newcastle Upon Tyne City Council	170	48	48	19
City Of Lincoln Council	151	29	89	2
Tameside Metropolitan Borough Council	71	12	382	98
Salford City Council	317	48	129	18
Leeds City Council	112	21	84	81
Oxford City Council	49	23	38	11
Southend-On-Sea Borough Council	83	0	25	9
Camden London Borough Council	162	0	185	35
Kirklees Council	266	7	102	17
Wakefield Council	184	3	182	112

Table II includes information about the sample set of tweets collected in 2011 in relation to the initial study and the set of tweets collected in 2013 expanding the study and capturing changes in use. For 2011, the table states the number of tweets captured for analysis between October and December 2011. For 2013, the table states the number of tweets captured over August 2013. This number is in essence the number of tweets made by the council over the specific periods of time. In addition, the number of re-tweets by the council is also captured. Re-tweets are those tweets that are the tweets of other users that the council has chosen to re-broadcast within their network.

Over the initial period of analysis, Salford City Council was the most active tweeting council with 317 tweets, whilst Oxford City Council was the least active. In the updated 2013 analysis Tameside Council was the most active tweeting council, with Southend-on-Sea Borough Council being the least active. From a re-tweet perspective Newcastle upon Tyne City Council and Salford City Council were most prolific in re-tweeting community messages in 2011, whilst Southend on Sea Borough Council and Camden London Borough Council made no re-tweets. For the updated analysis Wakefield Council were most prolific in re-tweeting community messages, with Lincoln City Council being the least prolific. In the 2013 analysis all councils were engaged in re-tweeting content. Significant increases in the amount of re-tweeted content have occurred in Tameside Metropolitan Borough Council, Leeds City Council and Wakefield Council. This suggests that these councils have adapted their practice over time and are now gaining a better understanding of how re-tweeted content can improve their service.

Table III contains information about the categorisation of the tweets sampled in relation to whether they are simple announcements from the council to the community, or whether the tweets are responses to community members. A response in Twitter generally (but not always) begins with the username of the Twitter user that one is responding to. This table demonstrates the level of engagement between council and citizen through Twitter. In the main this shows that the majority of councils are now clearly engaging in citizen conversation through Twitter. Leeds City Council

have moved their citizen communication into a different Twitter feed, so the numbers shown in this particular table do not show the full picture. The sample set in 2011 showed a number of councils were not really using Twitter in the way it was designed as a social communication channel which encourages two-way conversation. This is clearly not the case in 2013.

Tables IV, V and VI contain information about the textual content of each tweet categorised in relation to a range of council services.

TABLE III. DIRECTION OF COMMUNICATION

Council	Announcement 2011	Response 2011	Announcement 2013	Response 2013
Newcastle Upon Tyne City Council	115	10	47	1
City Of Lincoln Council	42	77	58	29
Tameside Metropolitan Borough Council	49	10	193	189
Salford City Council	184	85	53	76
Leeds City Council	56	35	76	8
Oxford City Council	16	10	33	5
Southend-On-Sea Borough Council	83	0	18	7
Camden London Borough Council	63	99	85	100
Kirklees Council	235	24	71	31
Wakefield Council	164	17	137	45

TABLE IV. FIVE MOST POPULAR CATEGORIES FOR ANNOUNCEMENT TWEETS

Council	Sport, Leisure & Entertainment	Business	Housing	Jobs & Careers	Parking, Roads, Travel & Transportation
Newcastle Upon Tyne City Council	13	3	1	1	5
City Of Lincoln Council	13	1	7	7	8
Tameside Metropolitan Borough Council	98	11	17	2	18
Salford City Council	15	4	2	13	3
Leeds City Council	28	8	5	6	1
Oxford City Council	8	3	11	3	2
Southend-On-Sea Borough Council	5	0	3	0	1
Camden London Borough Council	26	10	3	13	0
Kirklees Council	40	6	0	1	2
Wakefield Council	97	4	0	1	2

TABLE V. FIVE MOST POPULAR CATEGORIES FOR RE-TWEETS

Council	Sport, Leisure & Entertainment	Business	Jobs & Careers	Bins & Recycling	Education
Newcastle Upon Tyne City Council	5	4	0	0	0
City Of Lincoln Council	1	0	0	0	0
Tameside Metropolitan Borough Council	38	3	6	11	3
Salford City Council	10	0	1	2	1
Leeds City Council	33	3	6	6	10
Oxford City Council	5	0	0	0	0
Southend-On-Sea Borough Council	6	0	0	2	0
Camden London Borough Council	7	17	6	1	1
Kirklees Council	9	2	1	0	2
Wakefield Council	88	6	2	0	0

TABLE VI. FIVE MOST POPULAR CATEGORIES FOR RESPONSIVE TWEETS

Council	Bins & Recycling	Reporting	Parking, Roads, Travel & Transportation	Sport, Leisure & Entertainment	Business
Newcastle Upon Tyne City Council	1	0	0	0	0
City Of Lincoln Council	3	7	5	1	0
Tameside Metropolitan Borough Council	78	27	23	17	10
Salford City Council	26	19	10	7	2
Leeds City Council	2	0	1	1	2
Oxford City Council	0	3	0	0	0
Southend-On-Sea Borough Council	0	2	0	2	0
Camden London Borough Council	20	29	24	1	10
Kirklees Council	5	6	4	5	3
Wakefield Council	1	7	5	17	4

In 2011, in terms of announcements via Twitter the most popular reason for tweeting was to publicise entertainment, sport or leisure activities occurring in the local area. There was no change to this in the 2013 sample set. The only event which all councils engaged in tweeting content about during the August 2013 period was the release of educational results, with all councils commenting on the performance of their local students. The rest of the announcements varied by council, with for example, City of Lincoln council advertising conversations with their local council leader, Camden London Borough Council announcing government results of the development of a high speed rail network (HS2), and Tameside Metropolitan Borough Council making a range of announcements relating to temporary recycling arrangements (because of a fire in the local neighbourhood).

In relation to responsive tweets from the council to local citizens these primarily centred around transportation, entertainment, waste and housing services in 2011. In 2013, the primary areas of interest focused on waste management, reporting and transportation services. In general very few conversations within Twitter in 2013 lasted for any significant length of time. In general conversations resolved issues, or in the case of a number of councils moved more significant issues out into direct messages or email. Outside of the most popular categories of conversation, responsive tweets in other categories related to a single or couple of conversational instances about particular issues.

Finally, councils in 2011 were fairly inconsistent with regards to re-tweeting content across categories, with re-tweeted content seeming to centre primarily on local council priorities. This inconsistency also translated into the 2013 sample, however, all councils continued to provide some re-tweeted content related to sports, entertainment and leisure activities. Local priorities changed between 2011 and 2013 primarily relating to specific concerns of the period of time sampled. One example of this is Tameside Metropolitan Borough Council who during the course of August 2013 re-tweeted content (categorised as 'other') from the local fire service related to a substantial fire at a recycling plant in the local area.

V. LOCAL AUTHORITY CHALLENGES

The next few sections will take an in-depth look at challenges identified from the analysis of the tweets sampled in 2011. This material is extended and updated to reflect on the meaning of the 2013 research in relation to what has changed and whether the challenge continues to be valid. Sections J and K deal with recently identified challenges taking into account analysis of the 2011 and 2013 datasets.

A. Transforming Broadcast Communication

As noted by Rooksby and Sommerville [15] there are significant challenges in transforming different elements of local government services, away from broadcast only models of communication. Social media communication platforms provide direct public channels of communication between citizens and local government. This can lead to citizens directly criticising decisions made in local communities, and the transparency of local government [4]. It can also lead to citizens raising issues about local service provision or issues related to local areas. This is certainly apparent across the datasets from 2011 and 2013.

The challenge therefore is how to construct these social media channels such that appropriate individuals receive the messages related to their particular services, and how to manage this communication within these channels. In other circumstances, often the conversation is less public and more easily controlled – this leads to local government organisations (similar to other commercial organisations) approaching two way conversation in social media spaces, with some trepidation.

It is clear between 2011 and 2013 local government has significantly matured in relation to how this communication process is handled. In 2011, there was a feeling from the dataset that often the channel was used as the direct mechanism to achieving resolution, foregoing existing channels for resolution. In 2013, there are two specific mechanisms highlighted to demonstrate how councils are gaining an understanding of how to handle the channel. The first is the direction of individuals to recognised mechanisms

for reporting, e.g., email addresses for waste services or web based report forms. The second is the establishment of specific Twitter feeds for issue resolution, e.g., feeds for particular services recycling and or feeds for direct help, for example Leeds_Help.

Over both sample sets there was an indication that where local government is engaging directly in conversation with the local citizen base and helping to resolve issues, there is ample evidence that citizens are appreciating this service. In 2011, there was evidence of individuals commenting that social media platforms were the most effective ways in which they have communicated with government services. In 2013, there was no indication of this. Indeed, where the approach had matured into a direction of issue to either a web form or email service there was an indication that some individuals were resenting a multi-channel approach, e.g., asking questions as to why issues could not be raised and resolved directly through Twitter.

B. Frontend to Backend Integration

Local government must take clear steps to manage the social media channels they are using from the perspective of understanding information flow. Tweets show evidence that for some councils citizens see their Twitter spaces as the most effective way to gain answers to questions posed. However, in 2011 there were a number of cases of citizens not obtaining the information they required. One solution may see customer service staff taking responsibility for the engagement in answering citizen questions within social spaces, and other specific staff (e.g., marketing) taking responsibility for highlighting council achievements. In addition, customer requests coming in through social media spaces must be tracked, to help to facilitate knowledge of return on investment, as highlighted in Section C.

As noted in Section A as services have matured since 2011, the way in which the service is provided via Twitter is changing. Councils are adopting different approaches in realising good practice through this communication mechanism. Approaches such as encouragement of services, councilors and other council operatives to develop their own Twitter presence, demonstrate mechanisms for encouraging community communication. This also enables community members to sign up to information purely about the singular or groups of elements of service most interesting to them, enabling better information management from a citizen perspective. In addition, it is clear that councils are understanding the need to control integration between front and backend services with messages passed quickly to other services, either directly by the Twitter feed operators or through re-direction of the client.

For the most part the problems experienced of issues raised in relation to a lack of reply through the Twitter channel have disappeared between 2011 and 2013. However, it is clear that users are not slow in letting the feed know of a lack of response by the specific service providers, e.g., if bins continue not to be collected or if issues reported continue not to be resolved. It is positive to see that where issues are resolved in a multitude of different cases users are also using the Twitter channel to broadcast their thanks.

C. Leveraging Return on Investment

With government austerity measures in force, and a reduction in funding to local government in the UK, local councils are busy reducing inefficiencies. Austerity measures have continued over the 2011-2013 period with many councils experiencing substantial reductions in funding and asking members of the public to respond in terms of which local services require protection through public consultation. In some cases over 2011 and 2013 this public consultation extends to or is advertised through the local governments Twitter channel.

In the light of the above, for social media channels to be embraced they must make clear demonstrable impacts on local citizen engagement. One way in which this can occur, is by moving transactions from more expensive channels (e.g., face to face services) to online information provision. This is where it is important to ensure that the social network spaces are joined up to local government services – otherwise the cost of a transaction and speed of reply may be similar to other channels particularly if answers are not clear or incorrect. The most effective councils using these platforms make use of the private messaging spaces within social media platforms to provide targeted complete answers to citizen questions.

Over the previous two years this particular challenge seems to be becoming more problematic. One of the reasons for this is Twitter itself, which encourages short message communication. This means that often members of the public need to engage in multiple tweets to inform the council regarding a particular issue. It also involves the council in the first instance trying to diagnose the issue, in order to be able to respond. In addition, as an asynchronous communication channel there can be delays between tweet responses from the citizen to the council and vice versa – this creates challenges in itself. With the above this all leads to questions over the cost of service, especially in comparison to other channels, it could be determined that the cost of a phone call would be less than the cost of multiple tweet communications, demonstrated by some councils pushing citizens directly through their Twitter channel to the phone service to engage the council with their issue.

D. Engaging the Citizen

Engaging the citizen will be helped by a transformation of broadcast communication, but this is not the only challenge to providing an engaging service. Social media platforms are often at their best when the channel providers are utilising the channel effectively by posting new interesting content. Often individuals will post content that is delivered across multiple social media channels this can cause problems when users are signed up for multiple services, each service needs to be distinctive, and different to engender different user communities. Whilst platforms can be transformative in providing a greater relationship between citizen, and government, they can also be transformative in a negative way, if the use of the channel does not meet with citizen expectations, for example if the channel is used to broadcast less meaningful information.

Over the past two years across Twitter channels this particular challenge has continued to become of real interest with councils finding creative ways to positively engage with the local populace. Some of the more interesting and engaging content has come in the use of different media types and different social media channels, with Twitter being used as a navigation tool to other items. For example, viral videos about particular challenges in the local area (e.g., a dog waste production by Wakefield Council) creating positive citizen response and the way in which they engage in re-tweeting positive local citizen messages.

E. Privacy Challenges

Social media platforms are not private and often citizens within these spaces are not completely anonymous. This presents a clear challenge in ensuring details of customer cases are kept out of the public domain. In one particular poor example of this in 2011, a council engaged across multiple tweets in detailing reasons for action, in a specific individual's case related to their housing situation. This could be deemed to infringe on data protection when specific details are discussed.

The counter concern is where customers are willing to discuss their individual personal circumstances across public spaces. Councils must learn when to move conversation into more private spaces, and to find ways of educating local citizens about problems related to providing too much personal information within public social media spaces.

In the 2013 dataset there was certainly substantially more maturity demonstrated in the use of service with Councils channeling personal information through the direct messaging tool or through email communication. In addition, the growing use of privacy functions in Twitter limiting and providing protected services for individual tweets is helping to improve this particular issue.

F. Engaging with the Conventions

Social media spaces are not the same, all have particular nuances, and the community within different spaces behaves often in particular ways. Understanding the nuances of the platform will lead to greater success in platform use. For example Twitter encourages users to follow other users and build networks through interconnections of individuals, the community also has established conventions like #FF, which is short for #FollowFriday (Twitter users use this at the end of a week to suggest other individuals to follow). Recognising, understanding and using the structures available can help councils to fit into the community and be seen as an interesting component part. It is evident that now with mature services council are gaining a much better understanding of the platforms they are using and are understanding how to generate and promote community building. Re-tweets, modified tweets (MT), #FF and other conventions are now regularly forming part of council communications.

G. Small Numbers

In 2011 and 2013, council social media spaces can be deemed to be attracting relatively small numbers, in terms of percentages of citizens within local communities engaging with the service offered by and through them. However, as councils gain a better understanding of how to use and market their presence on social media networks effectively then these numbers should continue to grow. It may well be for certain community groups engagement through social media spaces is the only way to engage these groups in conversations about local council spaces, and services.

Whilst over the past two years numbers have grown, this challenge is still of direct relevance. If numbers of followers continue to grow then this will demonstrate community growth, but the question is how many active members of community are obtaining information about the service, and how many accounts linked as followers are not actively viewed on a regular basis? The number of followers is not a strong enough performance indicator of community size.

More mature use of social media within council spaces will continue to look directly at how networks are forming in social media spaces around the council. Understanding the community in terms of the different types of user present, and engaging in the space, is extremely important. Equally using the community as a network to advertise the strength of the service should be one thing councils continue to look to do.

H. Endorsement or Community Building

It is evident from the research that the practice that councils are using in building communities within social media spaces, differs. Decisions such as whether the council should 'follow' other users or 'join' other groups are not always that simple. However, social media networks are often formed from practices of engaging in community building. For example, presence as a friend may highlight your presence to other users who are interested in you. Similarly engaging in the process of #FF, or re-tweeting content, also provides a sense of social media community engagement, and may influence others in passing on your content. Another example is the highlighting of other videos for consumption within your YouTube channel. However, these community ideals come at a risk, the risk that the content you highlight, are associated with, or pass on is inappropriate or legally problematic. There is a question over whether associating local councils with other information or users provides a form of endorsement of them or their services, and therefore, whether such endorsement is appropriate given the linkage to public services.

The data set in 2013 demonstrates changes in relation to council behavior regarding re-tweets, with differences in the number of re-tweets by councils, and in the types of organizational content re-tweeted. There are some good examples of re-tweet engagement by councils in engaging in re-tweeting positive local community communications. Twitter also requires users to follow one another to engage in direct messaging and other services, this means that community growth occurs through necessity in some circumstances.

I. Answering the Difficult Questions

Social media spaces provide direct lines of communication to the general public, which are open 24/7 and easily communicable. These spaces are used by individuals for a variety of purposes from communication about particular things happening within local communities, through to information discovery about other places before travel. This creates complications regarding the types of issues raised within these spaces. In many ways, they are the most transparent places for citizens to find out more about what is happening within local communities direct from other citizens. This obviously means that some of the positives will be highlighted, but it also means that many of the issues often kept outside of general knowledge may also be highlighted, and associated with the spaces. For example, conversations about local drug use, anti-social behavior, areas of violence, and other criminal activity.

More recently some councils are using the creation of multiple Twitter channels to direct individuals to the most appropriate information. For example, providing information particularly focused on tourism services, to navigate individuals away from issues related to local fly tipping, or graffiti, and or issues in local neighbourhoods. It is however particularly evident that issues raised in relation to local areas have continued over the past two years with concerns raised that councils may wish to remain out of public channels. There is a fair amount of evidence in the 2013 sample set of citizens visually capturing their concerns using photographs further emphasizing issues in local areas.

J. Social Media Policy and Channel Closure

In 2011 it was noted that there was a limited amount of information regarding specific policy for how councils were going to use Twitter and other platforms as a service, generally categorised by a lack of clear social media policy. In 2013 whilst a social media policy could not be found for all councils, the more prolific councils with a range of social media services had developed clear policy to control this.

Councils are also through their Twitter page adequately communicating a brief description of the channel itself and information such as when the service is monitored. This is a good demonstration of how councils are adapting to ensure that individuals are aware that whilst the web operates 24/7, council services cannot be delivered in the same way.

As identified in [4] councils need to plan for closure of social media channels as much as they need to specify how they are using these in relation to their local community. For those councils pushing conversation directly out of Twitter through other service mechanisms there is a clear question regarding Twitter service lifetime.

K. Sentiment about a place

With the increase in engagement demonstrated by some councils with issues raised by the local population comes a significant challenge for councils, that of demonstrations of community sentiment. Recent developments by Sykora et al. (2013) [17] targeted at developing systems to measure the level of positive and negative feelings for the nation highlight one of the issues with social media. It is clear from reading through sample sets for the various councils involved that an appreciation for the feelings of the local population towards the council can be developed from such materials.

Twitter feeds for councils such as Salford City Council and Camden London Borough Council felt fairly negative in relation to the content exposed, whilst the Wakefield Council feed felt very positive. From a public relations perspective councils need to develop mechanisms for providing positive/negative balances within these community conversational spaces to ensure that the community engages in positive ways with the service, rather than in some circumstances fairly negative attitudes.

One of the concerns related to the above is how councils deal with community issues. One of the problems with moving individuals straight to direct messaging or straight to another mechanism of engagement is that the issue is not closed off within the Twitter space, generating a feeling of a number of community issues. Although Twitter only has a limited number of words, councils need to find a mechanism to positively close off issues, even when moved into other spaces. For example, closing text of "and your issue will be resolved", or something similar, would provide a positive end.

L. What is the difference?

Many of the above points raised as challenges for local authorities in these spaces, can also be linked directly to challenges for large organisations. However, the interesting question is whether there is anything that distinguishes the challenge for e-government. In this area, there are thought to be two major differences.

The first is the increased transparency and profile of the service, money to fund engagement of councils within social media spaces comes directly from the UK public purse. The service needs to have a clear demonstrable impact for citizens within the local community to deem this as successful. The UK press are quick to highlight customer service faillings in online spaces, take for example Tameside Council's experimentation with a virtual customer helpdesk in Second Life, this was described as "absolutely barmy" [18]. Although the value has to be demonstrated in large private organisations, they are not forced to be as transparent, for example with detail regarding cost of service.

It is interesting to see citizens commenting on the state of local government transparency through social media spaces, e.g., "#organisation is in breach of international law, but wants the £4.7bn contract for N London waste. Tell #Camden council 'No' #Palestine #Gaza@" and "People getting turned away... does not feel like an 'open council' tonight. #newcastle" These suggest that UK citizen's are willing to use the service directly to let councils know when they seem to be doing things thought not to be in the public interest.

The second percieved significant difference is that councils could be deemed to have a requirement to engage the citizen, whether this is in citizen democracy, or in conversation regarding quality of services. Other organisations may choose to engage customers but local

councils need to engage citizens. Therefore, understanding where particular customer groups are conversing, and using up to date mechanisms for engaging citizen groups should be a clear part of the strategic engagement for all local councils.

VI. DISCUSSION

The challenges outlined above in response to 2011 and 2013 datasets demonstrate that councils need to think carefully about the ways that they use social media channels to engage the citizen. Arising from the analysis, recommendations can be formed as to what councils should consider in having a presence within these spaces. The recommendations themselves are not new but it is evident from the research that councils are not completely engaging in understanding, and applying, the recommendations in practice.

Understand the channel – Each channel has particular nuances, those councils who understand how these operate seem from the data analysis to be perceived more preferably within the social media community. In addition, those demonstrating an understanding of the channel seem to obtain the greater amount of participation and the greater amount of two way communication. It is clear that as service matures understanding of channel operation also matures. It would seem engaging other councils or organisations with maturity in using social media spaces can be a good first step to channel development. In addition, watching as other councils change practice can help to identify positive mechanisms for transforming conversation through Twitter channels.

Engage the citizen – From the research in 2011 there was evidence that some councils were not engaging in any form of two way dialogue within social media spaces. Those councils who were engaging in a conversational manner with constituents, seemed to have been generating a positive response to their social media engagement. In 2013 all councils were engaging in some way with their local populace through the channel. One reason for this will be local citizens gaining knowledge of channel existence, and another reason will be the growing maturity of service.

Develop policy – It was evident in 2011 within the research that few councils had developed a charter for their and their citizens engagement within social media spaces. Developing clarity over what is and what is not acceptable in the spaces, may form a barrier to conversation but would provide a clearer sense of the general rules of engagement within the space. One argument may be that the spaces themselves often have guidance over what should and should not occur within the social media channel, however, in some cases, this may not be restrictive enough. In 2013 some councils still need to develop and clearly broadcast their policy regarding social media service usage.

Advertise the channel – The best form of advertising in social media spaces is the citizen. However, many organisations are finding innovative ways to highlight their involvement within the spaces. For example, Marseille FC offered their fans the opportunity to design a shirt if they gained a set number of Facebook followers. The best advert for any social media channel is the value added to customer experiences. If the channel is perceived as useful then it is likely that individual users will pass on that information, whilst if the channel provides in the main useless information, then it is likely to be ignored.

Integration – As outlined as a channel above, integration is extremely important for local councils. Managing the ways in which information flows into (in the form of physical posts), and out of the social media space is crucial (in the form of citizen driven requests or data). Social media use should not be in the hands of a singular person tagged with the responsibility of a social media producer, or equivalent. However, the ways in which councils are represented within the space, is extremely important. There is a clear balancing act between bring overly prescriptive and transparent (e.g., tagging each post with a service representatives name) through to allowing for total freedom of employee engagement within the space. In addition, there are significant questions, which arise linked to whether engagement within social media spaces should be part of the role of senior figures such as chief executives of local councils.

Withdrawal – It is evident over the past twelve years that developed spaces will fall in, and out of popularity. Early providers of social media platforms are finding reduced network engagement over time whilst newer platforms have come in, and filled their spaces. Therefore, councils need to manage their portfolio understanding how decisions are made to engage within particular spaces and when decisions should be made to withdraw from engagement within the space. This is a particularly complicated problem if there are constituent groups who maintain loyalty to particular social media spaces. In this space, councils should also consider audit and control mechanisms. If particular networks were to be forced to close it would be problematic if materials stored within those spaces were not stored in one form, or another. In addition, ensuring that any information that would generally be stored for legal purposes in the physical space, can be retrieved when engaging in virtual spaces.

VII. CONCLUSION AND FUTURE WORK

Tweets analysed over the past two years within this study offer a snapshot of council engagement with the public through the social media network platform of Twitter. This snapshot provides information of a specific time for a specific subset of councils. It is clear from this sample that over the past two years an initial indication in 2011 of a lack of true citizen engagement through a social media platform such as Twitter has been transformed with all councils demonstrating clear engagement. Councils are understanding how these services fit within their customer service portfolio and developing mechanisms to help resolve local problems brought forward through conversation in these spaces. Over the next few years these communication channels will continue to mature, as councils continue to re-think the ways in which they encourage individuals to interact with their service provision and as councils continue to review how their services operate.

Future work will continue to focus on usage of social media services by local government organisations. In

particular it may be useful to analyse particular groups of users and their engagement with such services. This type of analysis would provide an understanding of how such services can be used to positively benefit citizen experience.

Utilising approaches for greater automated analysis of tweets may bring greater understanding of the types of tweets provided and the ways in which councils are engaging with the population. In addition, content analysis provided through opinion analysis algorithms being developed at the University of Hull and other institutions will also offer greater dimensions on engagement.

The findings presented within this paper should be of use and be relevant to national and international managers of e-Government web services, government legal teams and senior managers in eGovernment.

REFERENCES

[1] D. Mundy and Q. Umer, "United Kingdom local authority challenges in the use of Twitter and other social media," Proc. Second International Conference on Social Eco-Informatics (SOTICS 2012), Oct 2012, Venice, Italy.

[2] Internet World Stats, "Facebook growth stats," http://www.internetworldstats.com/facebook.htm 12.12.2013.

[3] Twitter Blog, "Twitter turns six," https://blog.twitter.com/2012/twitter-turns-six 12.12.2013.

[4] D. Mundy and Q. Umer, "An analysis of UK council use of the social network – Twitter," Proc. 12th European Conference on E-Government, Barcelona, 2012 pp. 502-511.

[5] D. Landsbergen, "Government as part of the revolution: using social media to achieve public goals," Electronic Journal of e-Government, Volume 8, Issue 2, pp135-147, 2010.

[6] J. J. Zhao and S. Y. Zhao, "The impact of e-government use of social networking media on government transparency and public participation," Télescope, Vol. 18, 1-2, 2012, pp. 44-61.

[7] R. Cole, "Social media: what does it mean for public managers?," PM Magazine, Volume 91, Number 9, 2009.

[8] J. Hendricks and R. Denton, Communicator in Chief, Lexington Books, 2010, ISBN-13: 978-0739141052.

[9] J. Skinner, "Social media and revolution: the arab spring and the occupy movement as seen through three information studies paradigms," Sprouts: Working Papers on Information Systems, 11(169), 2012.

[10] G. Gulati and C. Williams, "Communicating with constituents in 140 characters or less: Twitter and the diffusion of technology innovation in the United States Congress," Presented at the Midwest Political Science Association, Chicago, Illinois.

[11] R. Waters and J. Williams, "Squawking, tweeting, cooing, and hooting: analyzing the communication patterns of government agencies on Twitter," Journal of Public Affairs, Volume 11, Issue 4, pp. 353–363, 2011.

[12] A. Tumasjan, T. Sprenger, P. Sandner, and I. Welpe, "Predicting elections with Twitter: what 140 characters reveal about political sentiment," Proc. Fourth International AAAI Conference on Weblogs and Social Media, Washington, D.C., 2010, pp. 178-185.

[13] R. Romsdahl, "Political deliberation and e-participation in policy-making," CLCWeb: Comparative Literature and Culture, Purdue University Press 7(2), 1-1, 2005, pp. 1-11.

[14] C. Danis, M. Bailey, J. Christensen, J.Ellis, T. Erickson, R. Farrell, and W. Kellogg, "Mobile applications for the next billions: a social computing application and a perspective on sustainability," Proc. 2nd Workshop on Innovative Mobile Technology and Services for Developing Countries (IMTS-DC 09), 30-31 July, 2009, Kampala, Uganda.

[15] J. Rooksby and I. Sommerville, "The management and use of social networks in a government department," Journal of Computer Supported Cooperative Work (CSCW), October 2012, Volume 21, Issue 4-5, pp 397-415.

[16] N. Williams, "Template Twitter strategy for government departments," http://neilojwilliams.net/missioncreep/wp-content/uploads/2009/07/17313280-Template-Twitter-Strategy-for-Government-Departments.pdf 12.12.2013.

[17] M. Sykora, T. W. Jackson, A. O'Brien, and S. Elayan, "EMOTIVE ontology: extracting fine-grained emotions from terse, informal messages," IADIS Intelligent Systems and Agents Conference, Prague.

[18] The Telegraph, "Council scraps £36,000 virtual town hall in Second Life", 21 July 2010, http://www.telegraph.co.uk/technology/internet/7901312/Council-scraps-36000-virtual-town-hall-in-Second-Life.html 12.12.2013

Towards a System that Relieves Psychological Symptoms of Dementia by Music

Chika Oshima
Japan Society for the Promotion of Science,
Faculty of Medicine, Saga University
Saga, Japan
chika-o@ip.is.saga-u.ac.jp

Koichi Nakayama
Department of Information Science, Saga University
Saga, Japan
knakayama@is.saga-u.ac.jp

Naoki Itou
Intermedia Planning, Inc.
Tokyo, Japan
n_itou@ipi.co.jp

Kazushi Nishimoto
Research Center for Innovative Lifestyle Design,
Japan Advanced Institute of Science and Technology
Ishikawa, Japan
knishi@jaist.ac.jp

Kiyoshi Yasuda
Kyoto Institute of Technology
Chiba Rosai Hospital
Kyoto and Chiba, Japan
fwkk5911@mb.infoweb.ne.jp

Naohito Hosoi
Sodegaura Satsukidai Hospital
Chiba, Japan
hosoi@mail.satsuki-kai.or.jp

Hiroshi Okumura
Department of Information Science, Saga University
Saga, Japan
oku@is.saga-u.ac.jp

Etsuo Horikawa
Faculty of Medicine, Saga University
Saga, Japan
ethori@med.saga-u.ac.jp

***Abstract*—MusiCuddle is a system to calm the symptoms of patients with mental instability who repeat stereotypical utterances. The system presents a short musical phrase whose first note is the same as the fundamental pitch (F0) of a patient's utterances. We performed a case study to investigate how a patient's behaviors changed with MusiCuddle. The results suggested that the phrases presented by MusiCuddle may provide patients with an opportunity to stop repeating stereotypical utterances. Then, we added a vocoder function to MusiCuddle so that patients would be able to attend to the music more. We examined whether the mood of university students changed or not according to music presented with the vocoder function. We found significant differences between major harmonies and minor harmonies for the "cheerful" and "negative" moods. Namely, when a person's voice is combined with cheerful sounds, he/she can become cheerful. However, when we conducted a case study to expect a patient's repetitive utterances changed or stoped by the sound from the MusiCuddle with the vocoder, the participant's utterances did not change. We discussed reasons of the result from an aspect of characteristic of a patient according to a cause disease of dementia.**

Keywords*—*MusiCuddle, vocoder, FTD, Harmony in a major and minor key

I. INTRODUCTION

We are structuring a music accompaniment system to calm the symptoms of patients with mental instability who repeat stereotypical utterances. "MusiCuddle [1][2]" is a system that presents a short musical phrase. The system determines a pitch at a predetermined interval on the basis of a sound extraction technique [3]. Then, the system plays a prepared Musical Instrument Digital Interface (MIDI) sequence (a phrase) the first note of which is the same as the F0 of the patient's utterance.

The concept of MusiCuddle is derived from the "iso-principle [4]," which is a theory of music therapy, and a case of an autistic child a famous music therapist treated by extracting approximate pitches of the child's screaming and improvising based on these pitches [5]. "Iso" simply means "equal," that is, the mood or the tempo of the music must initially have an "iso" relationship with the mood or tempo of the patients. If a client is distressed or agitated, then the quality of the music should initially match his or her mood and energy [6].

In this paper, first, we introduce the MusiCuddle and results of a case study with using the system [1][2]. We performed a case study in which one of the authors used MusiCuddle to present phrases to a patient with dementia who repeated stereotypical utterances. The symptoms of dementia are divided into core symptoms and behavioral and psychological symptoms of dementia (BPSD). BPSD includes agitation, aggression, wandering behavior, hallucinations, delusions, and repetitive stereotypical utterances. However, appropriate care is thought to alleviate and slow the progression of these symptoms. Music is a method known to alleviate the symptoms of dementia.

Second, on the basis of the results of the experiment, we added a "vocoder" to MusiCuddle. The vocoder allows an individual to hear his/her voice becoming a part of the instrumental sound according to a musical phrase presented by MusiCuddle. Because our target population repeat utterances quite frequently, it will be hard for them to listen to the music presented by MusiCuddle. Therefore, the utterances should be combined with music sounds in real time, as their attention will be more likely to shift to the music than when they listen to music in parallel with their utterances.

Furthermore, if the musical phrases from MusiCuddle can manipulate the mood of patients with mental instability

and make it more pleasurable, they may temporarily stop repeating utterances. There are studies showing that mood affects memory and cognitive processes [7]. For instance, Taniguchi [8] used music to manipulate subjects' mood. In [9], he considered a relationship between characteristic of music and mood induced by music. He proposed the Affective Value Scale of Music (AVSM) to indicate the property of musical pieces on the basis of 24 adjectives on five levels. Then, he conducted an experiment in which female students were rated on both the AVSM and the Multiple Mood Scale (MMS) [10], which is to evaluate subjects' mood by themselves for the five pieces, finding a significant relationship between the AVSM and the MMS. The result has shown that music can be a trigger to induce a mood.

Then, in this paper, we examined the contribution of harmonies in major and minor keys to mood induction for healthy subjects with the vocoder. They read a gloomy poem when their utterances were combined with music sounds by MusiCuddle with the vocoder. After reading the poem, they evaluated their current moods.

Finally, we performed a case study using the MusiCuddle with the vocoder for a patient with dementia who repeated stereotypical utterances. Her utterances were combined with music sounde in real time.

In the next section, we illustrate the MusiCuddle and experiments that the author presented music phrases to a patient with dementia using MusiCuddle. Section III describes the contribution of harmonies in major and minor keys to mood induction for healthy subjects by MusiCuddle with the vocoder. Section IV concludes this paper and outlines future works.

II. MusiCuddle: the first note of a phrase is the same as the F0 of the patient's utterance

We presented a music accompaniment system, "MusiCuddle" that presents a short musical phrase. Then, we conducted a case study using MusiCuddle for a patient with dementia.

A. Extract pitches from the utterances

Figure 1 shows a user interface of MusiCuddle [1][2]. MusiCuddle is a system that presents music when an operator (e.g., a caregiver) pushes any of the keys of the electronic keyboard or a button on the interface of the system. Previously, we have to select a folder of musical phrases and move it into the same folder as MusiCuddle. Once the play button is pushed, the system continuously extracts pitches (F0) from sounds (the intended patient's utterances). When the operator pushed the trigger's button again, the system determines a pitch at a predetermined interval. Then, the system selects a musical phrase file in the database on the basis of that extracted pitch. The first note of the musical phrase is the same as the pitch extracted from the patient's utterances.

We employ a pitch extractor to extract pitches (i.e., C, D, E) from the patient's utterances. This is based on the technique for extraction of sounds that have unstable pitches and unclear periods, such as natural ambient sounds and the human voice, into musical notes [3].

In the original system shown by [3], if the operator gave a start trigger, the system would initiate the processing to obtain the F0 (fundamental frequency) time series from the acoustical signals (i.e., a singing voice), which were being recorded via the microphone. The short-term F0 estimation by Fast Fourier Transform (FFT) and Inverse Fast Fourier Transform (IFFT) for the power spectral is repeated until the system catches an end trigger from the operator. The system then calculates a histogram of pitches with the F0 time series between the start and end triggers. Finally, only the most frequent pitch is selected and is output as the pitch of the period.

For our research, some processing designs were modified. Figure 2 shows the processing of the system. Considering the attitude of the operator, we would assume that the triggers would be input after the operator catches the utterance of the patient. Therefore, we omitted the start trigger. The system starts a short-term F0 estimation just after invocation of the system and continues it thereafter. When the operator inputs a trigger that is regarded as an end trigger, the system calculates a representative pitch for a predetermined period just before the trigger based on the above-mentioned method. Then the system plays a prepared MIDI sequence (a musical phrase) that corresponds to the representative pitch. These modifications of our system improve usability by reducing the time lag between the input of the trigger and the output of phrase.

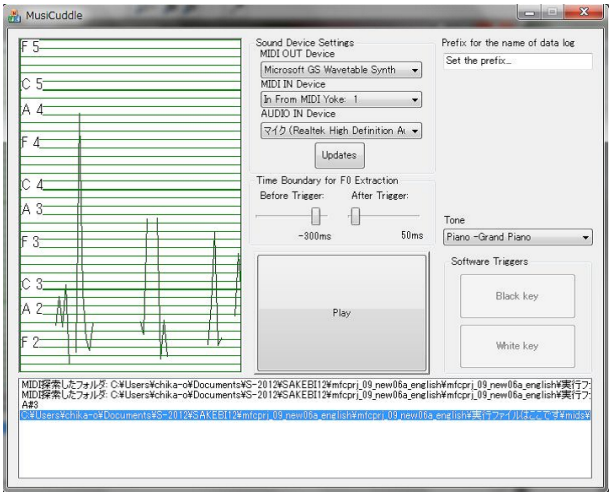


Fig. 1. User interface of MusiCuddle for a caregiver.

To extract the F0 against the mixed acoustical signal of the patient's utterance and the musical phrase output from the speaker, our system needs two of the same microphone (ideally one stereo microphone) and one speaker. Figure 3 shows the setting of the microphones. The microphones are set in front of the speaker to record the speaker's sound at the same level from both microphones. On the other hand, both microphones are displaced against the patient to record the levels of the patient's utterance that are clearly different. The system calculates the differential signals from the signals of both microphones to cancel the sounds of the MIDI sequence where they are localized in the center position. The F0 estimation is then determined with these differential signals.

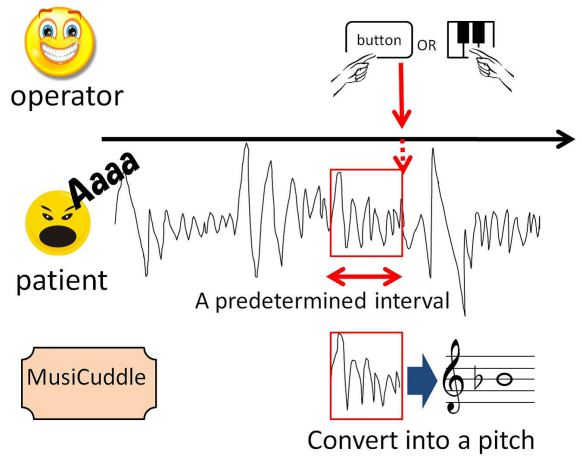


Fig. 2. How to convert an utterance into a pitch.

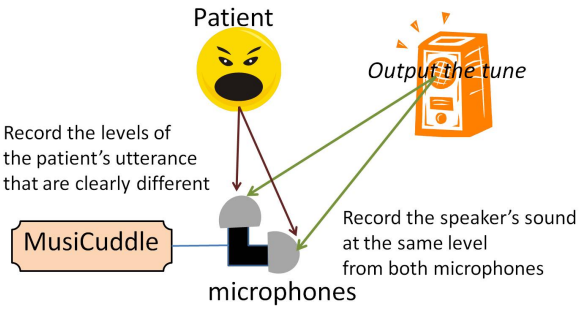


Fig. 3. Calculates the differential signals from both microphones to cancel the sounds of cadence.

B. Case study using MusiCuddle

We conducted a case study to investigate how a patient's behaviors changed with the application of MusiCuddle [1][2]. The symptoms of the patient targeted for this study was severe. This case study was conducted for a very short period, and we could enroll only one patient. Therefore, it is not appropriate to make any cognitive assessment [12] or to examine the patient's abilities with respect to activities of daily living [13]. Instead, we record the patient's utterances under conditions with and without the use of MusiCuddle and compare them in order to estimate the influence of MusiCuddle.

1) Ethical Considerations: This case study was approved by the Research Ethics Board of Saga University. The participant in the case study, who is a patient with dementia, her husband, and the hospital director were informed about the intentions of the case study and the treatment of personal information. Moreover, they were informed that they could withdraw from the case study at any time. Then, we obtained written consent from them.

When we conducted this case study, the hospital director and nurses worked on the same floor and could check on the condition of the participant. If the presentation of sounds from MusiCuddle had not been appropriate and the participant became more agitated, we would have had to abandon the case study immediately.

2) Participant: The participant was a 72 year-old, hospitalized patient with severe FTD (frontotemporal dementia). She repeats stereotypical utterances for many hours each day. Moreover, when she is agitated, she locks herself in a restroom for a long time while repeating stereotypical utterances. However, she is lucid enough to remember some nurses' names and greet them clearly. She can answer the date and the exact time. Her score on the HDS-R (Revised Hasegawa's dementia scale) [14] was 17 two years ago. The score shows she was mild dementia.

The following is an example of her usual utterances. This example was uttered in about thirty seconds. "P" means "Participant."

P: haittayo (repeated eight times) mashitayo imasen masen imasendesu masen (repeated three times)

"haittayo" means "have been entered" as well as "imasen" means "not being here." "mashitayo" may be fragment of "imashitayo." "imashitayo" means "being here."

Although she utters many kinds of sentences, most of them are rhythmical and fit into the same meter. Figure 4 describes some examples of her sentences. One of the authors dictated the rhythms of these sentences. These examples show that although the sentences are different, they fit into four-four time.

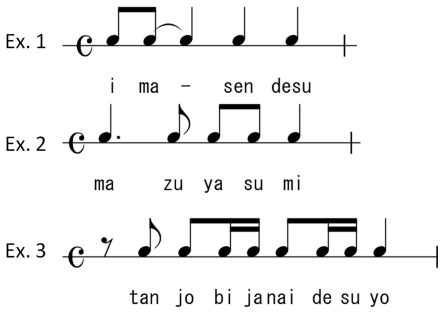


Fig. 4. The participant's sentences fit into four-four time.

She repeats stereotypical utterances especially when hungry. She often locks herself in a restroom from around eleven o'clock a.m. until lunchtime, and from around one o'clock p.m. until snack time while repeating stereotypical utterances nearly incessantly. However, she sometimes responds to nurses when they talk to her.

3) Preliminary experiment: Adopting the iso-principle, one of the authors attempted to utter according to the participant's utterances. We think that the author's utterances had the same tempo, rhythm, and pitch as those of the participant.

When the participant was agitated and repeated the same sentences, one of the authors repeated sentences in the same melody corresponding with the participant's repetition (the same tempo, rhythm, and pitches). Figure 5 shows these sentences in musical notation. Sentence A means the participant's

sentence. First, the author tries to repeat the participant's sentences in rhythm. Namely, both of them repeat Sentence A, "i-ma-se-n-yo (not being here)." Second, the author tries to repeat a different sentence using the same melody as the participant's in rhythm. Namely, the author repeats Sentence B, "go-han-de-su-yo (Time for lunch.)," although the participant is repeating Sentence A.

In the first trial, the participant turned around to pay attention to the author. However, she kept repeating the same sentence in harmony with the author's repetition. Her utterances became louder. In the second trial, the participant changed from the sentence A to the author's sentence, "go-han-de-su-yo" in the same melody. Then, the participant left the restroom and went toward a table for lunch. In a moment, however, she returned to the restroom and repeated the same sentences.

When the author repeated sentences in accordance with the participant's repetition and used the same melody, the participant kept repeating the same sentence in a loud voice. The author's utterances could have caused increased symptoms of agitation in the first trial.

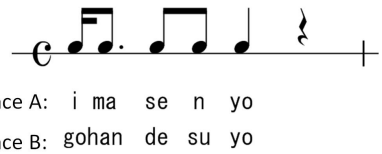


Fig. 5. The author repeated sentences in accordance with the participant's repetition.

4) Method: We stood by from ten o'clock a.m. to noon and from one o'clock p.m. to half-past two p.m. for 2 days. After this case study, we compared the participant's utterances with music with those without music (see Section II-B6) to estimate the influences of MusiCuddle. Therefore, we set two time periods, one with the use of MusiCuddle and one without MusiCuddle.

During the time with the use of MusiCuddle, we started MusiCuddle and selected a musical phrase. When the participant began to repeat stereotypical utterances, we presented the musical phrases arbitrarily by giving triggers to MusiCuddle.

The experiment was conducted in a hospital where the participant was hospitalized. Figure 6 shows the setting of the case study. The music was presented through a wireless cuboidal speaker with Bluetooth, which measured, 123×36×35 mm, and the patient's utterances were recorded through a wireless, columnar microphone with Bluetooth measuring about 75 mm in height and 24 mm in diameter. These devices were set on the door of the restroom.

Our system requires two of the same microphone (one stereo microphone). When the operator inputs the trigger, even when the previous musical phrase is being presented, the system extracts F0 against the mixed acoustical signal of the patient's utterance and the musical phrase being presented from the speaker. However, it is not so safe to use the stereo microphone in the hospital, because it is large in size and wired to its receiver. Thus, we did not use the stereo microphone in this experiment, and the operator did not input triggers when the previous musical phrase was being presented.

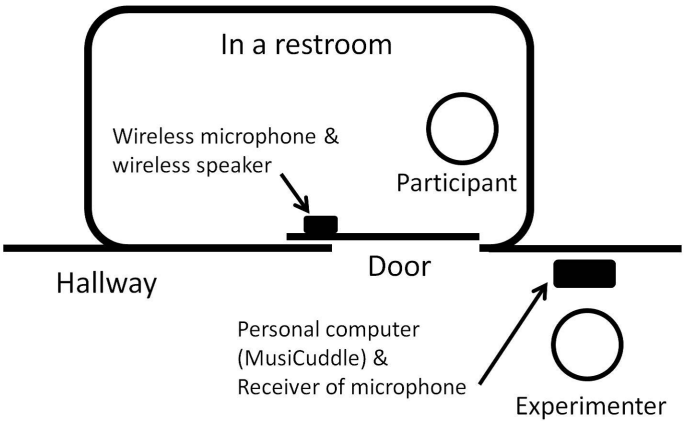


Fig. 6. A small wireless speaker and microphone are set on the door of the restroom.

5) How to use MusiCuddle: When the participant is agitated and repeats stereotypical utterances in a restroom, an operator (one of the authors) presents music by using the MusiCuddle system in front of the restroom. The operator listens to the patient's utterances outside of the restroom. When the operator finds a period during which the patient utters an almost stable pitch, she clicks the trigger button. Once the trigger button has been clicked, a musical phrase is retrieved from the database on the basis of the detected pitch, and it is automatically performed to overlap with the participant's utterances. In this case study, we prepared seven type of musical phrases: Four chords (Major seventh, Quarter note, and No volume change), Cadence, "Yuki," "Akaikutsu," "Hana," "Tsukinosabaku," and Stereotypical utterance (i-ma-se-n-yo) [1][2]. All of the musical phrases consist of very short phrases lasting 3∼30 seconds. The operator selected musical phrases considering the participant's reactions and condition. The operator clicked the trigger button again to perform the next musical phrase when the performance of the current musical phrase ended.

6) Analysis method: In this case study, we investigate how MusiCuddle influences the patient's stereotypical utterances. If the music presented from MusiCuddle distracted the patient's attention from her stereotypical utterances, her utterances would be disrupted and she would stutter. Therefore, we compare the participant's utterances while listening to music with those without music. Especially, we focus on the patient's stuttering to detect distraction of the patient's attention.

The participant's utterances are segmented into small sentences according to the method of repetition. One of the authors decided segmentation points according to the meanings of utterances by reference to the participant's breathing. For example, the following utterances (P1) are segmented like the next line (P2).

P1:imasendesuimasendesuhitoyasumimazuyasumiimasendesuyo
P2:imasendesu (not being here) / imasendesu (not being here) / hitoyasumi (taking a rest) / mazuyasumi (taking a rest) /

imasendesuyo (not being here)

Then, we analyzed the influences of the music on the patient's utterances. First, we determined whether each sentence was uttered "with music" or "without music" on the basis of the following conditions (see Fig. 7):

1) If the patient uttered a sentence while a musical phrase was being performed, the sentence was considered to be uttered with music.
2) If the patient began to utter a sentence just after a musical phrase had finished, the sentence was considered as being uttered with music.
3) If a musical phrase started after the patient had started uttering a sentence, the sentence was considered to be uttered without music.
4) Otherwise, the sentence was considered as being uttered without music.

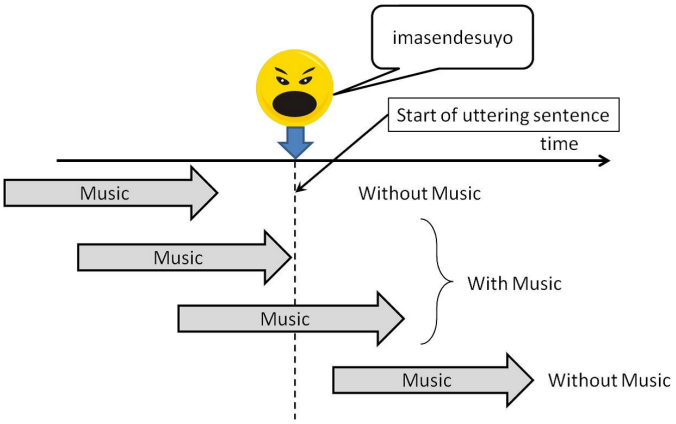


Fig. 7. Determination of whether each sentence was uttered "with or without music."

In the following example, the musical phrase was presented in the middle of "hirugohandeha (fragment of "It is not lunch time")." Therefore, "masende (fragment of "not being here")" and "hirugohandeha" were considered as being uttered without music, while "imasendesu (not being here)," "ima (fragment of "not being here")," and "gohanden (fragment of "It is not lunch time")" were considered as being uttered with music.

P: masende hirugohandeha imasendesu ima gohanden
[ ]
(Start music) (Stop music)

In this case, we consider that the sentence "hirugohandeha" was unaffected by music. Moreover, we consider that "gohanden" was affected by the music, because this sentence started to be uttered immediately after the presentation of the music.

Next, we find sentences on which the participant stuttered. She often repeats several stereotypical sentences without any slight changes many times (see Section II-B2). However, if the music distracted her attention from her repetition of stereotypical utterances, she stuttered, uttering only a part of stereotypical sentence or a sentence different from a stereotypical sentence. Therefore, if we find such a sentence including words that were part of an immediately preceding sentence (but not exactly the same as the immediately preceding sentence), she was considered to have stuttered.

We determined whether each sentence included words that were part of the immediately preceding sentence. In the following example, we consider that she stuttered, as "ima" is included in the immediately previous sentence "imasendesu:"

P: imasendesu ima

7) Results: The intended records documenting our analysis of the participant's utterances constitute only three parts of the entire recorded dataset; the lengths of the three parts are 16, 8, and 3 minutes. Although we recorded for a much longer time period, we could not use the other parts because of extraneous noises that masked the patient's utterances. It seems that the restroom's iron door blocked communication between the wireless microphone and the personal computer. Moreover, we could not record when the participant moved to some unexpected rooms. Therefore, there is a large gap between the time of using MusiCuddle and the total recording time.

Of the 27 minutes of intended records, the total time of music presentation was 6 minutes and 54 seconds, approximately one-fourth of the total recording time. For the first and second recording sessions, we presented the phrases using MusiCuddle. In contrast, we did not present phrases at all during the third recording session.

The participant emitted utterances at all times during the experiment. Table I shows the kinds of sentences. Six hundred eighty sentences were segmented (84 kinds) in 27 minutes. The most uttered sentence was "imasendesu" (201 times).

In many cases, the contents of the sentences were almost the same, even when their constituent words varied slightly. For example, "imasen" and "imasendesu" have the same meaning, "I am not here." Moreover, in certain instances, only parts of sentences were uttered ("ima," "imasende").

She repeated the same words many times, uttered different words in sequence, or uttered slightly different words continuously both in the rhythm of the presented-music and not. In the following example, she repeated the same words many times:

P: mazuyasumi mazuyasumi mazuyasumi mazuyasumi...(total number of repetitions was seven)

In the following example, she uttered different words in sequence:

P: gohandashimasendesu mazu imasendesu oyatuja imasendesu mazuyasumi

In the following example, she uttered slightly different words continuously:

P: imasende imasendesu imasen imasendesu ima imasendesu

Most of the sentences were rhythmical and fitted into four-four time (see Fig. 4). However, short sentences such as, "ima" and "mazu" fitted into four-one (irregular) time.

TABLE I. SENTENCES SEGMENTED FROM THE PARTICIPANT'S UTTERANCES IN THE CASE STUDY.

Estimated meaning	Sentences
I am (not) here.	imasu (1), imasendesu (201), imasen (48), imasende (43), ima (5), deimasendesu (1), sokoniimasen (1), imasenyo (1), uruchiimasendesu (1), ryugaimasende(1)
It is (not) lunch time.	mazugohandesu (78), mazugohan (14), mazugohande (6), hirugohannarimasendesu (5) gohandesu (4), hirugohandashimasendesu (2), hirugohannarimasende (2), gohandashimasendesu (2) haisugugohandashimasendesu (2), mazugohandesuyo (2), gohan (2), gohanden (1), gohannarimasendesu (1) gohannaidesuyo (1), hirugohande (1), gohandashimasende (1), gohanninarimasu (1) gokaimenogohandashimasendesu (1), mawarinogohangoyamoyashisendesu (1)
First,	mazu (34), mazuyasumi (33), mazudesu (32), ma (6), mazuya (3), mazudesu (2), mazude (1), mazugo (1), mazuyasu (1)
Not do	masende (10), masendesu (6), masendesuyo (2)
Bath time, Break	ofurohaittadesuyo (2), ofuro (1), ofurojaimasende (1), hitoyasumi (1)
(Not) Birth day	tanjobijanaidesu (3), tanjokainaidesuyo (3), tanjobijanaidesu (1), tanjobijaimasendesuyo (1) tanjobijaarimasendesu (1), tanjobijaarimasendesuyo (1)
Time	1ji40fundesuyo (13), yoruninarimasendesu (9), 1jihandesuyo (8), 3jihanninarimasendesu (8), yoruninarimasendesuyo (5) handesuyo (4), 3jihandesuyo (4), 1jihande (3), 2jihandesuyo (2), 3jininarimasendesuyo (2), 1jihandesune (1) 1ji10fundesune (1), 1ji (1), 1jihandesuyo (1), 3jihanni (1), 3jihanninarimasendesuyo (1), mou3jininarimasendesu (1)
Snack time	oyatudesuyo (1), oyatujaimasendesu (1), keikihanaidesuyo (1), keikihanaidemasendesu (1), keikihanaitodesu (1)
Soon	suguha (1), suguhanaidesu (1)
"Yuki"	zunzuntumoru (2)
Question	imashitaka (1)
Greeting	konnichiha (1)
Others	dojoninarimasende (1), ugoninarimasende (1), sonouchimasende (1), mashi (1), bokujaarigatoarigato (1), basyohanaidesuyo (1)

Values in parentheses show the numbers of times each sentence was uttered.

Table II shows the comparison between "with music" and "without music." The numbers of different sentences uttered were 114 with music and 179 without music. The total recording time was 27 minutes, and musical phrases were presented for 6 minutes 54 seconds of that time. Changes in the sentences uttered by the participant numbered about 16 per minute with music and 9 per minute without music. Therefore, we can say that the participant changed her utterances more often with than without music.

Next, we determined whether each sentence included words that were part of the immediately preceding sentence in order to determine on which sentences the participant stuttered. The results indicated that with music, 94 out of 114 sentences (82.5%) included words from the immediately preceding sentence (see Section II-B6). On the other hand, without music, that rate was 41.3%. This result shows that the rate of sentences including words from the immediately preceding sentence was higher with than without music. If the participant stuttered, we consider that the music distracted her attention from repeating her stereotypical utterances (see Section II-B6). The results indicate that MusiCuddle may give patients an opportunity to stop repeating utterances.

In the following example, a sentence changed into a completely different sentence when music was not presented (without music):

P: mazu imasendesu oyatsuja mazuyasumi
(without music)

The following is an example in which a sentence included the word from the immediately preceding sentence when music was presented (with music):

P: mazugohandesu **mazu**...
[ ]
(Start music) (Stop music)

TABLE II. THE NUMBERS OF CHANGES IN REPEATED SENTENCES.

	with music	without music
changing sentence (ALL)	114	179
include the words of the immediately previous sentence.	94	74
rate (%)	82.5	41.3

8) Discussion: The participant tended to stutter when each phrase was presented from MusiCuddle. The music might shift the participant's interest to music from the repetition of stereotypical utterances. On the other hand, when one of the authors repeated the participant's sentence using the same melody and rhythmic pattern, she did also pay attention to the author (see Section II-B3). However, the participant kept repeating the same sentence.

Namely, patients might attend to the phrase according to their similarity in pitch. Meanwhile, their attention may be deflected away from their repetitive stereotypical utterances if the melody is too strikingly different from their utterances. So, the phrases presented by MusiCuddle may provide patients with an opportunity to stop repeating stereotypical utterances.

III. MOOD INDUCTION USING MUSICUDDLE WITH A VOCODER: MAJOR VERSUS MINOR HARMONIES

The result of the case study using MusiCuddle suggested that the mental instability patient's attention might shift away from her repetitive stereotypical utterances to the music (see Section II-B). We expect that the utterances should be combined with music sounds in real time, as their attention will be more likely to shift to the music than when they listen to

music in parallel with their utterances. Therefore, we added a vocoder function to MusiCuddle.

A. Add a vocoder function to MusiCuddle

We added a vocoder function to MusiCuddle [1][2] so that patients would be able to attend to the music more. The vocoder is an audio processor that captures the characteristic elements of an audio signal and then uses this characteristic signal to affect other audio signals. The modulator extracts the fundamental frequencies of the voice and converts them into levels of amplitude on a series of band pass filters. Then, these band pass filter signals are passed onto the carrier wave and the final sound is created.

Fig. 8 shows the vocoder's connection to MusiCuddle. The patient's utterances are input to MusiCuddle and the vocoder (synthesizer) by two kinds of microphones. MusiCuddle extracts notes from these utterances and selects a phrase. Then, the phrase (MIDI sequence) is sent to the vocoder. The vocoder performs the MIDI sequence using the tone of the synthesizer and the patient's voice. The patient can hear his/her utterances combined with MIDI sequence.

B. Research aim

We want to examine whether the mood of the patient with mental instability changes or not according to music presented with the vocoder function. There is no research of mood induction using the vocoder function. Since it is difficult to gather the intended patients who repeat utterances continuously and it is difficult for them to express their moods in language, the subjects of this paper are healthy university students.

Moreover, we examine the difference between a mood induced by harmonies in a major key and a mood induced by harmonies in a minor key. Altshuler showed if a patient is gloomy, the quality of the music should initially be gloomier rather than happier [4]. Itoh [15] showed that individuals in a depressive state become relaxed when they listen to gloomy and calm music. After introducing these kinds of music, however, the mood of the music should gradually change to the target mood (Level attacks [4]). Takeuchi [16] conducted an experiment on university students in a state of depression and found that the group of subjects who heard music that progressed from sad to happy were put in a happier mood.

C. Pre-experiment

In this section, subjects evaluated their impressions of two musical phrases. These phrases were used in the main experiment (see Section III-D).

1) musical phrases: Hevner [17] indicated that the expressiveness of a modality, either major or minor, is more stable and more generally understood than that of any other musical element. He showed that major keys are strongly associated with happiness, gayety, playfulness and sprightliness and minor keys are deeply related to sadness, sentimental yearning, and tender effect. Moreover, consonant chords work for "delightful [18]" "cheerful [19]," and dissonant chords work for "exciting [18]" and "overcast [19] [11]."

In the main experiment, we examined the difference between a mood induced by harmonies in a major key and a mood induced by harmonies in a minor key with using the vocoder. The properties of two musical phrases should be similar although the mode (major or minor) and the harmonies are different. Therefore, we pick out these phrases from the same music piece, "Chaconne" rearranged by Busoni for a piano solo on the basis of "Chaconne from Partita No.2 for solo violin in D minor, BWV 1004" composed by Bach. Figs. 9 and 10 show two different kinds of phrases. We extracted the harmonies in the major key from bars 138-145 with one incomplete bar as well as the harmonies in the minor key from bars 1-8 with one incomplete bar. In the original score of Chaconne, there are many kinds of note values and some passing notes between the chords. However, we did not consider rhythm and passing notes. All notes in the scores were changed to whole notes due to the features of the vocoder (see Section III-A). The scores were transformed into two MIDI data files in advance. The tempo was a beat of 60 quarter notes for one minute. Namely, both phrases could be presented in about one minute.

The subjects listen to the phrases produced by the sound source of a synthesizer, "microKORG XL+ (Korg)." The synthesizer effect was made by the "ROCK" genre and "POLY SYNTH" category in microKORG XL+. In the main experiment, we used the same sound effect. However, we also used the vocoder function. Therefore, the feeling of sounds we got were different between in the pre-experiment and in the main experiment.

2) Method: The subjects were 132 engineering university students ranging from 18 to 20 years of age. Sixty-one of the subjects evaluated the harmonies in a major key (D major) first and then in a minor key (D minor). The rest of the subjects evaluated them in reverse order. We prepared AVSM [9] for the subjects to evaluate the affective value of the two phrases. The AVSM consists of 24 adjectives that can be divided into five dimensions: uplift (uplift and dysphoria), familiar, strong, lightness, and stateliness. The subjects were asked to evaluate the 24 adjectives (items) on a five-point scale: It does not apply to the adjective at all (1); It does not apply to it very much (2); I cannot say either way (3); It applies to it a little (4); It applies to it very much (5).

3) Result: We performed t-tests on the data for 24 items. Table III shows that there were significant differences between the major phrase condition and the minor phrase condition on 16 items. In particular, the evaluations of three items, "melancholy," "miserable," and "gloomy" became opposite. Their averages were more than 4-point in the evaluation for the minor phrase. Their averages were less than or equal to 3-point in the evaluation for the major phrase. These results showed that the phrases were suitable for use in the main experiment.

D. Experiment: mood induction using MusiCuddle with a vocoder

Each subject read a gloomy poem and indicated his/her current mood. Then, he/she read the same poem using MusiCuddle with the vocoder and indicated his/her mood again. The music presented from MusiCuddle included two kinds of phrases that were evaluated in Section III-C. We examined whether the mood induced in subjects using the vocoder differed according to the music from MusiCuddle.

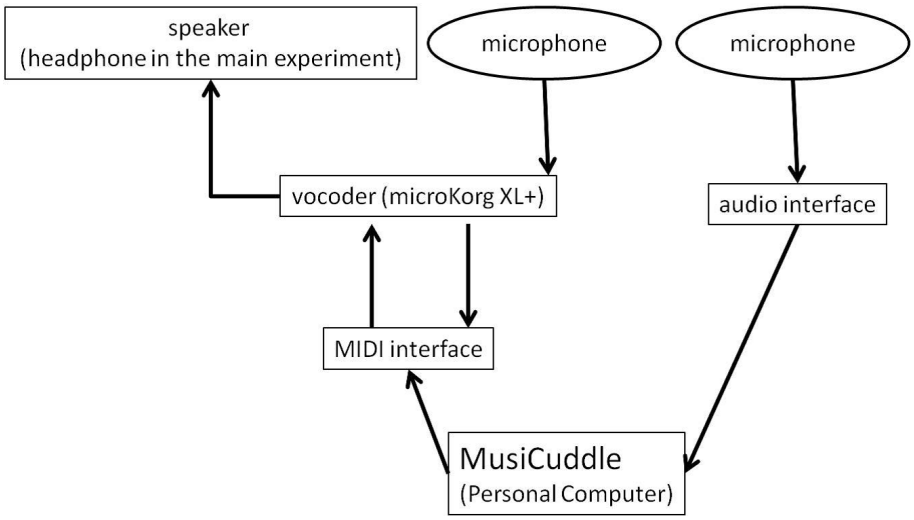


Fig. 8. Connection with a vocoder.

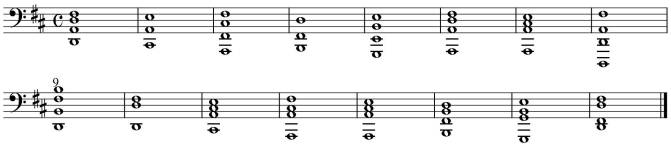


Fig. 9. Harmony in a major key.

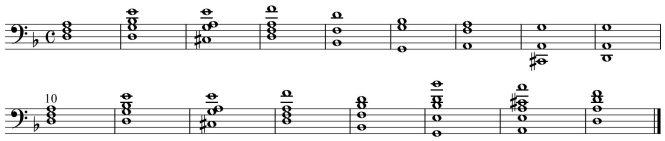


Fig. 10. Harmony in a minor key.

TABLE III. EVALUATIONS OF TWO HARMONIES.

demension	item which has a difference	Average minor	Average major	t-value
dysphoria	melancholy	4.01	2.48	10.90**
dysphoria	miserable	4.14	2.54	11.59**
dysphoria	sad	4.41	3.02	8.32**
dysphoria	gloomy	4.38	2.62	8.15**
uplift	cheerful	1.38	2.38	8.02**
uplift	delightful	1.38	2.82	10.07**
uplift	joyful	1.45	2.72	7.05**
uplift	bright	1.63	3.40	7.73**
familiar	tender	2.02	3.47	11.25**
familiar	calm	2.33	3.62	5.23**
familiar	sweet	2.37	2.77	2.38*
strong	vehement	2.55	1.95	3.60**
lightness	hilarious	1.50	2.11	3.91**
lightness	eathery	1.63	2.49	4.02**
stateliness	solemn	3.62	2.79	5.13**
stateliness	ceremonious	3.29	2.62	3.61**

* $<$ 5%, ** $<$ 1%.

1) Ethical Considerations: This experiment was approved by the Research Ethics Board of Saga University. The subjects were informed about the purpose of the experiment and the treatment of personal information. Then, we obtained written consent from them.

2) Method: The subjects were 12 engineering university students between 21 and 24 years old. Two female students were included in the subjects.

Fig. 11 shows the method of the experiment. The subjects participated in the experiment one by one. First, each subject read 28 words to him/herself. These words were selected from "Personality trait words [20]." They were expressed impressions of "darkness," "stay in one's shell," and "very sensitive." Second, each subject was asked to read a poem that a 20-year-old man had composed while in a gloomy mood and put on the Internet. We expected that the subjects would become gloomy while completing these tasks.

After reading the poem, each subject was asked to indicate his/her current mood by filling out a questionnaire. The questionnaire consisted of 40 mood-related items that were selected from the MMS [10]. The 40 items consist of 10 items on each of four dimensions: dysphoria/fatigue, active pleasure, and non-active pleasure. We lined up one set item that four items is extracted from each four dimensions. We could make 10 set items. The order of each set differed depending on the subject and on the number of times (one subject responded to the questionnaire twice). The subjects were asked to evaluate the 40 items on a four-point scale: I do not feel it at all (1); I do not feel it very much (2); I feel it a little (3); I feel it clearly (4).

Next, each subject read the same poem with headphones on. An experimenter pushed the trigger button for MusiCuddle when the subjects read the title of the poem. MusiCuddle calculated a representative pitch for a predetermined period just before the trigger to extract the pitch of each subject's

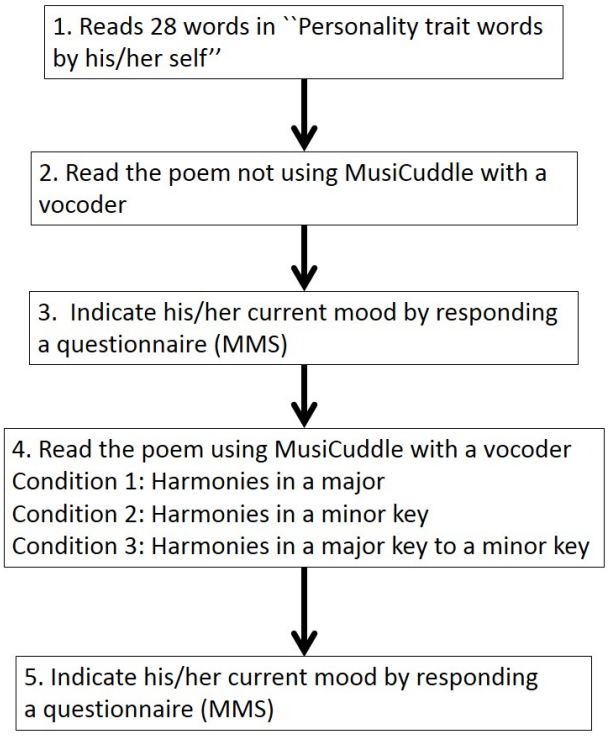


Fig. 11. Experimental method.

voice. Then, the musical phrase, the first note of which is the same as the F0 of the subject's utterance was presented. Since MusiCuddle selects an MIDI file of which the top note of the first chord is similar to the subject's F0, we transposed two phrases (Figs. 9 and 10) into other keys each, in which the top notes of the first chord are C2∼C5 before the experiment. However, the male subjects' voices were quite low, making it hard to hear the selected musical phrase (harmony). Therefore, in the experiment, MusiCuddle selected an MIDI file of which the top note of the first chord was similar to the subject's F0, but one octave higher. Each subject read the poem, hearing his/her voice combined with the musical phrase according to the vocoder function.

We prepared three musical phrase conditions: (1) harmonies in a major key, (2) harmonies in a minor key, and (3) harmonies in a minor key in the early part of the poem and in a major key in the latter part of the poem. In condition (3), the experimenter pushed the trigger button again halfway through the poem. The 12 subjects were assigned to one of the three conditions (four subjects per condition). After reading the poem, each subject completed the questionnaire again.

3) Result of the main experiment: The 12 subjects indicated their current mood by responding to the 40 items on the four-point scale twice. The first time, all subjects read the poem in the same condition, without MusiCuddle. We examined the null hypothesis "the medians of all conditions are equal" in the answers for each of the 40 questions using the Kruskal-Wallis one-way analysis. In the results, one of the items, "lack confidence" showed a significant difference of $p = 0.08$, although the others were $p > 0.10$. There was no evidence of differences in the remaining 39 items. Therefore, after this, we omitted "lack confidence" from the items for analysis.

In their second reading of the poem, the 12 subjects were assigned to one of the three conditions described previously. We conducted the Kruskal-Wallis one-way analysis of subjects' subsequent responses. Moreover, we calculated the differences between subjects' first and second responses to each of the 39 items. Then, we examined the Kruskal-Wallis one-way analysis for the differences in the 39 items.

The left side of Table IV shows the p values for subjects' second answers. In four of the 39 items (cheerful, well, slow-going ($p < 0.05$), and lively ($p = 0.06$)), the null hypothesis was rejected. Therefore, we performed multiple comparison analyses (Wilcoxon signed-rank test) for these items. The third, fourth, and fifth rows from the left of Table IV show the results. We can see that there was only a significant difference between the major and minor conditions for "cheerful" ($p = 0.03$).

Concerning the differences between subjects' first and second answers, we performed the Kruskal-Wallis one-way analysis and multiple comparison analyses. The right side of Table IV shows these results. In seven items, the null hypothesis was rejected. As a result of multiple comparison analyses, significant differences were observed between the major and minor conditions for "cheerful" ($p = 0.03$) and "negative" ($p = 0.06$).

Moreover, for the item "cheerful," there was a significant difference ($p = 0.03$) between four subjects' first and second answers in the major condition. Namely, we can say that the evaluations of "cheerful" for the major harmonies contributed to the result of multiple comparison analysis.

4) Discussion: We conducted an experiment in which subjects read a poem with/without using MusiCuddle with a vocoder function. When using the MusiCuddle, each subject could hear his/her voice, which was modified by harmonies in a major or minor key while reading. We examined the differences among the three conditions. The result showed that subjects' mood after reading the poem differed according to the condition. Moreover, the results of multiple comparison analyses showed that there were significant differences between subjects' cheerful mood for major harmonies and minor harmonies. In particular, it was clear that harmonies in a major key resulted in a more cheerful mood.

As another analysis method, we calculated the differences between subjects' first and second answers. Then, we examined the differences among the three conditions as well as the multiple comparison analyses. We found significant differences between major harmonies and minor harmonies for the "cheerful" ($p = 0.03$) and "negative" ($p = 0.06$) moods.

The iso-principle [4] shows that music's mood or the tempo must initially match patients' mood or tempo. If a patient is gloomy, then gloomy and/or sad music should initially be presented. However, the subjects of our experiment were induced cheerful mood by the major harmonies. The major harmonies were significantly "not melancholy (Ave.was 2.48),"

TABLE IV. CONTRIBUTION OF THE MUSIC IN THREE CONDITIONS TO MOOD INDUCTION.

items that have significant deferences	p	multiple comparison: second time		
		minor and major	minor and minor / major	major and minor / major
cheerful	0.02	0.03	0.43	0.17
well	0.05	0.11	1.00	0.11
lively	0.06	0.14	1.00	0.14
slowgoing	0.03	0.08	1.00	0.08
	p	differences between the first and the second time		
		minor and major	minor and minor / major	major and minor / major
cheerful	0.02	0.03	0.29	0.17
well	0.05	0.11	1.00	0.11
fresh	0.03	0.09	0.43	0.14
good mood	0.05	0.09	0.29	0.37
negative	0.03	0.06	0.09	0.40
worried	0.04	0.09	1.00	0.11
tired	0.06	0.11	0.40	0.09

The row of "p" shows the results of the Kruskal-Wallis test.

"not miserable (Ave. was 2.54)," and "not gloomy (Ave. was 2.62)" compared to the minor harmonies (see Table III).

On the other hand, traditionally, emotion was believed to stem from a physical reaction (The James-Lange theory). Then, Schachter and Singer [21] showed that emotional states may be considered a function of a state of physiological arousal and of a cognition appropriate to this state of arousal and a recognition of the factor of the emotion (Two-factor theory). These theories also support our tentative theory that a person who is gloomy can become cheerful when his/her voice is combined with cheerful sounds. Therefore, it is expected that music with the vocoder function calms the symptoms of patients with mental instability who repeat stereotypical utterances.

E. Case study using MusiCuddle with a vocoder

We performed a case study to investigate how a patient's behaviors changed with MusiCuddle using a vocoder. We expected the patient's repetitive utterances to change or stop as a result of the sound coming from MusiCuddle with the vocoder. This case study was approved by the Research Ethics Board of Saga University.

1) Method: The participant was an 81-year-old, hospitalized patient with frontotemporal dementia (FTD). She was hospitalized for depression six years ago and was discharged from the hospital. Next, she was hospitalized with a broken hip. She began to shout sometimes. Then, she moved to a dementia ward in the same hospital. Currently, she repeats stereotypical utterances for many minutes. However, she can often communicate with her care staff.

One of the authors (the MusiCuddle operator) stood by from ten o'clock a.m. to noon and again from one o'clock p.m. to half-past two p.m. In the first part of the case study, the operator played a normal electronic piano near the participant to examine the participant's interest in music.

Six months later, the second part of the case study was performed. We set two time periods, with/without the use of MusiCuddle with the vocoder. During the time MusiCuddle with the vocoder was used, the operator started MusiCuddle and presented the harmonies in either a minor key or a major key (see Section III-D). When the participant began to repeat stereotypical utterances, the operator gave triggers to MusiCuddle arbitrarily to present the harmonies.

The participant's utterances were recorded to examine the changes in her utterances. Two small wireless microphones were used to input her utterances to MusiCuddle and to the vocoder. These microphones were set behind her wheelchair. A small speaker was placed on the table near the participant.

2) Result: The participant tended to start repeating utterances only 15 minutes after going to the bathroom. She asked her care staff to take her to the bathroom, although she did not have to go to the bathroom. The following is an example of her typical utterances. "P" means "Participant."

P: Ne ne-chan ne ne-chan ne ne-chan (toots)

When the first case study was performed, other patients on the same floor were enjoying a karaoke session. When another patient sang songs, the participant temporarily transitioned from repeating utterances to singing the songs together with the other patient. After the karaoke session, the operator played the melodies of songs in which the participant was interested. Then, the participant began to sing another song making up her own lyrics.

In the second part of the case study, when the operator sang her favorite songs in front of the participant, she directed the operator to stop singing. The operator set two time periods, with/without MusiCuddle with the vocoder. The participant had to hear her utterances combined with harmonies in a minor or major key by MusiCuddle with the vocoder from a speaker. She did not push the speaker aside. So, she did not seem to hate the sound. However, there were no differences in the participant's utterances for the two time periods.

3) Discussion: In this case study, contrary to our expectations, the participant's utterances did not change. There are several possible explanations for the results.

1) It was necessary to use a refined speaker to ensure that the participant could hear the sound; because there were some patients with dementia on the floor, it was sometimes noisy. Since the subjects of the experiment (Section III-D) used headphones, they could hear the sound well. On the other hand, it is difficult for patients with mental instability to put headphones on. In the future, we should entertain the use of a directional loudspeaker.
2) If an FTD patient is able to hear the sound, can he/she recognize his/her voice in the sound? Is it

really necessary to recognize it? Even if he/she cannot recognize it, it may be enough to change his/her mood by the sound with the vocoder as long as the sound can shift his/her interest to music.

3) It is necessary to consider intended person for MusiCuddle with the vocoder. The participant in this case study was a patient with FTD. Generally, the following abilities in FTD patients are preserved: memory, perception, praxis, and spatial skills.
The participant repeated stereotypical utterances, telling staff members she wanted to go to the bathroom. The operator (one of the authors) presented the sound from MusiCuddle with the vocoder instead of responding to her request. However, she must have been unpleasant. She was assured she was asking her care staff to take her to the bathroom because she preserved some abilities. When an FTD patient requires something specific, it may not be appropriate to change his/her mood using a sound.

IV. Conclusion

In this paper, first, we introduced a system called "MusiCuddle" for patients with mental instability who repeat stereotypical utterances. MusiCuddle is a system that presents a short musical phrase when an operator pushes a button on the system's interface. The first note of the phrase is the same as the fundamental pitch (F0) of a patient's utterances.

We conducted a case study of a patient who repeated stereotypical utterances for many hours each day. The participant tended to stutter when each phrase was presented from MusiCuddle. The results suggest that FTD patients might attend to phrases according to their similarity in pitch, and their attention may be deflected away from their repetitive stereotypical utterances if the melody is too strikingly different from their utterances.

Then, we added a vocoder function to MusiCuddle so that patients would be able to attend to the music more. The vocoder allows a patient's utterances to combine with the phrase from MusiCuddle in real time. We examined whether the mood induced in subjects using the vocoder differed according to the music coming from MusiCuddle. Each subject read a gloomy poem, hearing his/her voice combined with the musical phrase according to the vocoder function. There are three conditions of the musical phrases: (1) harmonies in a major key, (2) harmonies in a minor key, and (3) harmonies in a minor key in the early part of the poem and in a major key in the latter part of the poem. The 12 subjects indicated their current mood by responding to the 40 items.

The results showed that subjects' mood after reading the poem differed according to the condition. We found significant differences between major harmonies and minor harmonies for the "cheerful" and "negative" moods. Namely, when a person's voice is combined with cheerful sounds, he/she can become cheerful. However, in the second case study, the participant's utterances did not change. It may not be appropriate to change his/her mood using a particular sound when an FTD patient requires something specific.

In the future, we will conduct experiments on patients with Alzheimer's disease who do not require something specific matter, but utter in the form of a monologue.

References

[1] C. Oshima, N. Itou, K. Nishimoto, K. Yasuda, N. Hosoi, H. Yamashita, K. Nakayama, and E. Horikawa, A Case Study of a Practical Use of "MusiCuddle" that is a Music Therapy System for Patients with Dementia who Repeat Stereotypical Utterances, *Proc. of Global Health 2012*, IARIA, pp. 14–20, 2012.

[2] C. Oshima, N. Itou, K. Nishimoto, K. Yasuda, N. Hosoi, H. Yamashita, K. Nakayama, and E. Horikawa, A Music Therapy System for Patients with Dementia who Repeat Stereotypical Utterances, *Journal of Information Processing*, Vol. 21, No. 2, pp. 283–294, 2013.

[3] N. Itou and K. Nishimoto, A Voice-to-MIDI System for Singing Melodies with Lyrics. *In: Proc. of the int. conf. on ACE'07*, pp. 183–189, 2007.

[4] I. M. Altshuler, The past, present and future of musical therapy, Podolsky, E. (Eds.). *Music therapy*, Philosophical Library, pp. 24–35, 1954.

[5] P. Nordoff and C. Robbins, *Creative Music Therapy*, the John Day Company, 1977.

[6] D. Grocke and T. Wigram, *Receptive Methods in Music Therapy: Techniques and Clinical Applications for Music Therapy Clinicians*, Educators and Students Jessica Kingsley Publishers, 2007.

[7] W. Aube, I. Peretz, and J.L. Armony, The effects of emotion on memory for music and vocalisations, *Memory*, 2013.

[8] T. Taniguchi, Music and Affection, Kitaooji Syobo Press, 1998 (in Japanese).

[9] T. Taniguchi, Construction of an Affective Value Scale of Music and Examination of Relations between the Scale and a Multiple Mood Scale, *The Japanese journal of psychology*, Vol. 65, No. 6, pp. 463–470, 1995.

[10] M. Terasaki, Y. Kishimoto, and A. Koga, Construction of a multiple mood scale, *The Japanese journal of psychology*, Vol. 62, pp. 350–356, 1992 (in Japanese).

[11] P. N. Juslin and J. A. Sloboda (Eds.), Music and Emotion: Theory and Research, Oxford University Press, USA, 2001.

[12] M. F. Folstein, S. E. Folstein, and P. R. McHugh, Mini-Mental State: A practical method for grading the cognitive state of patients for the clinician, *Journal of Psychiatric Research*, Vol. 12, pp. 189–198, 1975.

[13] S. J. Sherwood, J. Morris, V. Mor, and C. Gutkin, Compendium of measures for describing and assessing long-term care populations, Boston, MA: Hebrew Rehabilitation Center for the aged, 1977.

[14] Y. Imai and K. Hasegawa, The Revised Hasegawa's Dementia Scale (HDS-R) –Evaluation of its Usefulness as a Screening Test for Dementia. Hong Kong J Psychiatr. Vol. 4, No. 2, pp. 20–24, 1994.

[15] T. Itoh and M. Iwanaga, The effect of the relation between mood and music type on positive emotions, *The Journal of Japanese Music Therapy Association*, Vol. 1, No. 2, pp. 167–173, 2001 (in Japanese).

[16] T. Takeuchi, The influence of presentation sequences of pieces of music on depressed mood reduction –An experimental study with a musical mood induction procedure, *The Journal of Japanese Music Therapy Association*, Vol. 4, No. 1, pp. 76–86, 2004 (in Japanese).

[17] K. Hevner, Experimental studies of the elements of expression in music, *American Journal Psychology*, Vol. 48, pp. 246–248, 1936.

[18] K. Hevner, The affective character of the major and minor modes in music, *American Journal of Psychology*, Vol. 47, No. 1, pp.103–118, 1935.

[19] L. Wedin, Multidimensional study of perceptual-emotional qualities in music, *Scandinavian Journal of Psychology*, Vol. 13, pp. 241–57, 1972.

[20] T. Aoki, A psycho-lexical study of personality trait words: Selection, classification and Desirability ratings of 455 words, *The Japanese journal of psychology*, Vol. 42, No. 1, pp. 1–13, 1971 (in Japanese).

[21] S. Schachter and J. E. Singer, Cognitive, social and physiological determinants of emotional state, *Psychological Review*, Vol. 69, No. 5, pp. 379–99, 1962.

Improving Online Interactive Modules: An Iterative Design Model

Vanessa Slinger-Friedman and Lynn M. Patterson
Department of Geography & Anthropology
Kennesaw State University
Kennesaw, United States of America
vslinger@kennesaw.edu and lpatters@kennesaw.edu

Abstract— **With the promotion of online and interactive learning, instructors may be tempted to include arbitrarily the latest technologies into their modules without considering the ramifications on students and their learning processes. Instead, we propose that online and interactive modules should be carefully designed using multiple levels of review to ensure clarity of instruction, ease of use of interactive components, and engagement of the students with the material and any related activities. This paper documents the process to design and improve an online module using an iterative process involving instructional designers, student pilot tests, and a focus group. Feedback from these constituents enabled instructors to optimize instructional design to maximize learning opportunities and achievement in online environments.**

Keywords—iterative design process; student focus group; interactive online modules; cognitive learning; human geography

I. INTRODUCTION

Relatively recent changes in the focus of curriculum preparation from instructor-centered to learner-centered has put more attention on the deliberate design of courses and the development of content and assignments. In online courses, new technologies increasingly are being introduced to provide students with content and provide opportunities for students to interact with that content. However, designing new learning environments is challenging, and the center of attention tends to be on the delivery of new content through the technology rather than assessment of the impact of these new learning technologies on student learning [1]. Additionally, assessment of courses is often inadequate, poorly timed, and limited in the effect it has on course or module modification and improvement. While considered important, feedback on module design before it is delivered to students within the structure of a formal class is uncommon [2].

Using a cognitive theory framework, an online interactive learning module was developed for an introductory human geography course and as part of the development of a completely online textbook. The cognitive theory framework supports a multimedia design of educational materials [3][4][5] through which students can engage in meaningful learning when they actively process material through "selecting relevant words and pictures, organizing them into coherent pictorial and verbal models, and integrating them with each other and appropriate prior knowledge" [5]. This module is one of several developed that includes imagery, custom videos, readings, discussions, animations, interactive exercises, and assessments. In the module the integration of theory and applications takes place through activities in which theories and ideas are applied for use in practical situations to answer real-world geographic questions, bringing the course material "alive" for students.

The purpose of this paper is to document and analyze the iterative design process for the development and improvement of this online interactive human geography module. This iterative design process involved the input and feedback from Instructional designers (IDs), student pilot testers, and a student focus group. In the design process and improvement of the module, we obtained feedback in the three areas of (1) engagement, (2) clarity and ease of use of module elements, and (3) the ability of the module elements to assist in meeting learning outcomes.

This paper begins with the literature associated with a learner-centered approach utilizing a multimedia design, and the use of assessment and feedback to design and then improve modules that have a learner-centered focus. Specific examples of courses that have used an iterative design process or involved feedback from instructional designers and students in the course design process are included. Then, the design of the interactive module, based on a process of student-centered learning, created for the introductory human geography course is detailed. Next, we outline the use of an iterative design process using IDs, student pilot testers, and a student focus group that provided the basis for module redesign. We present the results and lessons learned from this iterative process. Finally, we conclude with the broader implications of this research on optimizing instructional design to maximize learning opportunities.

II. LITERATURE

Traditional curriculum preparation has conventionally focused on the instructor rather than the learner. That is, the instructor prepared the material for delivery and expected students to absorb the material through lectures, readings

and written exams. In recent years a paradigm shift has moved the emphasis from teaching to learning and to a student-centered curriculum with a greater emphasis on meeting learning outcomes and the ability for students to demonstrate skills and competencies within courses and modules. In the world of online courses this means that contemporary online learning development is moving away from courses with "pages of electronic text, to more deliberately planned learning designs, learning tasks, and processes structured in deliberate ways" [6].

Correspondingly, there is more of a focus on how learners learn and the design of effective learning environments based on best practices. The learner-centered model calls for active student participation and the use of multidimensional products to develop deeper understanding by students [7]. As technology has developed and become a more integral part of the distance learning environment, it has impacted the delivery of content, learning tasks, and assignments [8]. The ways by which information is presented and also the way in which students interact with that material are important. Furthermore, the medium employed can motivate and engage students as active and collaborative learners rather than just providing information to them. Multimedia instruction rather than "flat resources," such as static text documents, have been identified as an important element of high-level interactive engagement and student satisfaction [9][10].

The design of the online interactive module for this study is predicated on Mayer's research on cognitive theory-based assumptions regarding the way that people learn from words, pictures, and active processing of material (what Mayer considers the two elements of the "Dual Channel Assumption and the Active processing assumption) in computer-based multimedia presentations results in deeper understanding in learners [5]. This concept of knowledge transmission is based on a constructivist point of view where knowledge is constructed by the learner through activity [8]. This construction has led to the development of "new learning environments" or what Martens et al. [8] call "constructivist e-learning environments" in which activities are created to challenge students and provide them with realistic contexts so that students become intrinsically motivated to explore and control their own learning process.

Despite the shift in the teaching/learning paradigm and the rise of assurance of learning outcome assessments, there is a lack of available texts and other material to guide instructors involved in module design/redesign, and a lack of attention given to and results in terms of course improvement derived from the type of student feedback currently elicited [11][12]. Course-based student assessment, now commonly found at most accredited higher education institutions, is created to encourage instructors to examine their roles as course creators and to articulate their goals and objectives. The process of collecting student feedback through formal assessment of individual teachers and courses is widespread [12]. While there are other purposes for formal assessments, a significant function of this feedback is to provide instructors with information about their teaching, with the intention that they will use this feedback to improve their courses and enhance the effectiveness of their teaching. Most commonly, these student evaluation surveys take the form of automatically scanned standard questionnaires with questions like, 'The instructor is knowledgeable about the course material' and 'The instructor inspires interest in the course and course content', using a five-point scale ranging from 'strongly agree' to 'strongly disagree'. Customarily, course-based student assessment is done at the end of a semester, leaving no time for modification to be made to a module or course for that cohort of students or even the incoming cohort of students. Furthermore, research has shown that the test-retest reliability of students' evaluations is high, indicating that the performance of the teachers is not improving with experience, perhaps as a result of teachers and institutions not taking the student feedback sufficiently seriously [12]. While student evaluations and ratings of an instructor and course might be reliable, they do not in themselves lead to any improvement in the quality of teaching and the effectiveness of course content and course design [13]. Formative assessment is a well-recognized element of basic Web site design and publication, yet it has received limited attention in the literature on online module development [14].

In line with the cognitive theory for learning, course or module evaluation based on criteria that is co-operatively developed and focused on obtaining information about the quality and effectiveness of the module should be most constructive in course redesign. Additionally, some scholars believe evaluation should not be just a retrospective process, but it should be an integral part of module development, informing instructors before, during, and after the process, e.g., [2].

Designing and improving new learning environments is challenging. Much of the available research shows an emphasis on delivery of these new learning environments rather than on analysis, evaluation and process of development, e.g., [15]. Module authors need to consider design elements including perceived applicability, content organization and interaction, ease of use, and the potential for learner engagement [14].

Designers of new online learning environments rarely gain knowledge of how students will perceive the module and associated tasks before they are delivered to the students. Greenberg [16] asserts that quality assessments should be taking place during the design of the course and include the course creators. Some rare examples of this can be seen in Kingston et al. [17] and Lederman [18]. Kingston et al. [17] utilized mobile technologies and virtual fieldtrips to teach physical geography. Students who had taken the old module and completed the new module were given questionnaires and then participated in a focus group to investigate the effectiveness of the new technologies.

Improvements were made to the material based on student feedback. Lederman [18] also suggests that focus groups can be very useful for pre-testing educational materials as they "provide an opportunity for extensive commentary, unrestrained by the limits of a survey questionnaire or the student-teacher relationship which may affect course evaluations at the end of a class" ([18], page 126). Skye et al. [14] concluded that multiple methods of data collection could be used to provide information to module developers to improve modules and to address module ease of use, navigation and content. Based on the precedent set by these examples, in designing our module we used an iterative process that required design advice and inspection by IDs, student pilot tests, a student focus group, and refining of the module prior to incorporating the module into classes.

III. MODULE DESIGN

The interactive multi-media module design uses the concepts of space, place, and human-environment interaction to study the process of hydraulic fracturing (otherwise known as fracking) from a geographic perspective. The module requires approximately 30 to 45 minutes for completion. Hover texts and graphics provide the module with text explanations. Using a web-based format, the module lists the learning objectives and begins with a short reading of one to two paragraphs in length that provides an overview of the applied topic. Next, a three minute narrated animation illustrates the concept of fracking. This is followed by a short video, which discusses the geographic implications of the topic. Finally, a series of interactive exercises allows the student to explore the topic using geographic tools (e.g., visual examination, verbal descriptions, digital mapping, cognitive perceptions, and mathematical modeling). For several of the module elements described above, an interactive textbox appears to the right where the student is encouraged to take notes.

Different components in the module require different intensities of interactivity. For example the animation and videos require students to click to start or more forward. The multimedia uses animation, voiceover, and video to engage students in a different way than simply reading the content. Other activities have a higher intensity of interactivity. The clickable maps require students to be actively engaged in thinking about, manipulating, and actively participating in the learning through completing the exercise. An example of this is when students have to use a calculator embedded in the module to find the population change in Williston, North Dakota and Rifle, Colorado. Likewise, they have to click on the markers that show fracking violations in Pennsylvania and make active decisions about what information to use to contribute to the discussion on the blog tied to this activity.

The results of the interactive exercises are shared with the instructor and, in some cases, other students. The interactive exercises were developed using publically available software such as micromob, Scribblar, Google Earth and ArcGIS ™ (Figs. 1 and 2).

Figure 1. Interactive exercise using Google Earth

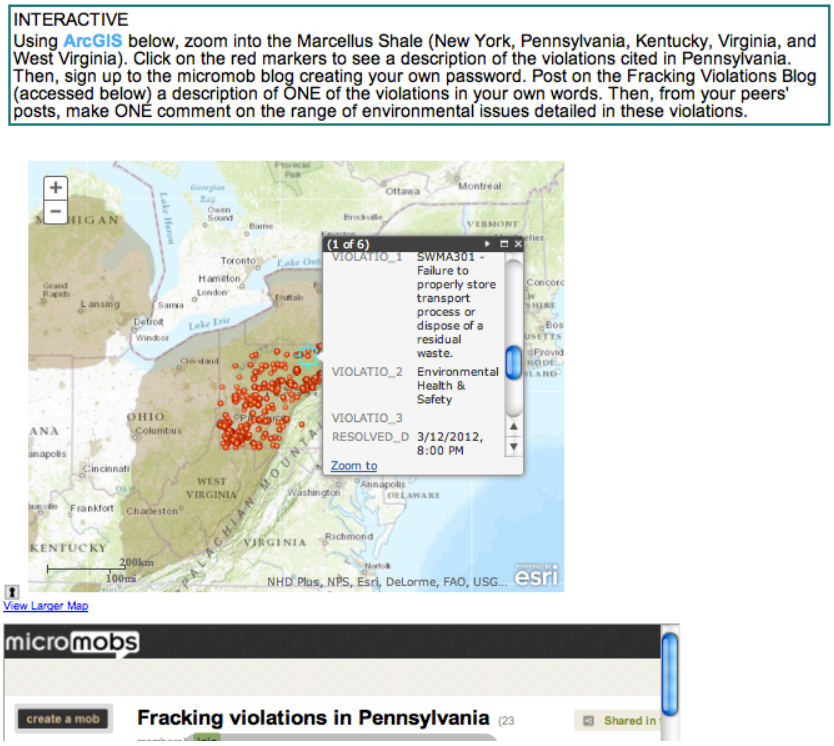


Figure 2. Interactive exercise using ARCGIS

IV. METHODOLOGY FOR ITERATIVE DESIGN DEVELOPMENT AND IMPROVEMENT OF AN ONLINE MODULE

In an iterative design process to develop and improve an online interactive human geography module, we engaged instructional designers, student pilot testers, and a student focus group. We focused on three areas for module assessment and improvement: (1) engagement, (2) clarity and ease of use of module elements, and (3) the ability of the module elements to assist in meeting learning outcomes. We also asked our IDs, student pilot testers, and focus group students to suggest general improvements to the module.

The online interactive module was created and evaluated in four stages (Fig. 3). At the beginning of the module's design, the instructors met with IDs from the University to gain knowledge about software (with a focus on free software) and design. After the module was designed the instructors met once again with the IDs to test for functionality of the module. Next, the module was pilot tested with three student volunteers. These students provided feedback on the module design and functionality. Finally, a group of 17 students participated in a large-scale pilot. As a group and simultaneously, each student completed the module in advance of the focus group interview.

The focus group students provided feedback on their engagement with the module, the clarity and ease of use of the different module elements, and how they thought the module elements assisted them in achieving the learning outcomes stated at the outset of the module (Fig. 4). Both members of the research team were present – one to serve as moderator and the other as a note taker who recorded speakers, comments and significant non-verbal behavior [19].

To address a potential a repercussion of focus groups, we had two mechanisms to minimize groupthink [20]. First, students each filled out a short questionnaire at the completion of the module. The questionnaire allowed us to obtain individual feedback that may not have come out in the group discussion but that might have been vital to improving the e-learning modules. Second, we asked the focus group members to jot down notes during the group interview. In the event students did not get a chance to share their comments, these notes were collected at the end.

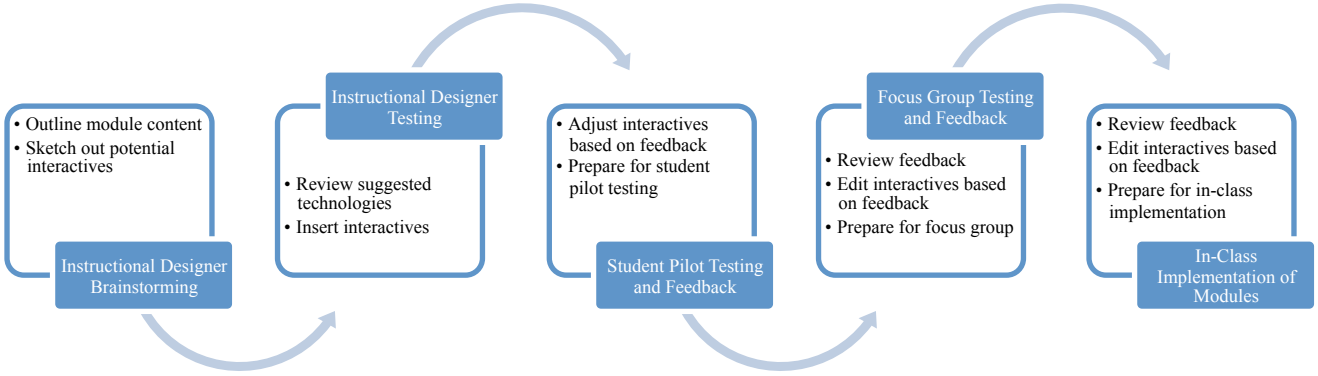


Figure 3. The Four Stages of Module Creation and Evaluation

Introductions

Facilitator introduces members of the research team and each of the group members introduce themselves. The facilitator provides the background and ground rules (confidential and anonymous reporting, honest opinions, etc.). The facilitator will inform the group that we would like to collect notes made by the participants during the session to ensure we collected as much feedback as possible, if the participants are willing.

Issues and Discussion Questions (Semi-structured)

Overall Impressions

- *Please share with us overall how you felt about the modules?*
- *What did you like about the modules? What didn't you like about the modules?*

Engagement

- *What about the material (videos, photos, readings) did you find the most engaging?*
- *How did the interactive exercises affect your interest in the content?*
- *Did any of the material or exercises make you want to learn more about the topic? If so, which and how?*

Clarity and Ease of Use of Elements

- *What concepts or parts of the module were the most clear? The least clear?*
- *What aspect of the interactive exercises did you find the clearest/easiest? What aspects were unclear/more difficult?*

Learning

- *Overall, how useful did you find the exercises?*
- *How did the interactive exercises assist you in understanding course content? In applying course content?*
- *How did the interactive exercises challenge you?*

Improvements

- *What improvements could we make to improve the elements of the modules?*

Summary of what we have heard

- *Have we missed anything?*

Collect notes (to review later).

Figure 4. Focus Group Questions

We reviewed the results of the surveys and interviews, coded the data and created categories to allow trends to emerge and to be able to develop summary statements that capture the essence of the responses [18][21]. The results of the coding offer two outcomes. First, the student responses helped us identify which of the interactive exercises had greater perceived value to students and which of the module elements need to be discarded or modified due to their inability to engage students and help them meet the stated learning outcomes for the module. Second, areas of confusion and lack of clarity were identified. The modules were then revised to address weaknesses.

V. Feedback and Discussion

Given the quality and content of the feedback from the three audiences, we concur with Greenberg [16] that assessment provided during the design of the course that include course creators can result in a quality outcome. Specifically, the feedback obtained for all design elements (content, clarity, ease of use, and engagement) were vital to ensuring the desired learning outcome could be achieved using these interactive exercises [14].

The iterative process allowed us to obtain feedback from different audiences and revise the module accordingly before we finalized it for in-class implementation. In the first three phases of feedback, we used the responses to modify the interactive exercises for clarity and ease of use. In the focus group stage, we not only focused on clarity and ease of use but also on engagement and ability to meet learning outcomes. In addition to these three areas, the focus group also provided feedback through general likes/dislikes, technical comments, and recommendations for the module. The students even recommended additional content they would like to see included.

The IDs' feedback from the pre-tests was mostly technical in nature and was intended to improve the student experience. For example:

- You might want to include an example of the word popper and explicitly tell students when they see blue text in following pages to hover over the word for additional information.
- You may want to add a comment above the animation, telling students to let each slide load before they try to watch it so there are no buffering issues.
- You may want to direct students to click "finish" after they take their notes so this is saved and ready to be submitted to you.
- I had issues getting the Google Earth plug in to download on my computer and ended up having to manually install it and restart the browser. You may want to take a second on the introduction page to tell students to download the plug in. This way they can have it installed and ready as they move through the lesson.
- You might want to expand the pixel dimensions of the calculator tool so students don't have to scroll in this box.
- I was able to comment on the blog, but once I submitted my comment I got kicked out of Softchalk to the actual blog page. I am trying to find alternative tools to help you accomplish the same thing. One option might be a tool called MicroMob (free, but students would have to register for an account).

The three student volunteers, who completed the pre-test, offered similar feedback but this time as representative users. Some comments from the pre-testers were:

- I tested the Module, but couldn't open Scribblar on my computer, but I think it was only a problem with my computer not the module itself.
- I was very impressed with the map interface, it worked well in conjunction with the activity.
- It was extremely helpful in learning about fracking and fracking sites when I was able to visualize what was talked about in the video from the lesson.
- On the cognitive map I was a little unsure what to do with the second drawing tool? So...I didn't really mess with it.
- I really like the blog section I thought it was cool but I wasn't entirely sure what to do. I think it's meant to be interactive with other students commenting on each others post, right?
- The map with the shale plays was a little difficult to read
- The layout might work better if you kind of separated the directions from the text? Or changed the lettering so it's easier to distinguish from the info?
- I really liked that you guys came at this from several different angles. I REALLY think that reinforces the information. Especially with the drawing.
- I also liked that you weren't exactly for or against fracking you gave different perspectives which allowed me (the student) to make up their own mind on where they stood. Unfortunately, I was already against fracking to begin with but had I never heard of if I'd be able to empathize with multiple sides.
- I like that I had to take notes during the YouTube video and that it gave me the option to print.
- I think it's important to have a basic understanding of the processes you're learning about so the intro video was very helpful and kind of set the foundation for the info that followed.

The focus group students provided a different set of insights, including their overall impression of the module, including likes and dislikes, their perceptions on clarity and ease of use, engagement with the material and their general perceptions of how the interactive components helped them to meet the learning objectives.

In general, the students comments related to likes and dislikes were distinctly separate. The students liked the interactivity and the way the interactive components made them “think about the material” and “made it clearer.” The dislikes mostly centered on technical issues such as wanting to break up the text and having to sign up for so many things (from free software).

Per the survey, the students indicated they would like to have more interactive exercises in their courses with 88% of students in agreement. The responses from both the survey and the interviews provided additional insight as to why the students would be interested:

- Because learning with visual and hand on exercise is much better than just reading a textbook.
- It helped get info across in a way a lecture might not- the added interactivity reinforced it by requiring me to put forth the lesson I had just learned.
- More and more learning is occurring online, but there is only so much one can take from online classes. Interactive courses like this may increase the level of participation.
- …Interactive material like this is great, especially for learning complex material in a brief snapshot, which this achieves remarkably.
- It takes learning to another level…drawing a map uses a different part of your brain and it forces you to kind of think critically about the subject matter.

From the survey, students overwhelmingly believed that the interactive exercises helped them to learn more (Fig. 5). They also indicated that the interactive activities helped them to describe the process of fracking (Fig. 6) and the activities helped them to use the geographic tools to explore fracking from a geographic perspective (Fig. 7). They were less certain, but still positive that the interactive exercises helped them to apply the geographic perspectives to fracking (Fig. 8).

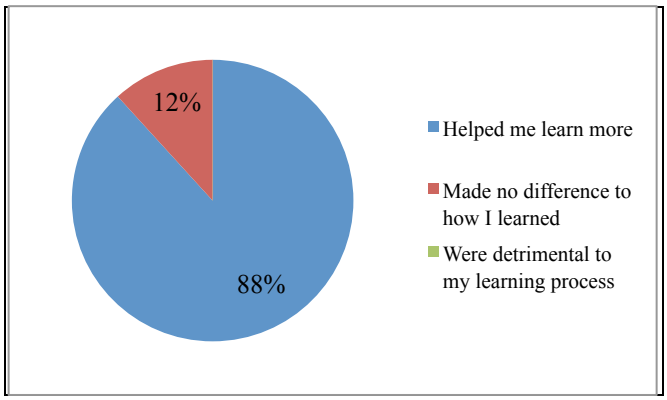


Figure 5. Overall, the interactive exercises _____.

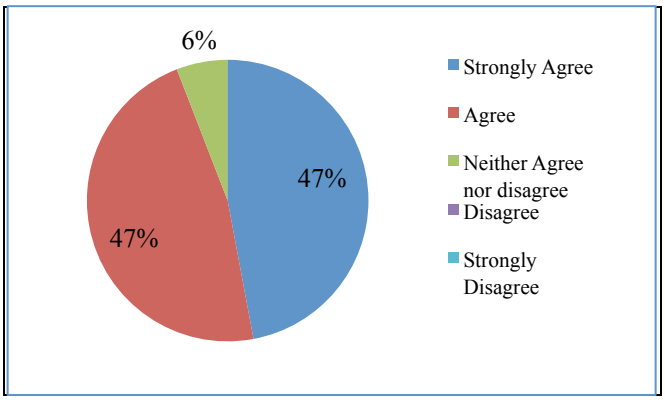


Figure 6. The interactive exercises helped me to describe the process of fracking.

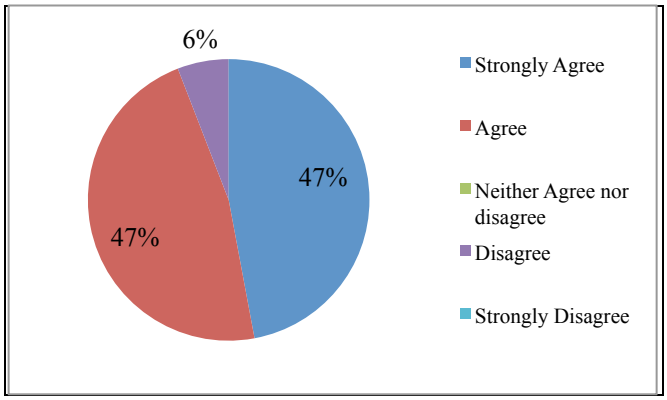


Figure 7. The interactive exercises helped me to use the geographic tools (visual verbal, cognitive, mathematical, digital) in exploring fracking from a geographic perspective).

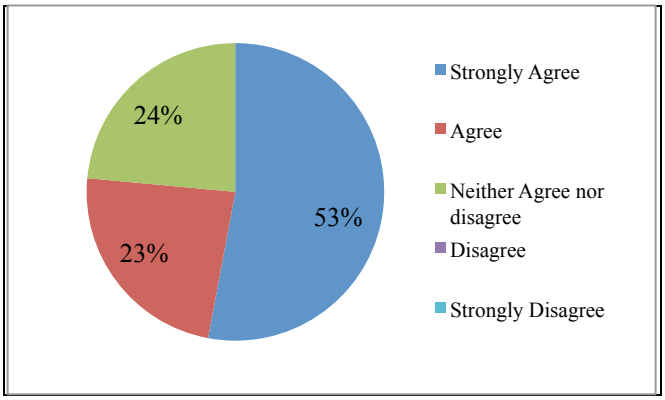


Figure 8. The interactive exercises helped me to apply the geographic perspectives (space, place, human-environment interaction) to fracking.

The feedback from the focus group interview was similar. When asked "if the material had just been presented as text, would you have learned as much?", the response was a resounding 'no!'. Opinions on the level of interactivity and how it helped students meet learning objectives was mixed. A few students responded that the lower intensity interactive really (video and animation) were instrumental in helping them learn the material.

Additional comments regarding the ability of the interactive exercises to assist students to meet learning outcomes included:

- Use of maps and aerial photography helped to contextualize this information. I can see where fracking occurs and what impacts on environment and population it has.
- First it [the module and its interactive activities] explained what space and place meant, then gave an example of each. Then the human-environment interaction.

The students rated the clarity of the interactive exercises in categories of Clear, Somewhat Clear and Unclear (Fig. 9) with the visual and verbal components offering the most clarity. For the mathematical and cognitive components, less than 50% of respondents were confident in how to utilize the interactive activities. In reviewing the comments associated with these activities, we realized that technical issues and perception caused a problem for the students. For instance, there was no mechanism to store the answers for the population growth calculation so when students had to answer questions, they had forgotten the specific numbers. Additionally, the calculator did not appear in one of the Internet browsers on campus, which limited the students' ability to fully utilize the module. While Scribblar™, the drawing software, allowed students to be working simultaneously on their cognitive maps, it did not allow for real-time edits. Students also requested the ability to return to view and discuss each other's maps. This feedback helped us to hone in on the elements in the module that had the least clear instruction and/or software to use.

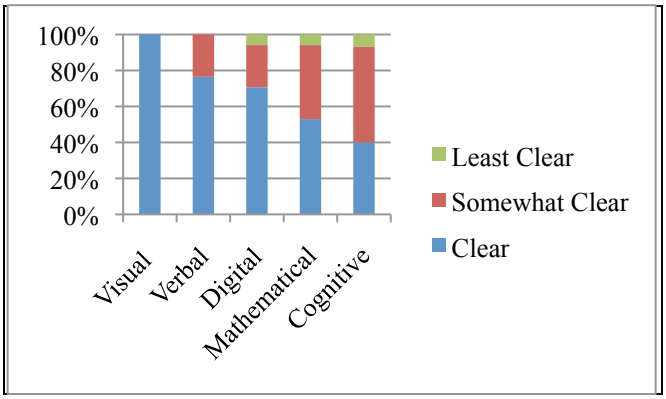


Figure 9. Student rating in terms of Clarity

The students also rated the interactive activities in terms of ease of use (Fig. 10). The mathematical and cognitive maps exercises rated more difficult than the others.

> The scribbler page was time consuming and a bit frustrating, but not bad. For the math exercise, it would have been much easier had the chart and the calculator been on the same section of the page so that the user doesn't have to scroll up and down while remembering the numbers.

The response to the math difficulty appeared to be technical while the cognitive interactive exercise may have been a combination of both technical frustration and conceptual reach as it is often a difficult concept to grasp. Furthermore, students had the most trouble with the free software program used in this section. We recognize this is the case and have modified that section of the module to improve clarity in both instruction and ease of use by changing the software used to capture the cognitive maps drawn by the students. We also provided a textbox for students to type in the results of their calculations for easy reference when they are answering the follow-up questions.

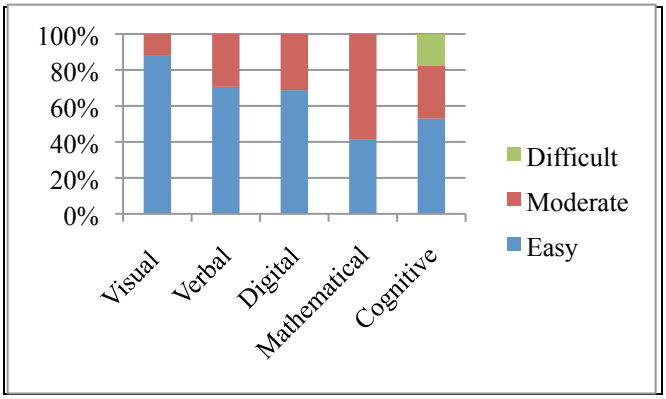


Figure 10. Students' rating of interactive activities in terms of Ease of Use.

The students' reactions to the multi-media interactive components and how it inspired creativity, was consistent with the notion that students have a higher satisfaction and engagement rate using these technologies [9, 10]. The students repeatedly referenced the animation, the news video and the cognitive map exercise. The animation and video seemed to be most helpful in describing the process and providing additional background, while the cognitive map drawing exercise was seen as helpful because it inspired creativity. The students also found value in the quizzes and calculator based exercise.

Regarding the Google Earth™ interactive, the students discussed how they explored and looked at other areas outside of the ones that they were directed to study to see how fracking was evident on the landscape. For the GIS-based exercise, the students also commented on some other data that could be incorporated to give additional insight in the impact of fracking to the region. These types of comments indicate that the interactive module did create a constructivist e-learning environment where students were motivated to explore the subject matter further on their own [8].

One comment we found particularly insightful was the recognition that the various components and levels of intensity of the interactives reinforced learning by touching on various learning styles.

> I think it hit on those components more so than you could in a normal classroom. My whole life teachers have told us there are various learning styles. Some people are audio, others are visual and some are Kinesthetic learners. This module kind of hit on all three of those so just in case you cannot keep your focus to listen to a seven minute video, you have a map to draw to reinforce the concepts you may have not totally gotten from just listening.

Further evidence of the students' engagement was how they provided us with multiple suggestions on additions to the material and interactive content. The students were engaged in the material to the degree that they began thinking of other ways to present and expand on the content covered in the module.

Finally, in both the focus group survey and the discussion, students provided excellent technical recommendations to improve the modules. Notation of these specific items were made and we have edited the module to integrate these suggestions to the best of our ability.

VI. LIMITATIONS AND RECOMMENDATIONS

We recognize that the development of the module and the focus group testing had limitations. First, we were dependent upon the use of free software in developing the interactive modules. This was cumbersome for the students and the instructors as we had to test out a variety of software packages to find the "right" match. Since the modules in the textbook will have many of the same interactive activities, once we have resolved the software issue and/or have

software developed specifically for this purpose, this limitation should be resolved.

Second, the students were asked to complete the module as part of a combined and timed class session. Because of this, student feedback to each other in the cognitive mapping interactive and the ArcGIS™ blogging exercise was limited and the full scope of the interactivity between students was virtually non-existent. We recommend spacing the focus group participation over a series of days so that the interactivity can be fully explored.

VII. CONCLUSION

The iterative process of gathering formative assessment feedback was essential to the improvement of the module. We suspect that often instructors assume that as long as the material is presented, the students should be able to navigate and succeed given the intent. A severe limitation of online content is that instructors do not gain immediate feedback from students unless there is a problem. And then, it is often too late for the instructors to efficiently address the problem without slowing the pace of the class down and confusing and frustrating the students. The idea of obtaining feedback prior to implementation is not innovative or unique; instead it is the exception rather than the rule, e.g., [14][2].

Through the iterative review process, we gained knowledge through the various audiences that would not have been attained through simple implementation of module in class. The IDs provided multiple levels of feedback. In the initial brainstorming session, they offered ideas on technologies and software that might be useful for the interactive components. Utilizing the IDs as a resource was key as it opened up possibilities we had not considered and their expertise in this arena offered insight into what works for students. They also tested the technologies for us. From an end-user perspective, the pilot student group provided us with feedback for functionality and learning. The focus group enabled us to see how a broader sample of multiple learners experienced the modules [18].

Educational delivery models for college courses have changed. Contemporary educational delivery models include online and distance education; however, there has been a gap in the assessment of these learning technologies of their impact on student learning [22]. In the development of learning modules for students in online courses there is room for an iterative design process whereby advice from IDs and feedback from students can and should be taken into consideration.

Ongoing assessment of a course can allow faculty to systematically incorporate feedback from all involved in the teaching and learning process, adding to, replacing, correcting, and improving an ever-growing body of learning materials and best practices. In each redesign of a module or, on a large scale, course shifts can be made towards making it more active and learner-centered [23].

Educational research of this nature tackles the fundamental question of how to optimize instructional design to maximize learning opportunities and achievement in online and distance learning environments [24]. Thus, by enlisting instructional designers and students in curriculum development, we expect to improve the module content and interactive activities by directing revision based on their feedback. While the overall assessment by the IDs, pilot testers and focus group students indicated that interactive activities in the module were perceived favorably by the students and likely required only tweaking for clarity and ease of use, an ongoing consideration should be whether or not the interactive activities are helpful in delivering content and developing desired skill sets. As such, research on perceived and actual learning outcomes should be conducted.

ACKNOWLEDGMENT

We wish to thank Tommy and Beth Holder for their support through the KSU Foundation Holder Award, the Center for Teaching and Learning at Kennesaw State University for providing technical assistance, and the College of Humanities and Social Sciences Office of Distance Education for research and travel funding.

REFERENCES

[1] V. Slinger-Friedman and L. M. Patterson, "Student Learning and Student Perceptions of Learning from Interactive Modules," Proc. *Fifth International Conference on Mobile, Hybrid, and On-line Learning* (eL&mL 2013). Held 23 February – 1 March, 2013, pp. 1-5. http://www.thinkmind.org/index.php?view=article&articleid=elml_2013_1_10_50018. Last accessed 12/9/2013.

[2] R. Donnelly and M. Fitzmaurice, Designing Modules for Learning, Dublin Institute of Technology, Ireland 2005, n.p. Available from http://www.aishe.org/readings/2005-1/donnelly-fitzmaurice-Designing_Modules_for_Learning.html. Last accessed 12/9/2013.

[3] B.L. Black, H. Heatwole, and H. Meeks, Using multimedia in interactive learning objects to meet emerging academic challenges in Learning Objects: Theory, Praxis, Issues, and Trends, Koohang, A. & Harman, K. Eds. Santa Rosa, California: Informing Science Press, 2007, pp. 209 – 257.

[4] M. Maag, "The Effectiveness of an interactive multimedia learning tool on nursing students' math knowledge and self-efficacy," CIN: Computers, Informatics, Nursing, vol. 22, no. 1, January-February 2004, pp. 26-33.

[5] R.E. Mayer, "Cognitive theory and the design of multimedia instruction: An example of the two-way street between cognition and instruction," New Directions for Teaching and Learning, vol. 2002, no. 89, 2002, pp. 55-71.

[6] R. Oliver, Learning objects: supporting flexible delivery of flexible learning in *Meeting at the crossroads: Proceedings of ASCILITE 2001* G. Kennedy, M. Keppell, C. McNaught & T. Petrovic, Eds. , pp 453-460. Melbourne: The University of Melbourne.

[7] R. Frye, G.R. McKinney, and J.E. Trimble, Tools & Techniques for Course Improvement: Handbook for Course Review and Assessment of Student Learning 2007. Available at http://www.wwu.edu/depts/vpue/assessment/documents/course_handbook.pdf. Last accessed 12/9/2013.

[8] R. Martens, T. Bastiaens, and P.A. Kirschner, "New learning design in distance education: The impact on student perception and motivation," Distance Education, vol. 28, no. 1, May 2007, pp. 81-93.

[9] M. Murray, J. Pérez, D. Geist, A. Hedrick, "Student interaction with online course content: Build it and they might come," Journal of Information Technology Education: Research, vol. 11, 2012, pp. 125 – 140.

[10] D. Jensen, B. Self, D. Rhymer, J.Wood, and M. Bowe, "A rocky journey toward effective assessment of visualization modules for learning enhancement in engineering mechanics," Educational Technology & Society vol. 5, no. 3, 2002, pp. 150-162.

[11] P.T. Knight, Being a Teacher in Higher Education. Maidenhead, UK: Society for Research in Higher Education and the Open University Press, 2002.

[12] J.T.E. Richardson, "Instruments for obtaining student feedback: a review of the literature," Assessment & Evaluation in Higher Education 30(4), pp. 387-415.

[13] D. Kember, D.Y.P. Leung, and K.P. Kwan, "Does the use of student feedback questionnaires improve the overall quality of teaching?" Assessment and Evaluation in Higher Education, 27, 2002, pp. 411–425.

[14] E.P. Skye, L.A. Wimsatt, T.A. Master-Hunter, and A.B. Locke., "Developing online learning modules in a family medicine residency," Fam Med vol. 43, no. 3, 2011, pp. 185-192.

[15] J.J.G. van Merrienboer and R. Martens, "Computer-based tools for instructional design," Educational Technology, Research and Development, vol. 50, 2002, pp. 5-9.

[16] G. Greenberg, "Conceptions of quality in course design for web-supported education," Proc. of the *26th Annual Conference on Distance Teaching & Learning*. Aug. 2010, Madison, WI. Available from http://www.uwex.edu/disted/conference/Resource_

library/proceedings/28667_10.pdf. Last accessed 12/9/2013.

[17] D.G. Kingston, W.J. Eastwood, P.I. Jones, R. Johnson, S. Marshall, and D.M. Marshall, "Experiences of using mobile technologies and virtual fieldtrips in Physical Geography: implications for hydrology education," Hydrology and Earth System Sciences, vol. 10, May 2012, pp. 1281-1286.

[18] L.C. Lederman, "Assessing the educational effectiveness: The focus group interview as a technique for data collection," Communication Education, vol. 38, 1990, pp. 117-127.

[19] P.S. Kidd and M. B. Parshall, "Getting the focus and the group: Enhancing analytical rigor in focus group research," Qualitative Health Research, vol. 10, May 2000, pp. 293-308.

[20] L. Janis, Victims of Groupthink: a psychological study of foreign-policy decisions and fiascoes. Boston: Houghton Mifflin, 1972.

[21] M. A. Carey, "Comment: Concerns in the analysis of focus group Data," Qualitative Health Research, vol. 5, no. 4, Nov. 1995, pp. 487-495.

[22] J. O'Malley, "Students perceptions of distance learning, online learning and the traditional classroom," Online Journal of Distance Learning Administration, vol. 2, no. 4, Winter 1999, pp. 1-9.

[23] National Center for Academic Transformation, Five principles of successful course redesign, 2005, pp. 1 – 10. Available from http://www.thencat.org/R2R/R2R%20PDFs/SuccCrsRed.pdf. Last accessed 12/9/2013.

[24] S.D. Johnson, S.R. Aragon, N. Shaik, and N. Palma-Rivas, "Comparative analysis of learner satisfaction and learning outcomes in online and face-to-face learning environments," Journal of Interactive Learning Research, vol. 11, no. 1, 2000, pp. 29-49.

Making Medication Prognoses for Prostate Cancer Patients by the Application of Linguistic Approaches

Hang Zettervall, Elisabeth Rakus-Andersson
Department of Mathematics and Science
Blekinge Institute of Technology
Karlskrona, Sweden
e-mail: Hang.Zettervall@bth.se
e-mail: Elisabeth.Andersson@bth.se

Janusz Frey
Department of Surgery and Urology
Blekinge County Hospital
Karlskrona, Sweden
e-mail: janusz.frey@ltblekinge.se

***Abstract*—Apart from the probabilistic model and the model of 2-tuple linguistic representations, a new extension of the fuzzy set, known as the hesitant fuzzy linguistic term set can be seen as the third representative of linguistic approaches. In this paper, we focus on multi-expert decision-making problems, in which a group of physicians are independently asked for assessing the effectiveness of a set of treatment therapies. Our goal is to rank the effectiveness of treatment modalities from the most recommended to the contraindicated. Two individual prostate cancer patients have been taken into account in the practical studies. For the first patient, the probabilistic model and the model of 2-tuple linguistic representations have been adopted to accomplish the medical application. Whereas, for the second patient, the approach of hesitant fuzzy linguistic term set has been used to make the medication prognoses. Moreover, the continuous fuzzy numbers in the Left-Right representations are used to mathematically express the experts' judgments and *s*-parametric membership functions are designed to represent the fuzzy linguistic terms.**

Keywords-multi-expert decision making; fuzzy group decision making; probabilistic model; 2-tuple linguistic representations model; hesitant fuzzy linguistic term set

I. INTRODUCTION

Prostate cancer is one of the most common oncological diseases in the world. Due to the vide heterogenicity of malignant potential, the prostate cancer treatment is multifactorial. Like in any other oncological disease, the cooperation of health professionals is required to make the consensual treatment decision. One way to facilitate the treatment decision-making process can be a multidisciplinary team meeting (MDT) - an event or a platform where decision makers from various relevant treatment / diagnostics fields meet to discuss further proceeding.

At the Urology Department of Blekinge County Hospital, Karlskrona, the MDT is a forum of health care providers including medical oncologists, urologists, urology sub-specialized nurses, radiologists and pathologists. The aim of the conference is to assess and establish treatment decisions for particular patients with a spectrum of problematic urological conditions that cannot be easily solved by means of available resources. Our long term aim is also to discuss the best and available treatment modalities of all newly diagnosed cases of prostate cancer. Quite often the decision making process is very clear and straight forward, but some cases lay outside the frames of guidelines and recommendations. Obviously, the final choice of treatment is also on discretion of the patient. This modern approach has however two pitfalls. One of them is when there is a discrepancy between forum members and the other one is when the patient is not interested in the treatment modality chosen by the panel. The best solution is to obtain a method for solving discrepancies and simultaneously to find a method that shows panel's results as treatment recommendations ranged from the strongly recommended to the contraindicated. Such approach should be very helpful particularly in such diseases as prostate cancer, which has a broad spectrum of treatment methods that can be tailored to the particular patient's needs and requirements.

Therefore, in view of the physicians' requirement, we wish to extend our earlier research presented in [1] by adopting the approach of hesitant fuzzy linguistic term set (HFLTS) [2] to make the medication prognoses.

In real life, we often are in such situations that we need to evaluate some information that cannot be expressed in numerical values. In such cases the linguistic approach [3] and its extensions [4]-[8] can be seen as good alternatives. Actually, in medical community, the information often is characterized vaguely and imprecisely, which makes it hard to be evaluated by singular numerical values. For example, the expressions such as "very painful", "slightly painful", "medium" and "not very painful" are just some examples of the linguistic evaluations of subjective pain feeling that can be easily formulated by the patient. Also in group decision making cases, when the experts assess the effectiveness of treatment therapies for prostate cancer patients, the semantic terms such as "contraindicated", "doubtful", "acceptable", "possible", "suitable", "recommended" and "strongly recommended" can be used. Comparing to the numerical quantity, the linguistic approach is regarded in [9], [10] as a more realistic, intuitionistic and natural method. Due to the advantages of the linguistic approach, an extensive application has been presented in [11], [12]. Reasonable results have been reported, e.g., the adoption of the probabilistic model and the model of the 2-tuple fuzzy linguistic representations illustrated that the linguistic

approaches supplied the physicians with treatment effectiveness ranked from the strongly recommended to the contraindicated [1]. The linguistic approaches also supported investors with the valuable information how the capital can be effectively invested [12], [13].

By applying three models, namely, the probabilistic model [14], the 2-tuple linguistic representations [15] and the hesitant fuzzy linguistic term sets [2], we intend to rank the effectiveness of treatment alternatives from the most recommended to the contraindicated. The entire process will be defined in the linguistic framework.

The construction of this paper is organized as follows. In Section II, the preliminaries are presented. Section III provides two practical studies of the medical applications. Finally, conclusion and discussion are given in Sections IV and V, respectively.

II. PRELIMINARIES

In this section, some preliminary items are presented. We start with the detailed description of the probabilistic model.

In [14], a general property of a multi-expert decision-making problem is considered as the introduction of a finite set of experts denoted by $E = \{e_1, \cdots, e_p\}$ who are asked for selecting assessments stated in another finite set of alternatives $A = \{a_1, \cdots, a_n\}$. The assessments are expressed by semantic words in an order structured linguistic term set $S = \{s_0, \cdots, s_g\}$, such that $s_k < s_l$ if and only if $k < l$. An example of the ordered structured linguistic term set S is given below.

Example 1: Suppose that we determine a linguistic term set $S = \{s_0, s_1, s_2, s_3, s_4, s_5, s_6\}$ consisting of s_0 = "contraindicated" = C, s_1="doubtful"= D, s_2= "acceptable" = A, s_3= "possible" = P, s_4 = "suitable" = S, s_5 = "recommended" = R and s_6= "strongly recommended" = SR.

A. The Probabilistic Model

According to [14], the probability model mainly contains four steps:

- In the first step, all the assessments are collected in a judgment table as shown in Table I. Here each judgment $L_{ij}, i = 1, \cdots, n$ and $j = 1, \cdots, p$ is expressed by the linguistic term selected from the linguistic term set S.

We should emphasize that each linguistic term s_l, $l = 0, \cdots, g$ via a fuzzy number is associated with a general s-parametric membership function [16]-[18] given by (1), where $z = [0,1]$ is a symbolic reference set for all effectiveness terms, $z_{\min} = 0$, and h_z is defined as the distance between of the peaks between two adjacent fuzzy sets.

TABLE I. THE JUDGMENT TABLE OF THE PROBABILISTIC MODEL

Alternatives	Experts		
	e_1	⋯	e_p
a_1	L_{11}	⋯	L_{1p}
a_2	L_{21}	⋯	L_{2p}
⋯	⋯	⋯	⋯
a_n	L_{n1}	⋯	L_{np}

If we set $z_{\min}$ and h_z in (1) as fixed values when choosing $l = 0, \cdots, g$, then we will obtain the membership functions for $s_0, \cdots, s_g$ as

$$\mu_{s_l}(z) = \begin{cases} 2\left(\frac{z-((z_{\min}-h_z)+h_z l)}{h_z}\right)^2 \text{ for} \\ (z_{\min}-h_z)+h_z l \le z \le \left(z_{\min}-\frac{h_z}{2}\right)+h_z l, \\ 1-2\left(\frac{z-(z_{\min}+h_z l)}{h_z}\right)^2 \text{ for} \\ \left(z_{\min}-\frac{h_z}{2}\right)+h_z l \le z \le z_{\min}+h_z l, \\ 1-2\left(\frac{z-(z_{\min}+h_z l)}{h_z}\right)^2 \text{ for} \\ z_{\min}+h_z l \le z \le \left(z_{\min}+\frac{h_z}{2}\right)+h_z l, \\ 2\left(\frac{z-((z_{\min}+h_z)+h_z l)}{h_z}\right)^2 \text{ for} \\ \left(z_{\min}+\frac{h_z}{2}\right)+h_z l \le z \le (z_{\min}+h_z)+h_z l. \end{cases} \quad (1)$$

- X_{a_i} is assumed as a random preference value for each alternative $a_i, i = 1, \cdots, n$, with associated probablility distribution P defined by [8] as

$$P\left(X_{a_i} = s_l\right) = P_E\left(\{e_j \in E \middle| L_{ij} = s_l\}\right). \quad (2)$$

It is worth highlighting that the statement of random preference X_{a_i} is a crucial procedure in the approach of probability. Since each X_{a_i} is stochastically independent of each other, it will make the comparisons of any two random preferences to be possible.

- The choice value $V(a_i)$ for each alternative $a_i, i = 1, \cdots, n$, is computed by the choice function implemented by

$$V(a_i) = \sum_{i \neq j} P\left(X_{a_i} \geq X_{a_j}\right)$$

$$= \sum_{i \neq j} \sum_{s_l \in S} \left[P\left(X_{a_i} = s_l\right) \sum_{\substack{L_{ij} \in S \\ s_l \geq L_{ij}}} P\left(X_{a_j} = L_{ij}\right) \right], \quad (3)$$

where the quantity $P\left(X_{a_i} \geq X_{a_j}\right)$ could be interpreted as the probability of "the performance of a_i is as least as good as that of a_j".

- Finally, by ranking the choice values obtained by the former step, we can select the optimal one by

$$a_{\text{optimal}} = \max_{a_i \in A}(V(a_i)). \quad (4)$$

B. The Model of 2-tuple Linguistic Representation

In this model, the physicians' judgments of the treatments are represented by the 2-tuples of the form of (s_l, α), where $s_l \in S$ is a semantic word to which a fuzzy set is assigned and $\alpha \in [-0.5, 0.5)$ is defined as a numerical value.

A 2-tuple linguistic representation model presented in [15] composes the following steps:

- Each judgment that is expressed by a semantic word in Table I is changed into a 2-tuple linguistic representation as (s_l, α). If $s_l \in S$, then $(s_l, 0)$ will reflect s_l. Next, $x_{a_i} = \{(s_l, \alpha)\}$ is defined as a finite set that consists of judgments of the 2-tuple linguistic representations for each alternative $a_i, i = 1, \cdots, n$.

- Two transformations are used in this model.

The first transform Δ^1 maps a 2-tuple representation $(s_l, \alpha) \in S \times [-0.5, 0.5)$ of an alternative a_i into a numerical value $\beta_{a_i}^{e_j} \in [0, g], i = 1, \cdots, n, j = 1, \cdots, p$, in which $\beta_{a_i}^{e_j} = l + \alpha$. The action of Δ^1 is formalized by

$$\Delta^1: \quad S \times [-0.5, 0.5) \rightarrow [0, g]$$
$$(s_l, \alpha) \rightarrow \beta_{a_i}^{e_j} = l + \alpha. \quad (5)$$

We explicate the performance of Δ^1 by the following example.

Example 2: Let $S = \{s_0, \cdots, s_6\}$. In Table II the assessment of a_1, given by expert e_3, is expressed by the semantic term s_2 = "acceptable" =A. By the model of 2-tuple linguistic representation we can employ the judgment (A, 0) presented in Table III for s_2 = "acceptable" =A

TABLE II. THE DECISION TABLE OF THE JUDGMENTS FOR EXAMPLE 2

Alternatives	Experts			
	e_1	e_2	e_3	e_4
a_1	s_0	s_1	s_2	s_3
a_2	s_2	s_0	s_1	s_4
a_3	s_3	s_4	s_5	s_1
a_4	s_2	s_1	s_2	s_0

TABLE III. THE JUDGMENT TABLE OF THE 2-TUPLE LINGUISTIC REPRESENTATIONS

Experts	Alternatives			
	a_1	a_2	a_3	a_4
e_1	$(C, 0)$	$(A, 0)$	$(P, 0)$	$(A, 0)$
e_2	$(D, 0)$	$(C, 0)$	$(S, 0)$	$(D, 0)$
e_3	$(A, 0)$	$(D, 0)$	$(R, 0)$	$(A, 0)$
e_4	$(P, 0)$	$(S, 0)$	$(D, 0)$	$(C, 0)$

and $\alpha = 0$. The 2-tuple linguistic representations for other judgments are aggregated in Table III.

Due to the first transformation, the 2-tuple representation of (A, 0) can be performed as a numerical value $\beta_{a_1}^{e_3} = l + \alpha = 2 + 0 = 2$, which belongs to the interval [0, 6]. Furthermore, $x_{a_1} = \{(C, 0), (D, 0), (A, 0), (P, 0)\}$ consists of the judgments of the 2-tuple linguistic representations for alternative a_1.

In addition, we use the notation, $\overline{\beta_{a_i}}$, to represent the arithmetic mean of the sum of $\beta_{a_i}^{e_j}$, in which i = 1, …, n and j = 1, …, p. The computation of $\overline{\beta_{a_i}}$ is given by

$$\overline{\beta_{a_i}} = \frac{1}{p} \sum_{j=1}^{p} \beta_{a_i}^{e_j}. \quad (6)$$

Example 3: From Table III we obtain $x_{a_1} = \{(C, 0), (D, 0), (A, 0), (P, 0)\} = \{(s_0, 0), (s_1, 0), (s_2, 0), (s_3, 0)\}$, which leads to $\beta_{a_1}^{e_1} = 0 + 0 = 0$, $\beta_{a_1}^{e_2} = 1 + 0 = 1$, $\beta_{a_1}^{e_3} = 2 + 0 = 2$ and $\beta_{a_1}^{e_4} = 3 + 0 = 3$. According to (6), the arithmetic mean of $\overline{\beta_{a_1}}$ is equal to $\frac{1}{4}(0 + 1 + 2 + 3) = 1.5$.

The second transformation Δ^2 can be regarded as an inverse of the first one, i.e., it maps the numerical value $\overline{\beta_{a_i}} \in R$ into a 2-tuple (s_l, α) by

$$\Delta^2: \quad \overline{\beta_{a_i}} \rightarrow S \times [-0.5, 0.5)$$
$$\overline{\beta_{a_i}} \rightarrow (s_l, \alpha). \quad (7)$$

Here s_l has the closest index label to $\overline{\beta_{a_i}}$, the interval of $[0, g]$ represents the space consisting of the semantic label indices in the linguistic term set $S = \{s_l\}$, $l = 0, \cdots, g$.

Example 4: Let $S = \{s_0, \cdots, s_6\}$. According to (6), $\overline{\beta_{a_2}} = \frac{1}{4}(2 + 0 + 1 + 4) = 1.75$. Since 1.75 is closer to s_2 than to

s_1, then we choose s_2 as the semantic word. The difference between 1.75 and 2 is 0.25, and 1.75 lies to the left of 2. Therefore, we choose –0.25 to be the value of α. By means of the second transformation, $\Delta^2(1.75) = (s_2, -0.25)$, which is depicted in Figure 1.

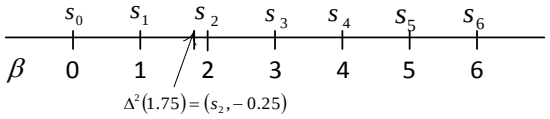


Figure 1. The 2-tuple linguistic representation of $\beta = 1.75$

- The third step contains the computation of the arithmetic mean $\bar{x}^e_{a_i}$ of 2-tuples for each alternative $a_i, i = 1, \cdots, n$, This is formalized by

$$\bar{x}^e_{a_i} = \Delta^2\left(\overline{\beta_{a_i}}\right). \quad (8)$$

Since the arithmetic means, supplied from the previous step, are presented by 2-tuples, a computational technique to compare the arithmetic mean for each alternative proposed in [15] is given as follows.

- Let (s_k, α_1) and (s_l, α_2) be two 2-tuples linguistic representations, with each one representing a counting of information as follows:
  1) if $k < l$, then (s_k, α_1) is smaller than (s_l, α_2).
  2) if $k = l$, we check the following conditions:
     if $\alpha_1 = \alpha_2$, then (s_k, α_1) and (s_l, α_2) represents the same information.
     if $\alpha_1 < \alpha_2$, then (s_k, α_1) is smaller than (s_l, α_2).
     if $\alpha_1 > \alpha_2$, then (s_k, α_1) is greater than (s_l, α_2).

- At last, by comparing the arithmetic values with each other and ranking the alternatives, the optimal alternative(s) will be obtained.

C. The Hesitant fuzzy Linguistic Term Sets

For better understanding of the later application of the hesitant fuzzy linguistic term set (HFLTS) in making medication prognosis for the second prostate cancer patient, we need shortly review the conception of the hesitant fuzzy linguistic term set [2].

Definition 1: Let $S = \{s_0, \ldots, s_g\}$ be a linguistic term set. A hesitant fuzzy linguistic term set, denoted by H_s, is an ordered finite subset of the consecutive linguistic terms of *S*.

The empty HFLTS and the full HFLTS for elements s ∈ S are defined as follows [2]:

- Empty HFLTS: $H_s(s) = \emptyset$,
- Full HFLTS: $H_s(s) = S$.

Example 5: Let us assume that $S = \{s_0, s_1, s_2, s_3, s_4, s_5\}$ is a linguistic term set describing the effectiveness of some treatment schemes for prostate cancer patients. If s_0 = "contraindicated" = "C", s_1 = "acceptable" = "A", s_2 = "possible" = "P", s_3 = "suitable" = "S", s_4 = "recommended" = "R", and s_5 = "strongly recommended"= "SR", then a HFLTS might be $H_s(s) = \{s_2, s_3, s_4\} = \{P, S, R\}$ = between" P" and "R".

We still suppose that $A = \{a_i\}$, i =1, …, n represents a set including n types of treatment alternatives, $E = \{e_j\}$, j = 1, …, p, denotes a collection of p experts and $S = \{s_l\}$, l = 0, …, g, consists of $g + 1$ linguistic assessments. We use the combination of comparative terms and the words selected from S to express the judgments P_{ij} (the judgments of e_j referring to treatment a_i). Especially, S contains the elements ordered in such a way that $s_q \leq s_r$ if and only if $q \leq r$, $q, r = 0, \ldots, g$ [14]. It is worth highlighting that each s_l is represented by a continuous fuzzy number in the Left-Right form, (*L-R* form) [19]. The aggregated preferences from individual experts are presented in Table IV.

TABLE IV. THE HESITANT JUDGMENT TABLE

Alternatives	Experts		
	e_1	…	e_p
a_1	P_{11}	…	P_{1p}
a_2	P_{21}	…	P_{2p}
…	…	…	…
a_n	P_{n1}	…	P_{np}

Sets H_S^{ij} contain these elements of S which consider the judgments P_{ij}. $H_S^{ij} \subseteq S$,i = 1, …, n, j = 1, …, p. By utilizing the operation of union on sets H_S^{ij} on each row, the new generated HFLTS, U_{a_i}, i = 1, …, n, becomes a subset of the linguistic term set *S* and obtains all conceivable effectiveness assessments. Subsequently, the union of all the elements in U_{a_i} yields the effectiveness of each alternative denoted by $\text{Eff}(a_i) = W(a_i)$. We illustrate this by Example 6.

Example 6: Consider three alternatives $\{a_1, a_2, a_3\} \subset A$ which represent three different kinds of treatment schemes. Three experts $\{e_1, e_2, e_3\} \subset E$ express their preferences about these treatment alternatives by combining comparative terms and words selected from the linguistic term set $S = \{s_0, s_1, s_2, s_3, s_4, s_5\}$, in which s_0 = "contraindicated" = "C", s_1 = "acceptable" = "A", s_2 = "possible" = "P", s_3 = "suitable" = "S", s_4 = "recommended" = "R", s_5 = "strongly recommended" = "SR". Table V displays the collection of preferences and Table VI shows the HFLTS's subset of *S*.

TABLE V. THE HESITANT JUDGMENT TABLE FOR EXAMPLE 6

Alternatives	Experts		
	e_1	e_2	e_3
a_1	$\leq A$	$[A, P]$	R
a_2	$[A, S]$	$[P, S]$	C
a_3	$[S, R]$	$[C, P]$	$[P, S]$

TABLE VI. THE SETS H_S^{ij} AS THE SUBSETS OF S DUE TO EXAMPLE 6

Alternatives	Experts		
	e_1	e_2	e_3
a_1	$\{s_0, s_1\}$	$\{s_1, s_2\}$	$\{s_4\}$
a_2	$\{s_1, s_2, s_3\}$	$\{s_2, s_3\}$	$\{s_4\}$
a_3	$\{s_3, s_4\}$	$\{s_0, s_1, s_2\}$	$\{s_2, s_3\}$

We use the operation of union on the sets H_S^{ij} for each row, denoted by $U_{a_i} = \cup H_S^{ij}$, $i, j = 1, \ldots, 3$, to obtain all possible effectiveness assessments for each alternative.

$U_{a_1} = \{C, A\} \cup \{A, P\} \cup \{R\} = \{C, A, P, R\} = \{s_0, s_1, s_2, s_4\}$, $U_{a_2} = \{A, P, S\} \cup \{P, S\} \cup \{C\} = \{s_0, s_1, s_2, s_3\}$ and $U_{a_3} = \{S, R\} \cup \{C, A, P\} \cup \{P, S\} = \{C, A, P, S, R\} = \{s_0, s_1, s_2, s_3, s_4\}$. Hence, the effectiveness assessments of a_1, a_2 and a_3 can be given as follows:

$Eff(a_1) = W_{a_1} = s_0 + s_1 + s_2 + s_4$,
$Eff(a_2) = W_{a_2} = s_0 + s_1 + s_2 + s_3$,
$Eff(a_3) = W_{a_3} = s_0 + s_1 + s_2 + s_3 + s_4$.

In the HFLTS, the judgment expressions based on comparative terms like, e.g., between … and …, greater than … or less than …, [2] will be used to supply the preferences. Symbolically, we denote "between acceptable and possible" as [A, P], "less than possible" as $\leq$ P and "greater than suitable" as $\geq S$. Single words such as "contraindicated" are abbreviated as "C".

After obtaining the sets U_{a_i} containing all conceivable effectiveness assessments, we would like to utilize the algorithm for calculating the sum of fuzzy numbers in the Left-Right (*L-R*) form, and later on to transform the *L-R* form into the interval form [19]. Finally, by adopting the technique of ranking fuzzy numbers in compliance with [20], we hopefully can select the most consensual alternative or alternatives.

We recall the information about fuzzy numbers expressed in the *L-R* form. We suppose that s_q and s_r are two fuzzy numbers in the *L-R* form, in which $q, r = 0, \ldots, g$. We describe $s_q = \left(m_{s_q}, \alpha_{s_q}, \beta_{s_q}\right)_{LR}$ and $s_r = \left(m_{s_r}, \alpha_{s_r}, \beta_{s_r}\right)_{LR}$ in which m_{s_q} and m_{s_r} are called the mean values, α_{s_q} and α_{s_r} are defined as the left spreads, β_{s_q} and β_{s_r} are known as right spreads, respectively. The union of s_q and s_r is calculated by

$$s_q + s_r = \left(m_{s_q} + m_{s_r}, \alpha_{s_q} + \alpha_{s_r}, \beta_{s_q} + \beta_{s_r}\right)_{LR} \quad (9)$$

Being able to rank the fuzzy numbers obtained from (9), we need first transfer them into interval forms.

We review the fuzzy number transformation from the *L-R* form into interval form in [19]. Assume $W_{a_i} = \left(m_{W_{a_i}}, \alpha_{W_{a_i}}, \beta_{W_{a_i}}\right)_{LR}$ is a fuzzy number in the *L-R* form. The interval form of W_{a_i} is given by

$$W_{a_i} = \left[b_{W_{a_i}}^-, m_{W_{a_i}}, b_{W_{a_i}}^+\right]_{int} \quad (10)$$

in which $m_{W_{a_i}}$ is the mean value, $b_{W_{a_i}}^- = m_{W_{a_i}} - \alpha_{W_{a_i}}$ and $b_{W_{a_i}}^+ = m_{W_{a_i}} + \beta_{W_{a_i}}$ are defined as the left and the right border, respectively. The membership function associated with the fuzzy number $W_{a_i} = \left[b_{W_{a_i}}^-, m_{W_{a_i}}, b_{W_{a_i}}^+\right]_{int}$ can be given by the following *s*-functions [16]-[18]:

$$y = \mu_{W_{a_i}}(z) = \begin{cases} Left\left(\mu_{W_{a_i}}(z)\right) & \text{for} \quad z \leq m_{W_{a_i}}, \\ Right\left(\mu_{W_{a_i}}(z)\right) & \text{for} \quad z \geq m_{W_{a_i}}, \end{cases} \quad (11)$$

in which

$$Left\left(\mu_{W_{a_i}}(z)\right) = \begin{cases} 2\left(\dfrac{z - b_{W_{a_i}}^-}{m_{W_{a_i}} - b_{W_{a_i}}^-}\right)^2 & \text{for} \quad b_{W_{a_i}}^- \leq z \leq c_{W_{a_i}}^1, \\ 1 - 2\left(\dfrac{z - m_{W_{a_i}}}{m_{W_{a_i}} - b_{W_{a_i}}^-}\right)^2 & \text{for} \quad c_{W_{a_i}}^1 \leq z \leq m_{W_{a_i}} \end{cases} \quad (12)$$

and

$$Right\left(\mu_{W_{a_i}}(z)\right) = \begin{cases} 1 - 2\left(\dfrac{z - m_{W_{a_i}}}{b_{W_{a_i}}^+ - m_{W_{a_i}}}\right)^2 & \text{for} \quad m_{W_{a_i}} \leq z \leq c_{W_{a_i}}^2, \\ 2\left(\dfrac{z - b_{W_{a_i}}^+}{b_{W_{a_i}}^+ - m_{W_{a_i}}}\right)^2 & \text{for} \quad c_{W_{a_i}}^2 \leq z \leq b_{W_{a_i}}^+, \end{cases} \quad (13)$$

where $c_{W_{a_i}}^1 = \frac{b_{W_{a_i}}^- + m_{W_{a_i}}}{2}$ and $c_{W_{a_i}}^2 = \frac{b_{W_{a_i}}^+ + m_{W_{a_i}}}{2}$ are arithmetic mean values.

Ranking fuzzy numbers in a decision-making environment is a very important and complex procedure. So far, the approaches to ranking fuzzy numbers have been proposed in [21]-[25]. Some of them are difficult to perform and others lead to different outcomes for a same problem. Therefore, a revised approach, based on [26] was explicated by Wang and Lee in [20]. In [20], the authors argued that "multipying the value on the horizontal axis with the value on the vertical axis often degrades the importance of the value on horizontal axis in ranking fuzzy numbers." Instead, Wang and Lee proposed a technique to overcome the shortcomings. The revised method is given by the following criteria:

- If $\bar{z}(W_{a_i}) > \bar{z}\left(W_{a_j}\right)$, then $W_{a_i} > W_{a_j}$.
- If $\bar{z}(W_{a_i}) < \bar{z}\left(W_{a_j}\right)$, then $W_{a_i} < W_{a_j}$.
- If $\bar{z}(W_{a_i}) = \bar{z}\left(W_{a_j}\right)$, then $W_{a_i} = W_{a_j}$, thereby we check the following conditions:
- If $\bar{\mu}(W_{a_i}) > \bar{\mu}\left(W_{a_j}\right)$, then $W_{a_i} > W_{a_j}$.
- If $\bar{\mu}(W_{a_i}) < \bar{\mu}\left(W_{a_j}\right)$, then $W_{a_i} < W_{a_j}$.
- If $\bar{\mu}(W_{a_i}) = \bar{\mu}\left(W_{a_j}\right)$, then $W_{a_i} = W_{a_j}$,

in which

$$\bar{z}(W_{a_i}) = \frac{\int_{b_{W_{a_i}}}^{m_{W_{a_i}}} zLeft\left(\mu_{W_{a_i}}(z)\right)dz + \int_{m_{W_{a_i}}}^{b_{W_{a_i}}^{+}} zLeft\left(\mu_{W_{a_i}}(z)\right)dz}{\int_{b_{W_{a_i}}}^{m_{W_{a_i}}} Left\left(\mu_{W_{a_i}}(z)\right)dz + \int_{m_{W_{a_i}}}^{b_{W_{a_i}}^{+}} Left\left(\mu_{W_{a_i}}(z)\right)dz} \tag{14}$$

and

$$\bar{\mu}_{W_{a_i}}(z) = \frac{\int_0^1 \mu Left\left(\mu_{W_{a_i}}(z)\right)^{-1} d\mu + \int_0^1 \mu Right\left(\mu_{W_{a_i}}(z)\right)^{-1} d\mu}{\int_0^1 Left\left(\mu_{W_{a_i}}(z)\right)^{-1} d\mu + \int_0^1 Right\left(\mu_{W_{a_i}}(z)\right)^{-1} d\mu} \tag{15}$$

Here, $\left(\bar{z}(W_{a_i}), \bar{\mu}_{W_{a_i}}(z)\right)$ is the centroid point of the fuzzy number W_{a_i}, $Left\left(\mu_{W_{a_i}}(z)\right)$ and $Right\left(\mu_{W_{a_i}}(z)\right)$ are called the left and the right membership functions of W_{a_i}, $Left\left(\mu_{W_{a_i}}(z)\right)^{-1}$ and $Right\left(\mu_{W_{a_i}}(z)\right)^{-1}$ are known as the inverse functions of $Left\left(\mu_{W_{a_i}}(z)\right)$ and $Right\left(\mu_{W_{a_i}}(z)\right)$, respectively.

III. PRACTICAL STUDIES

In this section, we want to present two practical studies in medical group decision-making task. The physicians from a MDT group (urologists and medical oncologists) are independently asked for providing the opinions on some treatment schemes for two separate prostate cancer patients. The methods of probabilistic model, the 2-tuple linguistic model are considered for the first prostate cancer patient and the approach of hesitant fuzzy linguistic term sets is applied to the second patient.

A. The Probabilistic Model

Let us suppose that $E = \{e_1, e_2, e_3, e_4\}$ denotes a collection consisting of four physicians. And another set $A = \{a_1, a_2, a_3, a_4, a_5, a_6\}$ contains sex types of treatment schemes for a prostate cancer patient, where a_1= "wait and see", a_2= "active monitoring", a_3= "symptom based treatment", a_4 = "brachytherapy", a_5 = "external beam radiation therapy" and a_6= "radical prostatectomy". Also, $L = \{s_0, s_1, s_2, s_3, s_4, s_5, s_6\}$ includes seven linguistic terms, in which s_0= "contraindicated", s_1= "doubtful", s_2= "acceptable", s_3= "possible", s_4 = "suitable", s_5= "recommended" and s_6= "strongly recommended".

By inserting $z_{\min} = 0$, $h_z = 0.167$ and l = 0 in (1), we obtain the function for s_0 = "contraindicated" expanded by

$$\mu_{s_0}(z) = \begin{cases} 2\left(\frac{z+0.167}{0.167}\right)^2 & \text{for} \quad -0.167 \le z \le -0.0835, \\ 1 - 2\left(\frac{z}{0.167}\right)^2 & \text{for} \quad -0.0835 \le z \le 0, \\ 1 - 2\left(\frac{z}{0.167}\right)^2 & \text{for} \quad 0 \le z \le 0.0835, \\ 2\left(\frac{z-0.167}{0.167}\right)^2 & \text{for} \quad 0.0835 \le z \le 0.167. \end{cases} \tag{16}$$

By following the same procedure for l =1, 2, 3, 4, 5 and 6 we generate membership functions

$$\mu_{s_1}(z) = \begin{cases} 2\left(\frac{z}{0.167}\right)^2 & \text{for} \quad 0 \le z \le 0.0835, \\ 1 - 2\left(\frac{z-0.167}{0.167}\right)^2 & \text{for} \quad 0.0835 \le z \le 0.167, \\ 1 - 2\left(\frac{z-0.167}{0.167}\right)^2 & \text{for} \quad 0.167 \le z \le 0.2505, \\ 2\left(\frac{z-0.334}{0.167}\right)^2 & \text{for} \quad 0.2505 \le z \le 0.334, \end{cases} \tag{17}$$

$$\mu_{s_2}(z) = \begin{cases} 2\left(\frac{z-0.167}{0.167}\right)^2 & \text{for} \quad 0.167 \le z \le 0.2505, \\ 1 - 2\left(\frac{z-0.334}{0.167}\right)^2 & \text{for} \quad 0.2505 \le z \le 0.334, \\ 1 - 2\left(\frac{z-0.334}{0.167}\right)^2 & \text{for} \quad 0.334 \le z \le 0.4175, \\ 2\left(\frac{z-0.501}{0.167}\right)^2 & \text{for} \quad 0.4175 \le z \le 0.501, \end{cases} \tag{18}$$

$$\mu_{s_3}(z) = \begin{cases} 2\left(\frac{z-0.334}{0.167}\right)^2 & \text{for} \quad 0.334 \le z \le 0.4175, \\ 1-2\left(\frac{z-0.501}{0.167}\right)^2 & \text{for} \quad 0.4175 \le z \le 0.501, \\ 1-2\left(\frac{z-0.501}{0.167}\right)^2 & \text{for} \quad 0.501 \le z \le 0.5845, \\ 2\left(\frac{z-0.668}{0.167}\right)^2 & \text{for} \quad 0.5845 \le z \le 0.668, \end{cases} \tag{19}$$

$$\mu_{s_4}(z) = \begin{cases} 2\left(\frac{z-0.501}{0.167}\right)^2 & \text{for} \quad 0.501 \le z \le 0.5845, \\ 1-2\left(\frac{z-0.668}{0.167}\right)^2 & \text{for} \quad 0.5845 \le z \le 0.668, \\ 1-2\left(\frac{z-0.668}{0.167}\right)^2 & \text{for} \quad 0.668 \le z \le 0.7515, \\ 2\left(\frac{z-0.835}{0.167}\right)^2 & \text{for} \quad 0.7515 \le z \le 0.835, \end{cases} \tag{20}$$

$$\mu_{s_5}(z) = \begin{cases} 2\left(\frac{z-0.668}{0.167}\right)^2 & \text{for} \quad 0.668 \le z \le 0.7515, \\ 1-2\left(\frac{z-0.835}{0.167}\right)^2 & \text{for} \quad 0.7515 \le z \le 0.835, \\ 1-2\left(\frac{z-0.835}{0.167}\right)^2 & \text{for} \quad 0.835 \le z \le 0.9185, \\ 2\left(\frac{z-1.002}{0.167}\right)^2 & \text{for} \quad 0.9185 \le z \le 1.002, \end{cases} \tag{21}$$

and

$$\mu_{s_6}(z) = \begin{cases} 2\left(\frac{z-0.835}{0.167}\right)^2 & \text{for} \quad 0.835 \le z \le 0.9185, \\ 1-2\left(\frac{z-1.002}{0.167}\right)^2 & \text{for} \quad 0.9185 \le z \le 1.002, \\ 1-2\left(\frac{z-1.002}{0.167}\right)^2 & \text{for} \quad 1.002 \le z \le 1.0855, \\ 2\left(\frac{z-1.169}{0.167}\right)^2 & \text{for} \quad 1.0855 \le z \le 1.169. \end{cases} \tag{22}$$

We sample all functions (16)–(22) in a family of fuzzy numbers restrictions, which are plotted in Figure 2.

By using the probabilistic model, we collect all the experts' judgments in Table VII, whereas the random preference value of each judgment is given in Table VIII.

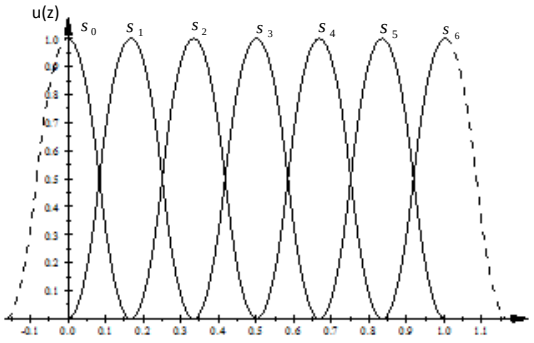


Figure 2. S-parametric membership functions for linguistic fuzzy sets $s_0 - s_6$

TABLE VII. THE COLLECTION OF THE JUDGMENTS FOR THE FIRST PATIENT

Alternatives	Experts			
	e_1	e_2	e_3	e_4
a_1	s_0	s_0	s_0	s_0
a_2	s_6	s_6	s_5	s_5
a_3	s_0	s_0	s_0	s_0
a_4	s_3	s_2	s_4	s_4
a_5	s_3	s_1	s_3	s_4
a_6	s_4	s_5	s_4	s_5

TABLE VIII. THE AGGREGATION OF RANDOM PREFERENCES FOR THE FIRST PATIENT

	Random Preferences						
	s_0	s_1	s_2	s_3	s_4	s_5	s_6
X_{a_1}	1	0	0	0	0	0	0
X_{a_2}	0	0	0	0	0	0.5	0.5
X_{a_3}	1	0	0	0	0	0	0
X_{a_4}	0	0	0.25	0.25	0.5	0	0
X_{a_5}	0	0.25	0	0.5	0.25	0	0
X_{a_6}	0	0	0	0	0.5	0.5	0

By using (3), we calculate the choice value for a_1 as the following structure:

$$\begin{aligned} V(a_1) &= \sum_{1 \ne j} P\left(X_{a_1} \ge X_{a_j}\right) \\ &= \sum_{1 \ne j} \sum_{s_l \in S} [P\left(X_{a_1} = s_l\right) \sum_{\substack{L_{ij} \in S \\ s_l \ge L_{ij}}} P(X_{a_j} = L_{ij})] \\ &= P\left(X_{a_1} \ge X_{a_2}\right) + \cdots + P\left(X_{a_1} \ge X_{a_6}\right) \\ &= 0 + 1 + 0 + 0 + 0 = 1. \end{aligned}$$

TABLE IX. THE COLLECTION OF CHOICE VALUES

The Collection of Choice Values for Each Alternative					
$V(a_1)$	$V(a_2)$	$V(a_3)$	$V(a_4)$	$V(a_5)$	$V(a_6)$
1	5	1	3	2.625	4.25

For other a_i, i = 2,3,4,5,6, $V(a_i)$ are calculated in the similar way as

$$V(a_2) = 1+1+1+1+1 = 5,$$
$$V(a_3) = 1+0+0+0+0 = 1,$$
$$V(a_4) = 1+0+1+0.75+0.25 = 3,$$
$$V(a_5) = 1+0+1+0.5+0.125 = 2.625,$$
and
$$V(a_6) = 1+0.25+1+1+1 = 4.25.$$

The collection of choice values for each $a_i, i = 1,\cdots,6$ is aggregated in Table IX. We choose the optimal therapy alternative by means of (4) as

$$a_{\text{optimal}} = \max_{a_i \in A}\{V(a_i)\} = \max\{1,5,1,3,2.625,4.25\} = 5 = V(a_2).$$

The value of 5 indicates the choice value of a_2 to be maximal. This means that the second therapy alternative is the most efficacious.

We want to confirm the result by applying the model of 2-tuple fuzzy linguistic representations.

B. The Model of 2-tuple Linguistic Representation

According to the algorithm for the model of 2-tuple representation, the judgment that is transformed into 2-tuples is given in Table X.

TABLE X. THE JUDGMENTS EXPRESSED IN THE 2-TUPLES REPRESENTATION MODEL FOR THE FIRST PATIENT

Experts	Alternatives					
	a_1	a_2	a_3	a_4	a_5	a_6
e_1	(C, 0)	(SR, 0)	(C, 0)	(P, 0)	(P, 0)	(S, 0)
e_2	(C, 0)	(SR, 0)	(C, 0)	(A, 0)	(H, 0)	(R, 0)
e_3	(C, 0)	(R, 0)	(C, 0)	(S, 0)	(P, 0)	(S, 0)
e_4	(C, 0)	(R, 0)	(C, 0)	(S, 0)	(S, 0)	(R, 0)

We calculate the arithmetic mean for the first alternative a_1 by means of (5).
$x_{a_1} = \{(C,0),(C,0),(C,0),(C,0)\}$ is a finite set consisting of four 2-tuple linguistic representations for the alternative a_1. By adopting (5), the arithmetic mean value for a_1 is calculated as:

$$\bar{x}^e_{a_1} = \Delta^2\left(\frac{1}{4}(0+0+0+0)\right) = \Delta^2(0) = (s_0, 0).$$

For the second alternative the arithmetic means value is given as follows:

$$\bar{x}^e_{a_2} = \Delta^2\left(\frac{1}{4}(6+6+5+5)\right) = \Delta^2(5.5) = (s_5, 0.5).$$

By the same reasoning, when setting i = 3,4,5,6 in (5), we implement

$$\bar{x}^e_{a_3} = \Delta^2\left(\frac{1}{4}(0+0+0+0)\right) = \Delta^2(0) = (s_0, 0),$$
$$\bar{x}^e_{a_4} = \Delta^2\left(\frac{1}{4}(3+2+4+1)\right) = \Delta^2(2.5) = (s_2, 0.5),$$
$$\bar{x}^e_{a_5} = \Delta^2\left(\frac{1}{4}(3+1+3+4)\right) = \Delta^2(2.75) = (s_3, -0.25),$$
and
$$\bar{x}^e_{a_6} = \Delta^2\left(\frac{1}{4}(4+5+4+5)\right) = \Delta^2(4.5) = (s_4, 0.5).$$

The collection of the arithmetic mean values for all alternatives is presented in Table XI.

TABLE XI. TABLE OF THE ARITHMETIC VALUES

The Collection of the Arithmetic Mean Values					
$\bar{x}^e_{a_1}$	$\bar{x}^e_{a_2}$	$\bar{x}^e_{a_3}$	$\bar{x}^e_{a_4}$	$\bar{x}^e_{a_5}$	$\bar{x}^e_{a_6}$
$(s_0, 0)$	$(s_5, 0.5)$	$(s_0, 0)$	$(s_2, 0.5)$	$(s_3, -0.25)$	$(s_4, 0.5)$

According to the computational technique presented earlier, we compare the above 2-tuples that represent the arithmetic values for all the alternatives. We obtain the result presented as $a_2 > a_6 > a_5 > a_4 > a_1 = a_3$, which shows that alternative a_2 is the most efficacious treatment scheme for this particular patient. This result converges to the previous result from "the probabilistic model".

C. The Hesitant Fuzzy Linguistic Term Sets

In this medical application, another prostate cancer patient is considered. We have five health professionals constitute the expert group $E = \{e_1, e_2, e_3, e_4, e_5\}$. The set $A = \{a_1, a_2, a_3, a_4, a_5, a_6, a_7\}$ contains seven types of treatment schemes alternatives, in which a_1= "active expectance", a_2= "active monitoring", a_3= "symptom based treatment", a_4 = "brachytherapy", a_5 = "external beam radiation therapy", a_6 = "adjuvant hormonal therapy" and a_7= "radical prostatectomy".

Furthermore, a linguistic term set $S = \{s_0, s_1, s_2, s_3, s_4, s_5\}$ includes six linguistic terms, in which s_0= "contraindicated" = "C", s_1= "acceptable" = "A", s_2= "possible"= "P", s_3 = "suitable" = "S", s_4= "recommended" = "R", and s_5= "strongly recommended" = "SR". Each linguistic term is associated with a fuzzy number restricted by general *s*-parametric function [16]-[18] given by (1). By choosing l = 0, …, 5, we obtain a family of six membership functions that map the effectiveness of the treatment therapies. Functions s_l are presented in Figure 3.

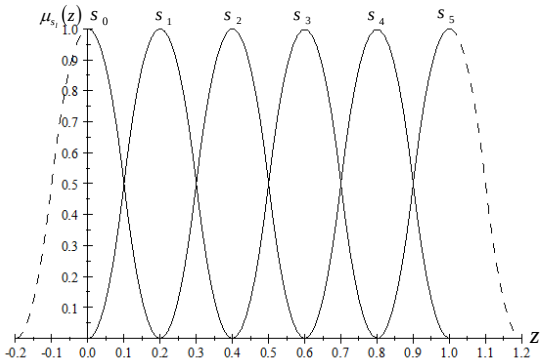


Figure 3. The family of hesitant membership functions $s_0 - s_5$

Here, s_0= "contraindicated" = "C" = $(0, 0.2, 0.2)_{LR}$, s_1= "acceptable" = "A" = $(0.2, 0.2, 0.2)_{LR}$, , s_2= "possible"= "P" = $(0.4, 0.2, 0.2)_{LR}$, s_3 = "suitable" = "S" = $(0.6, 0.2, 0.2)_{LR}$, s_4= "recommended" = "R" = $(0.8, 0.2, 0.2)_{LR}$ and s_5= "strongly recommended" = "SR" = $(1.0, 0.2, 0.2)_{LR}$. By combining comparative terms with single words, the experts express the preferences of the treatment therapies in a broader spectrum. These assessments are aggregated in Table XII.

TABLE XII. THE HESITANT JUDGMENT TABLE DESIGNED BY EXPERTS FOR THE SECOND PATIENT

Alternatives	Experts				
	e_1	e_2	e_3	e_4	e_5
a_1	[A, P]	≥ S	C	[A, P]	≤ P
a_2	≤ A	C	[A, P]	≤ A	R
a_3	[A, S]	C	C	C	≤ A
a_4	[A, S]	[P, S]	C	[A, S]	≤ S
a_5	[S, R]	[P, S]	≥ S	≥ P	≤ A
a_6	[A, S]	[P, S]	C	[P, R]	C
a_7	[S, R]	[C, P]	[P, S]	[P, R]	[P, R]

The assessment "[A, P]" denotes a comparative term, which indicates the terms $s \in S$ between "acceptable" and "possible". It is also a hesitant fuzzy linguistic term set H_S^{11}= $\{s_1, s_2\}$ in which $\{s_1, s_2\} \subset S$. "$\geq S$" can be interpreted as "greater than suitable", which symbolizes another hesitant fuzzy linguistic term set H_S^{12}= $\{s_0, s_1, s_2, s_3\}$. Furthermore, "$\leq$ A " means "less than acceptable", which assigns H_S^{21}= $\{s_0, s_1\}$.

In order to obtain the assessments as comprehensive as possible and prevent the information loss, for individual alternative, we perform the operation of union on fuzzy sets to aggregate all possible preferences in one set. Therefore,

$U_{a_1} = \{\mathrm{A, P}\} \cup \{\mathrm{S, R, SR}\} \cup \{\mathrm{C}\} \cup \{\mathrm{A, P}\} \cup \{\mathrm{C, A\ P}\}$
$= \{\mathrm{C, A, P, S, R, SR}\} = \{s_0, s_1, s_2, s_3, s_4, s_5\}$,

$U_{a_2} = \{\mathrm{C, A}\} \cup \{\mathrm{C}\} \cup \{\mathrm{A, P}\} \cup \{\mathrm{C, A}\} \cup \{\mathrm{R}\} = \{\mathrm{C, A, P, R}\} = \{s_0, s_1, s_2, s_4\}$,

$U_{a_3} = \{\mathrm{A, P, S}\} \cup \{\mathrm{C}\} \cup \{\mathrm{C}\} \cup \{\mathrm{C}\} \cup \{\mathrm{C, A}\} = \{\mathrm{C, A, P, S}\} = \{s_0, s_1, s_2, s_3\}$,

$U_{a_4} = \{\mathrm{A, P, S}\} \cup \{\mathrm{P, S}\} \cup \{\mathrm{C}\} \cup \{\mathrm{A, P, S}\} \cup \{\mathrm{C, A, P, S}\}$
$= \{\mathrm{C, A, P, S}\} = \{s_0, s_1, s_2, s_3\}$,

$U_{a_5} = \{\mathrm{S, R}\} \cup \{\mathrm{P, S}\} \cup \{\mathrm{S, R, SR}\} \cup \{\mathrm{P, S, R, SR}\} \cup \{\mathrm{C, A}\} = \{\mathrm{C, A, P, S, R, SR}\} = \{s_0, s_1, s_2, s_3, s_4, s_5\}$

$U_{a_6} = \{\mathrm{A, P, S}\} \cup \{\mathrm{C}\} \cup \{\mathrm{P, S}\} \cup \{\mathrm{C}\} \cup \{\mathrm{P, S, R}\} = \{\mathrm{C, A, P, S, R}\} = \{s_0, s_1, s_2, s_3, s_4\}$

and

$U_{a_7} = \{\mathrm{S, R}\} \cup \{\mathrm{C, A, P}\} \cup \{\mathrm{P, S}\} \cup \{\mathrm{P, S, R}\} = \{\mathrm{C, A, P, S, R}\} = \{s_0, s_1, s_2, s_3, s_4\}$.

We recall the union of two fuzzy numbers which can be performed by (9). Thereby, the effectiveness of a_1 and a_5 can be calculated as

$Eff(a_1) = W_{a_1} = s_0 + s_1 + s_2 + s_3 + s_4 + s_5 = (0, 0.2, 0.2)_{LR} + \cdots + (1.0, 0.2, 0.2)_{LR} = (3.0, 1.2, 1.2)_{LR} \equiv [1.8, 3.0, 4.2]_{int}$

and

$Eff(a_5) = W_{a_1} = s_0 + s_1 + s_2 + s_3 + s_4 + s_5 = (0, 0.2, 0.2)_{LR} + \cdots + (1.0, 0.2, 0.2)_{LR} = (3.0, 1.2, 1.2)_{LR} \equiv [1.8, 3.0, 4.2]_{int}$.

The membership functions for a_1 and a_5 are given by

$$y = \mu_{W_{a_1}}(z) = \mu_{W_{a_5}}(z) = \begin{cases} Left\left(\mu_{W_{a_1}}(z)\right) & \text{for} \quad z \leq 3.0, \\ Right\left(\mu_{W_{a_1}}(z)\right) & \text{for} \quad z \geq 3.0, \end{cases} \tag{23}$$

in which

$$Left\left(\mu_{W_{a_1}}(z)\right) = Left\left(\mu_{W_{a_5}}(z)\right) = \begin{cases} 2\left(\dfrac{z - 1.8}{3.0 - 1.8}\right)^2 & \text{for} \quad 1.8 \leq z \leq 2.4, \\ 1 - 2\left(\dfrac{z - 3.0}{3.0 - 1.8}\right)^2 & \text{for} \quad 2.4 \leq z \leq 3.0, \end{cases} \tag{24}$$

and

$$Right\left(\mu_{W_{a_1}}(z)\right) = Right\left(\mu_{W_{a_5}}(z)\right)$$
$$= \begin{cases} 1 - 2\left(\frac{z-3.0}{4.2-3.0}\right)^2 & \text{for} \quad 3.0 \le z \le 3.6, \\ 2\left(\frac{z-4.2}{4.2-3.0}\right)^2 & \text{for} \quad 3.6 \le z \le 4.2. \end{cases} \quad (25)$$

The membership function of W_{a_1} and W_{a_5} is depicted in Figure 4.

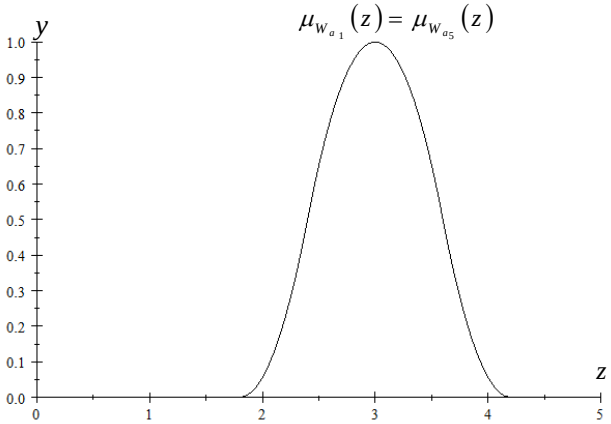


Figure 4. The membership function of W_{a_1} and W_{a_5}

The effectiveness of W_{a_2} is given by $Eff(a_1) = W_{a_2} = s_0 + s_1 + s_2 + s_3 + s_4 = (0, 0.2, 0.2)_{LR} + \cdots + (0.8, 0.2, 0.2)_{LR} = (1.4, 0.8, 0.8)_{LR} \equiv [0.6, 1.4, 2.2]_{int}$. The membership function of W_{a_2} is given by

$$Left\left(\mu_{W_{a_2}}(z)\right)$$
$$= \begin{cases} 2\left(\frac{z-0.6}{1.4-0.6}\right)^2 & \text{for} \quad 0.6 \le z \le 1.0, \\ 1 - 2\left(\frac{z-1.4}{1.4-0.6}\right)^2 & \text{for} \quad 1.0 \le z \le 1.4, \end{cases} \quad (26)$$

and

$$Right\left(\mu_{W_{a_2}}(z)\right)$$
$$= \begin{cases} 1 - 2\left(\frac{z-1.4}{2.2-1.4}\right)^2 & \text{for} \quad 1.4 \le z \le 1.8, \\ 2\left(\frac{z-2.2}{2.2-1.4}\right)^2 & \text{for} \quad 1.8 \le z \le 2.2. \end{cases} \quad (27)$$

The membership function of W_{a_2} is plotted in Figure 5.

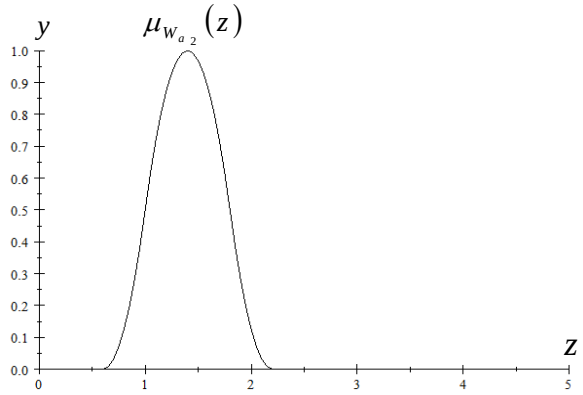


Figure 5. The membership function of W_{a_2}

For a_3 and a_4, the result is shown by $Eff(a_3) = W_{a_3} = s_0 + s_1 + s_2 + s_3 = (0, 0.2, 0.2)_{LR} + \cdots + (0.6, 0.2, 0.2)_{LR} = (1.2, 0.8, 0.8)_{LR} \equiv [0.4, 1.2, 2.0]_{int}$

and

$Eff(a_4) = W_{a_4} = s_0 + s_1 + s_2 + s_3 = (0, 0.2, 0.2)_{LR} + \cdots + (0.6, 0.2, 0.2)_{LR} = (1.2, 0.8, 0.8)_{LR} \equiv [0.4, 1.2, 2.0]_{int}$

The membership functions for W_{a_3} and W_{a_4} are given by

$$Left\left(\mu_{W_{a_3}}(z)\right) = Left\left(\mu_{W_{a_4}}(z)\right)$$
$$= \begin{cases} 2\left(\frac{z-0.4}{1.2-0.4}\right)^2 & \text{for} \quad 0.4 \le z \le 0.8, \\ 1 - 2\left(\frac{z-0.4}{1.2-0.4}\right)^2 & \text{for} \quad 0.8 \le z \le 1.2, \end{cases} \quad (28)$$

and

$$Right\left(\mu_{W_{a_3}}(z)\right) = Right\left(\mu_{W_{a_4}}(z)\right)$$
$$= \begin{cases} 1 - 2\left(\frac{z-1.2}{2.0-1.2}\right)^2 & \text{for} \quad 1.2 \le z \le 1.6, \\ 2\left(\frac{z-2.0}{2.0-1.2}\right)^2 & \text{for} \quad 1.6 \le z \le 2.0. \end{cases} \quad (29)$$

Figure 6 represents the membership function of W_{a_3} and W_{a_4}.

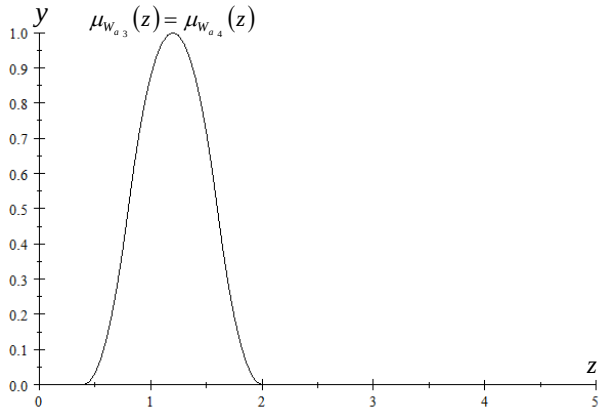


Figure 6. The membership function of W_{a_3} and W_{a_4}

And finally, the effectiveness of a_6 and a_7 is yielded by $Eff(a_6) = W_{a_6} = s_0 + s_1 + s_2 + s_3 + s_4 = (0, 0.2, 0.2)_{LR} + \ldots + (0.8, 0.2, 0.2)_{LR} = (2.0, 1.0, 1.0)_{LR} \equiv [1.0, 2.0, 3.0]_{int}$

and

$Eff(a_7) = W_{a_7} = s_0 + s_1 + s_2 + s_3 + s_4 = (0, 0.2, 0.2)_{LR} + \ldots + (0.8, 0.2, 0.2)_{LR} = (2.0, 1.0, 1.0)_{LR} \equiv [1.0, 2.0, 3.0]_{int}$.

The membership functions for W_{a_6} and W_{a_7} are given by

$$Left\left(\mu_{W_{a_6}}(z)\right) = Left\left(\mu_{W_{a_7}}(z)\right) = \begin{cases} 2\left(\frac{z-1.0}{2.0-1.0}\right)^2 & \text{for} \quad 1.0 \le z \le 1.5, \\ 1-2\left(\frac{z-2.0}{2.0-1.0}\right)^2 & \text{for} \quad 1.5 \le z \le 2.0, \end{cases} \tag{30}$$

and

$$Right\left(\mu_{W_{a_6}}(z)\right) = Right\left(\mu_{W_{a_7}}(z)\right) = \begin{cases} 1-2\left(\frac{z-2.0}{3.0-2.0}\right)^2 & \text{for} \quad 2.0 \le z \le 2.5, \\ 2\left(\frac{z-3.0}{3.0-2.0}\right)^2 & \text{for} \quad 2.5 \le z \le 3.0. \end{cases} \tag{31}$$

Figure 7 represents the membership function of W_{a_6} and W_{a_7}.

We first use (14) to calculate the horizontal coordinate of the centroid point of each fuzzy number W_{a_i}. If there exists identical horizontal coordinates, then (15) will be used to compute the vertical coordinate.

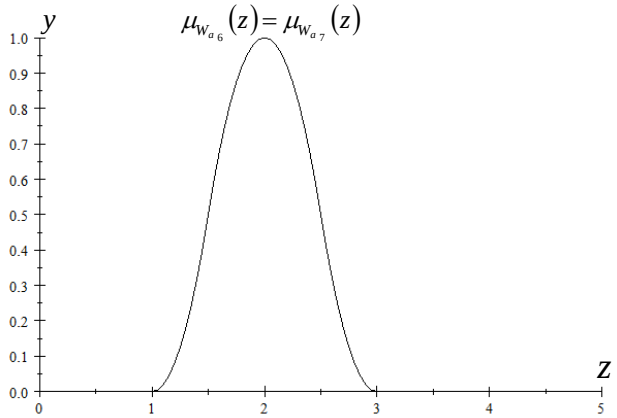


Figure 7. The membership function of W_{a_6} and W_{a_7}

By the insertion of the left respective the right membership functions and the borders of W_{a_1} and W_{a_5} in (14), we obtain the horizontal coordinate of W_{a_1} and W_{a_5} presented as follows:

$$\bar{z}(W_{a_1}) = \bar{z}(W_{a_5}) = \frac{\int_{1.8}^{3.0} zLeft\left(\mu_{W_{a_1}}(z)\right)dz + \int_{3.0}^{4.2} zLeft\left(\mu_{W_{a_1}}(z)\right)dz}{\int_{1.8}^{3.0} Left\left(\mu_{W_{a_1}}(z)\right)dz + \int_{3.0}^{4.2} Left\left(\mu_{W_{a_1}}(z)\right)dz} = \frac{3.6}{1.2} = 3.0.$$

By the same procedure, we obtain the horizontal coordinates for the remained alternatives presented below:

$$\bar{z}(W_{a_2}) = \frac{\int_{0.6}^{1.4} zLeft\left(\mu_{W_{a_2}}(z)\right)dz + \int_{1.4}^{2.2} zLeft\left(\mu_{W_{a_2}}(z)\right)dz}{\int_{0.6}^{1.4} Left\left(\mu_{W_{a_2}}(z)\right)dz + \int_{1.4}^{2.2} Left\left(\mu_{W_{a_2}}(z)\right)dz} = \frac{1.12}{0.8} = 1.4$$

$$\bar{z}(W_{a_3}) = \bar{z}(W_{a_4}) = \frac{\int_{0.4}^{1.2} zLeft\left(\mu_{W_{a_3}}(z)\right)dz + \int_{1.2}^{2.0} zLeft\left(\mu_{W_{a_3}}(z)\right)dz}{\int_{0.4}^{1.2} Left\left(\mu_{W_{a_3}}(z)\right)dz + \int_{1.2}^{2.0} Left\left(\mu_{W_{a_3}}(z)\right)dz} = \frac{2.073}{0.8} \approx 2.5913$$

and finally,

$$\bar{z}(W_{a_6}) = \bar{z}(W_{a_7}) = \frac{\int_{1.0}^{2.0} zLeft\left(\mu_{W_{a_6}}(z)\right)dz + \int_{2.0}^{3.0} zLeft\left(\mu_{W_{a_6}}(z)\right)dz}{\int_{1.0}^{2.0} Left\left(\mu_{W_{a_6}}(z)\right)dz + \int_{2.0}^{3.0} Left\left(\mu_{W_{a_6}}(z)\right)dz} = \frac{2.0}{1.0} = 2.$$

Since no identical horizontal coordinates are found, we do not need to compute the values of the vertical coordinates. By means of the criteria introduced in [35] we obtain the following results: 3 > 2.5913 > 2 > 1.4, i.e. $\bar{z}(W_{a_1}) = \bar{z}(W_{a_5}) > \bar{z}(W_{a_3}) = \bar{z}(W_{a_4}) > \bar{z}(W_{a_6}) = \bar{z}(W_{a_7}) > \bar{z}(W_{a_2})$. The first a_1 = "active expectance" and the fifth alternative a_5 = "external beam radiation therapy" have the most optimal effectiveness found for the second patient according to our computation.

IV. CONCLUSION

According to the physicians' requirements, we seek the arrangements of the effectiveness of treatment alternatives from the most recommended to the contraindicated for two separate prostate cancer patients.

Three approaches, such as the probabilistic model, the model of 2-tuple linguistic representations and the hesitant fuzzy linguistic term sets have been applied to two multi-expert decision-making cases. The convergence results from the first two approaches verify the high reliability of adopting the linguistic approach in solving group decision making problems. Moreover, the independent assumed preferences of each alternative make the computation of comparing the probabilities easy to be performed. Especially, the use of the 2-tuple linguistic representation model prevents the loss of information and makes the result more precise. The use of *s*-parametric membership functions not only increases the accuracy rate of the comparative analysis, but also facilitates the transformation process from the linguistic preferences to the numerical values. In the approach of hesitant fuzzy linguistic term set, the horizontal coordinates of the centroid point of fuzzy numbers are adopted for ranking the fuzzy numbers in a decision-making environment. The calculating process has its complexity but the technique is reliable.

V. DISCUSSION

We found all three methods very interesting in decision-making process when panelists were not unanimous. The results seem to be reasonable from the clinical point of view. The process of sampling the data by filling the questionnaires was easy and quickly accomplished, especially in the probabilistic model and the model of 2-tuple linguistic representations. However, we do encountered some issues in filling questionnaires for the hesitant fuzzy linguistic term sets, mostly because it is not as intuitive as the two aforementioned methods. The filling of hesitant fuzzy linguistic term sets questionnaire needs a 2-3 minutes preparation, just to maintain the homogenicity of the answers. The authors have a feeling that the hesitant fuzzy linguistic term sets is a reliable method, but probably better to use in other conditions than in prostate cancer decisions making, mostly because of its logistical and practical problems. We hope to soon introduce one of the models in our clinical practice to assess the method in a real life conditions. Hopefully, this approach can allow us to find better treatment strategies and to give prostate cancer patients more flexibility concerning the treatment options. This should be a great complement to the current guidelines and scientific society recommendations.

ACKNOWLEDGMENT

The authors are very thankful for the support and suggestions provided by physicians from the Urology Department of Blekinge County Hospital in Karlskrona, Sweden.

REFERENCES

[1] H. Zettervall, E. Rakus-Andersson, and J. Frey, "Applied Multi-Expert Decision-Making Issue Based on Linguistic Models for Prostate Cancer Patients." Proceedings of the Fifth International Conference on Bioinformatics, Biocomputational Systems, and Biotechnologies (BIOTECHNO 2013), Lisbon, Portugal, March 2013, pp.71–76.

[2] R. M. Rodriguez, L. Martinez, and F. Herrera, "Hesitant fuzzy linguistic term sets for decision making," IEEE Trans. on Fuzzy Systems, vol. 20, Issue. 1, February 2012, pp.109-119.

[3] L. A. Zadeh, "The concept of a linguistic variable and its applications to approximate reasoning," Inform. Sci., pt. I, no. 8, 1975, pp.199–249.

[4] D. Dubois and H. Prade, Fuzzy Sets and Systems: Theory and Applications. New York: Kluwer, 1980.

[5] K. Atanassov, "Intuitionistic fuzzy sets," Fuzzy Sets and Systems, vol. 20, 1986, pp. 87-96.

[6] S. Miyamoto, "Remarks on basics of fuzzy sets and fuzzy multisets," Fuzzy Sets and Systems, vol. 156, 2005, pp. 427-431.

[7] R.R. Yager, "On the theory of bags," International Journal of General Systems, vol. 13, 1986, pp. 23-37.

[8] S. Miyamoto, "Multisets and fuzzy multisets," In Z.-Q. Liu, S. Miyamoto (Eds.), Soft Computing and Human-Centered Machines, Springer, Berlin, 2000, pp. 9-33.

[9] M. Delgado, J. L. Verdegay, and M. A. Vila, "Linguistic decision making models," Int. J. Intell. Syst., vol. 7, 1992, pp. 479–492.

[10] F. Herrera and J. L. Verdegay, "Linguistic assessments in group decision," Proceedings of the 1st European Congress on Fuzzy and Intelligent Technologies, Aachen, 1993, pp. 941–948.

[11] J. Kacprzyk and M. Fedrizzi, Multiperson Decision Making Models Using Fuzzy Sets and Possibility Theory, Kluwer Academic Publishers, Dordrecht, 1990

[12] M. Tong and P.P. Bonissone, "A linguistic approach to decision making with fuzzy sets," IEEE Trans. Systems, Man, and Cybernetics, vol. 10, 1980, pp.716–723.

[13] D. F. Li, "TOPSIS-based nonlinear-programming methodology for multiattribute decision making with interval-valued intuitionistic fuzzy sets," IEEE Trans. Fuzzy Syst., vol. 18, no. 2, April 2012, pp.299-311.

[14] V. N. Huynh and Y. Nakamori, "Multi-expert decision-making with linguistic information: a probabilistic-based model," Proceedings of the 38th Hawaii International Coference on System Sciences, 2005.

[15] F. Herrera and L. Martinez, "A 2-tuple fuzzy linguistic representation model for computing with words," IEEE Trans. Fuzzy Syst., vol. 8, 2000, pp.746–752.

[16] E. Rakus-Andersson and L. Jain, "Computational intelligence in medical decisions making," Recent Advances in Decision Making, Springer, Berlin Heidelberg, 2009, pp.145-159.

[17] E. Rakus-Andersson, H. Zettervall, and M. Erman, "Prioritization of weighted strategies in the multi-player games with fuzzy entries of the payoff matrix," Int. J. of General Systems, vol, 39, Issue 3, 2010, pp.291-304.

[18] H. Zettervall, E. Rakus-Andersson, and H. Forssell: "The Mamdani controller in prediction of the survival length in elderly gastric cancer patients," Proceedings of Bioinformatics, Rome, Italy, 2011, pp.283-286.

[19] E. Rakus-Andersson and H. Zettervall, "Different approaches to operations on continuous fuzzy numbers," Developments of Fuzzy Sets, Intuitionistic Fuzzy Sets and Generalized Nets, Related Topics. Foundations, vol. I, Eds: Atanassov K., Countas P., Kacprzyk, J., Melo-Pinto, P. et al., EXIT – The Publishing House of The Polish Academy of Sciences, 2008, pp.294-318.

[20] Y. J. Wang and H. S. Lee, "The revised method of ranking fuzzy numbers with an area between the centroid and original points," Computers and Mathematics with Applications, vol. 55, 2008, pp.2033-2042.

[21] G. Bortolan and R. Degani, "A review of some methods for ranking fuzzy numbers," Fuzzy Sets and Systems, vol. 15 1985, 1-19.

[22] G. Dias, "Ranking alternatives using fuzzy numbers," A computational approach, Fuzzy Sets and Systems, vol. 56, 1993, pp.247-252.

[23] K. M. Lee, C. H. Cho and H. Lee-Kwang, "Ranking fuzzy values with satisfaction," Fuzzy Sets and Systems, vol. 64, 1994, pp.295-311.

[24] I. Requena, M. Delgado and J. I. Verdagay, "Automatic ranking of fuzzy numbers with the criterion of decision-maker learnt by an artificial neural network," Fuzzy Sets and Systems, vol. 64, 1994, pp.1-19.

[25] P. Fortemps and M. Roubens, "Ranking and defuzzication methods based on area compensation," Fuzzy Sets and Systems, vol. 82, 1996, pp.319-330.

[26] T. C. Chu and C. T. Tsao, "Ranking fuzzy numbers with an area between the centroid and the original points," Computers and Mathematics with Applications, vol. 43, 2002, pp.111-117.

Real-time Visualization and Exploration of Protein Empty Space with Varying Parameters

Ondřej Strnad, Barbora Kozlíková, Vilém Šustr, and Jiří Sochor
Faculty of Informatics
Masaryk university, Brno, Czech Republic
Email: xstrnad2@fi.muni.cz, xkozlik@fi.muni.cz, xsustr@fi.muni.cz, sochor@fi.muni.cz

***Abstract*—Long-term research in the area of protein analysis proved the importance of an empty space situated inside these macromolecular structures. This empty space influences the protein function, characteristics or reactivity. Many algorithms enabling computation of these empty spaces (or voids) have been published and their results were evaluated by protein engineers to confirm their chemical relevance. However, not all detected voids inside protein are of the same importance. Thus, the examination and assessment of all voids must follow to reveal the important ones. In this phase the visual representation of voids is very valuable and substantially decreases the time spent in this evaluation phase. In this paper we present an extension of the algorithm for the visualization and further evaluation of protein voids in real-time. The user-driven approach enables to compute and display empty space that satisfies the input parameters instantly. The values of these parameters can be changed by the user anytime and the changes are immediately displayed and prepared for further exploration. Our improvements involve an exclusion of selected atom or group of atoms (ligands, ions) from the computation, which can change the size and shape of the detected void. Another improvement is related to the detection of the binding site which is usually located in one of the largest voids. So the algorithm suggests and visually separates (by different coloring) the largest void of given area. Several improvements were also made in the field of real-time exploration – currently the interaction on large structures is fluent. In consequence, the current version of the algorithm provides the biochemists with very adjustable and precise algorithm for detection of inner voids in a user-defined region of protein structures.**

Keywords-protein, empty space, void, visualization, real-time, cavity, volume, Voronoi diagram, Delaunay triangulation

I. Introduction

This paper is an extended version of the conference paper entitled ”Real-Time Visualization of Protein Empty Space with Varying Parameters” presented at the BIOTECHNO 2013: The Fifth International Conference on Bioinformatics, Biocomputational Systems and Biotechnologies, March 24 – 29, 2013, Lisbon, Portugal [1]. In this extended version, we provide a substantially revised paper describing the algorithm for real-time visualization of protein empty space supplied with functions for its further exploration. These additional features include the visual evaluation of detected voids, automatic detection and highlighting of the largest void (it requires the computation of the volume of voids) or several improvements of the implementation, which facilitate the fluent user interaction even with large macromolecules.

The biochemical research concentrates on the identification and characterization of structures and processes, which are present in all living cells. Investigation of these structures, their function, causes and effects, are performed on a daily basis in labs. This process is usually very time and resources demanding so the current trend in this field is based on searching for various in-silico predictive software tools, which enable to explore and evaluate the structure prior to in-vitro testing in the laboratory.

Proteins belong to the category of structures that are present in all living organisms so their importance is indisputable. Analysis of their inner arrangement and effects of such arrangement to their function form a basis of many biochemical disciplines, such as protein engineering. In this case, the protein is studied with respect to its function. The order of its amino acids in the polypeptidic chain and its folding determines the characteristics of given protein (see Figure 1).

Figure 1. Polypeptidic chain of DhaA haloalkane dehalogenase visualized using the Cartoon method.

Directed mutagenesis of some amino acids causes changes in the protein structure as well as in its behavior. Using this approach, various important characteristics of proteins can be influenced – such as their activity, stability or enantioselectivity [2][3].

The other approach, which is utilized in the drug design, is based on the transportation of small molecule of substrate (ligand, ion or solvent molecule) to the protein binding site. The binding site (also called an active site) is formed by a cavity, which is deeply buried inside the protein structure and it is a place where chemical reaction between protein and substrate undergo. The product of such reaction then forms a basis of some chemical compound, such as new medication.

We advocated that studying of protein inner structure is important in many areas of biochemical research. Moreover, theses studies can be performed using software tools, which are also able to visualize the protein structure along with results of the analysis. When analyzing the protein inner structure, biochemists concentrate on an empty space, which is present in the protein. The amount of the empty space determines also the size of substrates entering the active site or possible mutations. Moreover, the binding site is also defined by an empty space marked as a void (or more specific, a cavity).

Many algorithms that compute the empty spaces in proteins (often also marked as voids), have been published and the chemical relevance of computed results was proven by biochemists. However, not each void can play a role of a binding site. So it is necessary to enable the user also visual exploration of detected voids to better assess their importance. To fulfill these requirements, we developed a novel approach to the computation, visualization and further evaluation of voids in protein. The low complexity of the algorithm enables the user to perform all these steps in real-time, which substantially decreases time demands on the evaluation phase. Our user-driven approach enables to compute and display empty space that satisfies the input parameters, which influence the minimal width of voids. The values of these parameters can be changed anytime during the exploration and changes are immediately projected.

A. *Empty space classification*

The empty space present in protein structures can be further qualified according to various criteria and marked as a cavity, pocket, tunnel, channel, pore or other specific structure (see Figure 2). As mentioned above, inner cavities can play a role of binding sites and thus serve as the destination for a small ligand molecule that can follow a pathway from the outside environment of the protein. In such cavity the chemical reaction between protein and ligand can take place. Pockets are defined as a (concave, cleft, hole)-shaped region on the protein surface [4]. They are also of high importance because they can also serve as potential ligand binding sites. Thanks to their large binding surface area they are more easily accessible than the deeply buried binding sites.

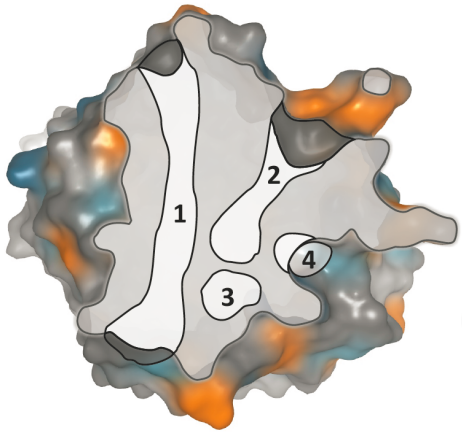


Figure 2. Illustration of channel (1), tunnel (2), cavity (3) and pocket (4).

Channels and pores are specific pathways crossing the whole protein. Channels can be used for the transport of substrates, products, water molecules and other compounds through the protein. Pores are present in transmembrane channel proteins and are essential in mediating the transport of ions and molecules through biological membranes [5].

Our long-term research in this field was concerned mainly with the detection of tunnels. These structures represent a path leading from a specific protein cavity (the binding or active site) to the molecular surface. Prior to tunnel detection it is crucial to determine the starting point for their computation – the active site. Thus it is necessary to analyze and evaluate protein cavities and assess that one containing the active site. Small substrate molecules entering the active site determine the limitation of properties of computed tunnels (e.g., their width or curvature).

The derivation of an empty space from the 3D structure of a protein's amino-acid sequence introduces a very complex task. Proteins with known three-dimensional structure are stored in the PDB database [6]. This archive contains information about experimentally-determined structures of proteins, nucleic acids, and complex assemblies. Some of the structures involved into the PDB database were analyzed and their binding sites were detected. These sites are stored in the CSA (Catalytic Site Atlas) database [7] or in the UniProt database [8]. However, active sites of most of the structures still have not been revealed or published. This situation creates the necessity of using other semi-automated tools or even manual detection of the active site.

B. *Cavities*

Our novel approach to the detection of cavities is able to identify all cavities in a given region (its size is defined by the *distance* parameter, which determines the radius of

a sphere of interest with user-defined center) or in the whole protein structure. First of all, all cavities inside the structure or region of interest are detected. However, when operating with large protein complexes or even ribosomes, computation of all cavities in such molecule is very time and memory consuming. Thus we introduce a novel method for detection and visualization of inner cavities focusing on minimizing the memory and time requirements. This technique is designed to operate in real-time, enabling users to interactively change the inner and outer size of a spherical probe utilized for detection of cavities.

Detection of cavities should be followed by their proper visualization. Otherwise, it is very complicated for biochemists to evaluate the resulting cavities. To be able to determine the cavity containing the active site properly, biochemists should be equipped with a powerful visualization tool enabling not only displaying of detected voids but also allowing real-time alternations of parameters of cavities. When combining this method with other chemical properties of given protein (such as the knowledge of partial charges of atoms), biochemists are able to recognize the ligand binding site easily. To be more specific, when the atoms surrounding the cavity have neutral or small partial charges, this cavity probably will not be marked as an active site.

A brief outline of the paper follows. Section II presents existing algorithms for detection and classification of empty spaces in protein structures, and common approaches to computation and visualization of protein surfaces. Section III gives a detailed description of the proposed real-time algorithm for visualization of protein voids. The next section, IV, discusses several possibilities of visual evaluation of detected voids. Section V concludes the paper with computational results and their analysis. Last section, VI, introduces some ideas for the future improvements of the proposed algorithm.

II. Related work

Detection and classification of the empty space inside proteins has been in the scope of biochemists for the last decades. Many algorithms have been proposed and published in this field. Although the aim of this article is to, above all, present a novel approach to detection and visualization of inner cavities, in this section we will introduce a related research focused on the detection of empty space in general. Techniques mentioned in this section can be mostly adapted to the computation of more specific structures, which were mentioned (tunnels, channels, etc.).

Algorithms detecting an empty space inside proteins are based on the similar principle – they are all based on computational geometry using the three-dimensional protein structure (positions and radii of atoms) as the input. These algorithms can be divided into two groups according to their approach to space representation:

- algorithms based on a grid approach
- algorithms utilizing Voronoi diagram and Delaunay triangulation

The main difference between these two approaches lies in their precision, speed and memory consumption. More detailed description of these approaches along with their representatives follow.

A. Detection of empty space

1) Grid method: In this approach, the entire protein is enclosed in an axis aligned bounding box, which is subsequently regularly sampled to a voxel grid. Each vertex of the voxel grid is classified according to its collision with an atom. Non-colliding voxels form the empty space used for construction of cavities, tunnels and other structures. The quality of results is strongly influenced by the sampling density. Too sparse sampling can lead to a situation where all vertices of voxels are colliding with an atom and no empty space is detected. On the other hand, too dense a sampling causes an enormous increase in time and memory demands. The main advantage of this approach is its simplicity; the disadvantage, as already mentioned, comes from its computational complexity $\mathcal{O}(n^3)$, with n depending on the sampling density.

The grid approach was adopted for tunnel computation in CAVER 1.0 [9]. Another tool using the grid approach for computation of specific cavities (pores) inside proteins is called CHUNNEL [10]. Each voxel is marked according to its distance to the nearest atom. On this structure, the Dijkstra algorithm is launched and the tunnel with highest voxel values (the widest tunnel) is detected.

Kleywegt et al. [11] presented their grid approach applied to the detection of cavities. Their implementation is presented in the VOIDOO application. The first step of the algorithm maps the protein onto a 3D grid with a spacing between 0.5 and 1.0 Ångströms. Each point of the grid is noted by the zero value. Then, each grid point is processed and when the distance to the nearest atom is less than the sum of the atom radius and the probe radius, its value is set to one. This method is also known as the flood-fill algorithm. Finally, points inside cavities still have a zero value so they can be easily detected and their volume can be measured.

2) Voronoi diagram and Delaunay triangulation: Another approach to protein 3D space inspection is based on the Voronoi diagram (VD) and its dual structure – the Delaunay triangulation (DT). The main benefit of this approach is based on the division of space, which is independent on any user-defined settings. That is the main reason why this method overcomes the main disadvantage of the previous grid approach. The detailed description of VD construction can be found, e.g., in [12]. The dual structure to VD, the Delaunay triangulation (tetrahedrization in the three-dimensional space), can be constructed by connecting neighboring points sharing the Voronoi edge (see Figure 6). Tetrahedra of the Delaunay tetrahedrization fulfills the

condition that no point is presented inside the circumsphere of any tetrahedron.

Voronoi diagrams and Delaunay tetrahedra were utilized by various software tools for tunnel and channel computation [13] (see Figure 3), such as CAVER 2.0 [14], MolAxis [15] or MOLE [16].

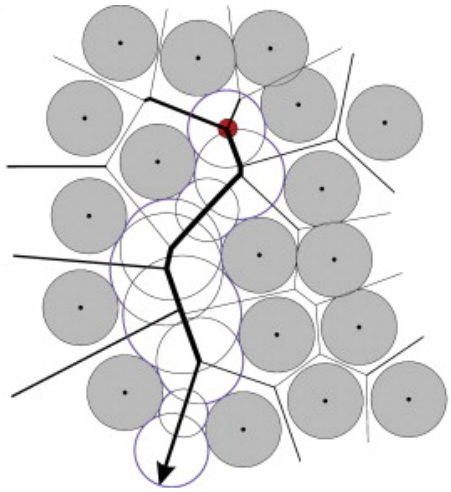


Figure 3. Tunnel detected using the Voronoi diagram (taken from [13]).

Another approach to cavity detection using the Delaunay triangulation and the alpha complex was implemented in the CAST application [17] (CASTp is its online version). This tool is able to measure the area and volume of cavities analytically as well.

In [18], Voronoi diagrams were extended to the Additively weighted Voronoi diagrams (AVD). AVDs were originally designed for environments containing non-uniform objects. This is also the case of protein structures because atoms of various chemical elements vary also in the size of spheres representing them. Compared to traditional VDs, AVDs gain the more adequate space subdivision through the specification of weight w attached to each site point. According to their weight values the respective points attract ($w > 1$) or repel ($w < 1$) the corresponding Voronoi edges. Resulting Voronoi edges have curvilinear shapes (see Figure 4). AVD construction is more complex in comparison to traditional VD and thus the time and space complexity increases substantially. AVD were used in the protein visualization tool called Voroprot [19].

The second part of the related work is connected with the protein outer environment and its differentiation from the inner part of the protein structure. The outer environment can be considered a void as well. To distinguish the inner environment from the outer one, the protein has to be encapsulated in some bounding object – a protein surface. Detection of voids inside protein structures is highly influenced by the protein surface, which gives an overview of the protein's compactness. It makes sense to detect empty

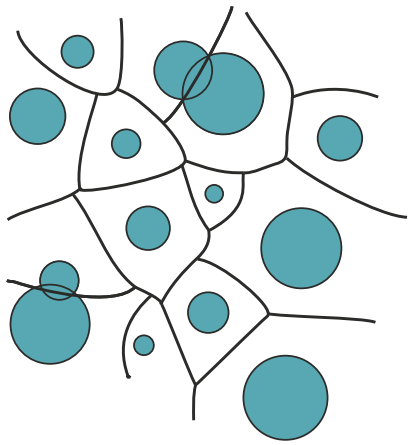


Figure 4. Additively weighted Voronoi diagram on a set of atoms.

space only within the volume that is defined by the protein surface. Therefore, in the next section, we concentrate on existing approaches to detection of protein surfaces.

B. Protein surface and its detection

Computation and visualization of surfaces play an important role not only in the case of detection of voids. The resulting shape of the surface also gives the biochemists a fair overview of protein constitution, presence of pockets and clefts etc (see Figure 5). Generating a smooth molecular surface is also important for a number of applications, including molecular recognition, drug design, electrostatics, molecular graphics, etc. Thus, protein surface detection has been in researchers' scope for decades and many approaches have been proposed. Two main groups of existing algorithms employ either analytical or numerical approach.

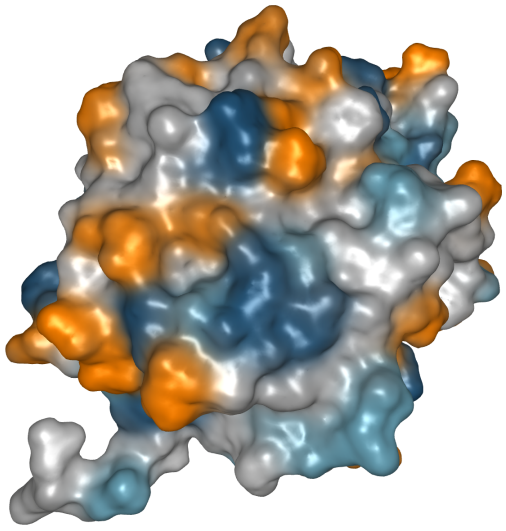


Figure 5. Visualization of molecular surface, which enables to observe the overall shape of the structure.

1) Analytical surface construction: The input set contains objects (atoms) that should be encapsulated by the surface. The analytical approach describes the surface using a set of mathematical equations. For protein exploration there are two basic analytical approaches to generation of surfaces. The Reduced surface [20] is constructed by rolling a probe of specific radius over the protein starting in the outer environment. Inwards facing parts of the probe surface combined with parts of atoms' surfaces on the boundary create the resulting solvent-accessible surface. The second approach is based on the alpha-shapes theory [21]. Totrov and Abagyan [22] introduced the contour-buildup algorithm for analytic calculation of molecular surface. The core part of this algorithm is based on the sequential build up of multi-arc contours on the van der Waals spheres representing individual atoms.

The main disadvantage of the analytical representation of the surface comes from its complexity. Thus its utilization on large datasets (e.g., macromolecular structures) cannot be performed in real-time or can even fail.

2) Numerical surface construction: The accuracy of numerically based algorithms is strongly dependent on initial user settings. The basic principle is the division of the scrutinized space into a uniform voxel grid. Each voxel is classified according to its intersection with objects in space. Subsequently, the marching cubes algorithm [23] can be utilized for visualization of the surface. The marching cubes method was designed primarily for a simple and fast construction of isosurfaces in volume data sets. This approach is widely used, e.g., in MRI or other medical applications. One of the representatives of this approach is the LSMS algorithm [24]. This algorithm is able to compute not only the protein surface – it can detect the interior cavities as well.

III. Real-time visualization of protein voids

In comparison with existing algorithms for visualization of empty voids, our novel approach does not require any additional time for their recomputation when the input parameters change. The empty space corresponding to these changes is visualized instantly. The original version of our approach was able to perform the real-time update of voids on small and middle size molecular systems. But when dealing with large macromolecules (e.g., ribosomes) we had to introduce several improvements. Current version of our algorithm handles also cases operating on macromolecules.

In the rest of this section, the basic principle of our algorithm along with current improvements will be described. The main aim of our approach is the real-time visualization of inner voids. Naturally, these voids must be computed in the first phase. For this task we utilize the standard Voronoi diagram, which omits the differences between radii of atoms (contrary to AVD approach) since our priority is the speed of the algorithm. Of course, VD does not provide users with as precise results as AVD does, but from our experience the difference is acceptable for purpose of cavities detection. The main difference between the results obtained by VD and AVD is in the exact representation of the surface of voids. However, the set of detected cavities is equal.

In the visualization phase, the algorithm is divided into five basic steps. Steps 1 and 2 represent preprocessing and they are performed only once during the initialization phase. Steps 3 to 5 (represented by subsections C to E) are iteratively repeated for any change of input parameters and are considered as one of the main contributions of this paper.

```
Algorithm Real-Time Visualization of Protein Voids

Input: set of atoms
1. compute Delaunay triangulation
2. convert it to a graph
3. while user is changing parameters do
4.      determine center point of the bounding box
5.      select empty space inside the protein
6.      visualize selected empty space
7. end while
```

A. Construction of Delaunay triangulation

input: set of atoms A
output: Delaunay triangulation T

The input set A consists of all atoms of the protein. Since we do not take into account the difference between atom radii, the atomic centers were selected as representatives of atoms. These atomic centers then form the input set of points marked as P, which is subsequently processed. For the set P, the Delaunay triangulation T is constructed using the QuickHull 4D algorithm described in [25].

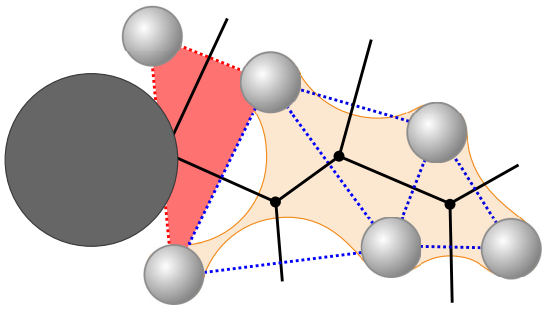


Figure 6. Tetrahedron (red) accessible by a probe (dark gray) is removed from the triangulation (blue dotted). The molecular surface defined by the probe is highlighted (orange).

The triangulation T is afterwards refined so that all tetrahedra intersecting the molecular surface of the protein are removed. In other words, all surface tetrahedra that are accessible from the outside by a probe with radius 2.8Å(double the van der Waals radius of oxygen) are removed from T (see Figure 6). This ensures that the tunnel throat will not contain any excessive boundary spheres.

B. Construction of the graph G

input: Delaunay triangulation T
output: evaluated graph G

For each tetrahedron $t_i \in T$ a node N_i is inserted into a newly constructed graph G. An edge e_{jk} connecting nodes N_j and N_k is added into G if their referenced tetrahedra t_j and t_k share a face f_{jk}. For every edge $e_{jk} \in G$, we define its center point $c(e_{jk})$ and width $w(e_{jk})$ as follows. The center point $c(e_{jk})$ is defined as a point in f_{jk} where sphere with maximal possible radius not intersecting any atom from t_j or t_k can be placed. The width $w(e_{jk})$ is then defined by the radius of such a sphere. The evaluation process is illustrated in Figure 7.

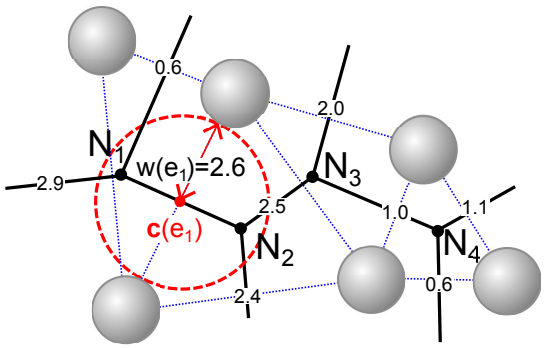


Figure 7. Illustration of a part of a graph G. Thick lines represent Voronoi edges. Every edge is evaluated by the value representing its distance to the nearest atom.

C. Selection of center point

input: Delaunay triangulation T
output: center point $\mathbf{C}$

The algorithm was designed to operate with large macromolecules. In this case, computation and visualization of all inner voids usually leads to complex and ambiguous results, which the biochemist cannot properly explore, thanks to the huge amount of visualized data. In order to avoid this situation, we allow computing inner voids from a starting point $\mathbf{C}$ that represents the center of the bounding sphere. The empty space is then visualized only inside this bounding sphere, which represents the area of interest. The point $\mathbf{C}$ set by the user can be determined in two ways. The user can enter the space coordinates of the point directly or, in most cases, the binding site loaded from the CSA (Catalytic Site Atlas) database [7] can define the point $\mathbf{C}$. Once the center point $\mathbf{C}$ is set, it can be stored for further iterations of the algorithm.

D. Selection of relevant edges

input: graph G, point $\mathbf{C}$, distance d, parameter w_{min}
output: set of filtered edges E

In this phase, the iteration process is started. The goal is to select a set E of edges from G, which satisfy the condition of thickness (driven by the parameter w_{min} representing the minimal width of the edge) and proximity (parameter d defining the bounding sphere radius). For remark, every edge e_{jk} connecting two nodes N_j and N_k is evaluated by a width $w(e_{jk})$.

The set of filtered edges E consists of all edges having the $w(e_{jk})$ greater or equal to w_{min} and with the distance to $\mathbf{C}$ lower than d. More formally, let G_E is the set of all edges from G. The set of filtered edges is then $E = \{e_{jk} \in G_E | dist(\mathbf{C}, c(e_{jk})) < d \wedge w_{min} \leq w(e_{jk})\}$.

E. Visualization

input: set of edges E, selected visualization method(s)

Firstly, the set E has to be transformed into geometrical objects, which are possible to render. Every edge e_{jk} is transformed into a sphere s_{jk} with center in $c(e_{jk})$ and with radius equal to $w(e_{jk})$. The set S of all such spheres is then prepared as an input for selected visualization method(s).

For our case of protein visualization, we utilized two basic methods effectively describing the empty space inside macromolecules.

1) Rendering of spheres: This option represents the most intuitive visualization method as well as the fastest one (see Figure 8). It displays all spheres of the set S. From the construction introduced above, all spheres fill the empty space inside the molecule and do not intersect with any atom. Using this method, the empty space is highlighted, but it looks distracting and for user it can sometimes be difficult to distinguish between an atom of the molecule and a sphere highlighting the empty space. Thus also another approach was introduced.

2) Grid sampling: This technique enables users to visualize a continuous surface of discrete voids, which gives more intuitive and user friendly results. To construct such surface, a grid based approach is performed. All spheres from the set S are enclosed into an axis aligned bounding-box. Afterwards, this bounding-box is regularly sampled with a user defined $density$. It is obvious that the higher density leads to more precise surface. On the other hand, the number of samples directly influences the memory and time complexity of the computation. We found out that for exploring of local neighborhood the empirically obtained $density = 200$ (i.e., grid 200x200x200) is optimal. Subsequently, each vertex of each cell in the grid is evaluated according to its intersection with any sphere from S. When all vertices are processed, the fully evaluated grid serves as the input for the marching cubes algorithm. For a notice, this algorithm operates with a predefined set of configurations, thus it its very straightforward and fast when constructing the resulting surface (see Figure 8).

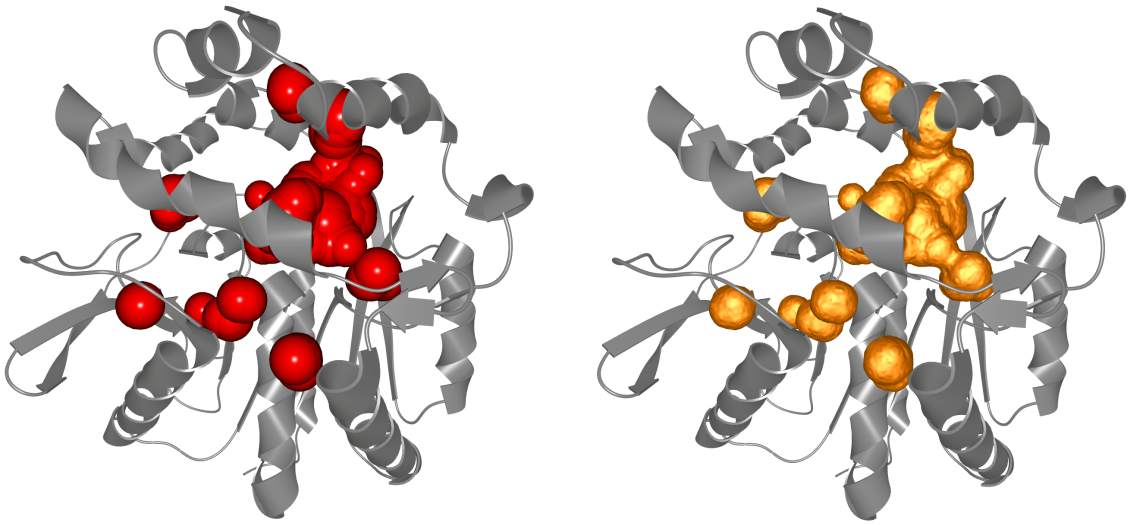


Figure 8. DhaA haloalkane dehalogenase with computed cavities (parameter settings: distance = 13.4 Å, minimal width = 1.5 Å). Empty space visualized as a set of spheres (left) or a surface (right).

IV. Visual Evaluation of Detected Voids

The main aim of the visualization methods for displaying computed voids is the subsequent evaluation of these voids. Without proper visualization it is very complicated to observe the size and shape of individual voids along with their surrounding amino acids. For better exploration of the cavity inner environment, we designed a clip plane with an arbitrary orientation – see Figure 9. It displays the largest cavity of the DhaA haloalkane dehalogenase, which corresponds to the binding site of this structure (it is specified by the cross with axes).

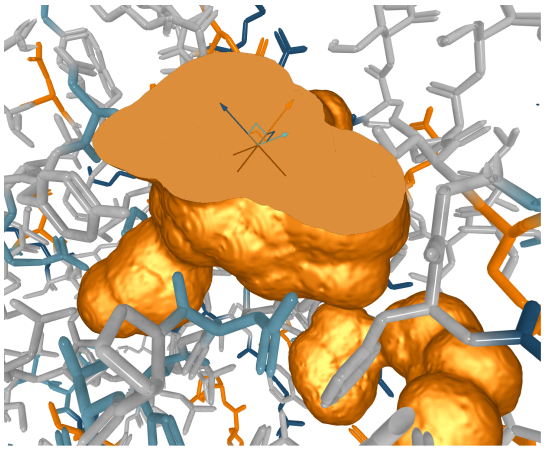


Figure 9. Largest cavity of DhaA haloalkane dehalogenase containing the active site is cut by a clip plane to explore the inner environment. The clip plane can be shifted in both directions in order to go through the whole cavity.

The fast recomputation and real-time observation of the impact of parameter changes also helps to explore voids with respect to changes of the input set. This means that users are able to include or exclude any atom from the computation and the results are instantly displayed to the user. This situation is illustrated in Figure 10. Molecules of water and ions of iodine, which are originally present in the PDB file of given structure are excluded from the computation of voids.

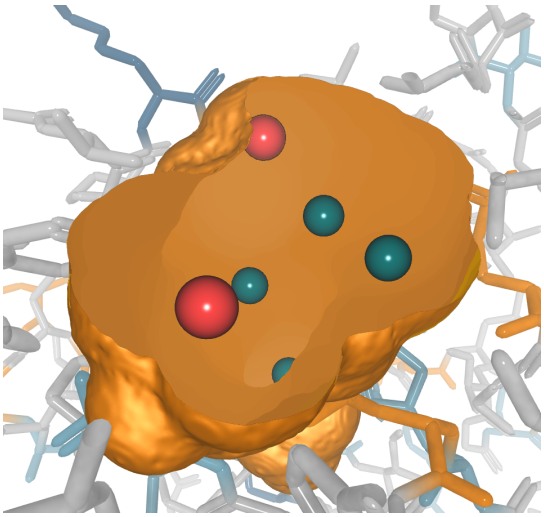


Figure 10. Cavity containing molecules of water (blue spheres) and red spheres representing ions of iodine.

Figure 11 illustrates the high importance of atoms included into the computation. The left part of the figure shows the largest cavity of the DhaA haloalkane dehalogenase. In this case only the amino acids forming the polypeptidic chain were taken into account. Blue spheres in the neighborhood of the cavity represent water molecules. The middle

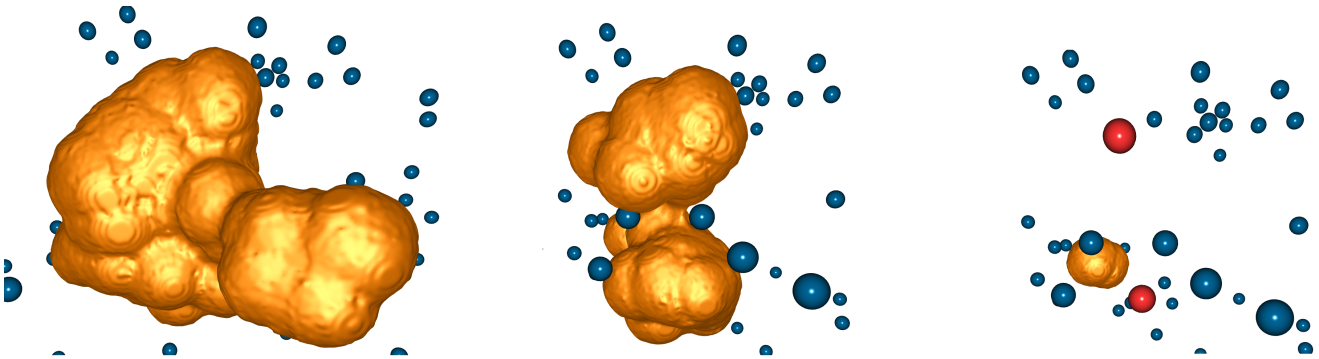


Figure 11. Largest cavity of DhaA haloalkane dehalogenase with different computation settings. Left - only protein amino acids are taken into account during computation. Blue spheres represent water molecules present in the PDB file of the structure. Middle - amino acids and water molecules (blue) present in the structure are considered. This causes reduction of the cavity. Right - amino acids, waters and ions of iodine (red) present in the structure are involved into computation. The cavity is even more contracted.

part displays the same cavity when the surrounding waters are considered during computation. Finally, in the right part the cavity almost disappeared because water molecules and also ions of iodine were taken into account.

The improved version of our algorithm is also able to compute the volume of detected cavities and according to this metric to determine and mark the largest cavity. Figure 12 shows the result of such evaluation when the largest cavity is highlighted using red color.

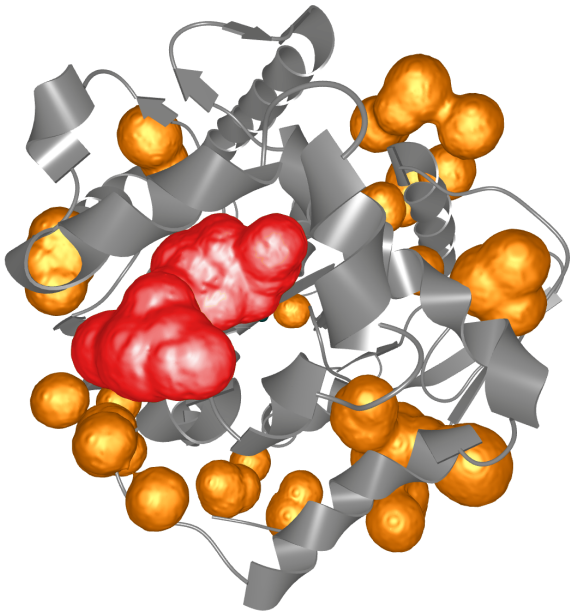


Figure 12. All cavities of the DhaA haloalkane dehalogenase with minimal width 1.9 Å. The largest cavity is detected and marked with red color.

V. Results and Discussion

In this paper, we presented a novel method for real-time visualization of empty space inside macromolecules, which concentrates on user-driven evaluation of computed voids. The method is not limited by the size of the molecule (the number of atoms) as the encapsulation of displayed voids into a bounding sphere allows to restrict the amount of processed data respectively. The current version is able to process also the whole macromolecule in real-time. Thanks to introducing adaptive refinement of the surface we are able to react on changing of parameters instantly because of visualization of the rough surface (with the low density of the sampling grid). When parameters are set and not changing, the algorithm immediately starts to calculate more precise surface. This calculation is performed in three levels of detail. So when the parameters remain the same for a few seconds, the smooth surface is generated.

The implementation has not any special hardware or software requirements, the algorithm was implemented in 32-bit Java environment. The performance was tested on a common single-threaded 2.66GHz computer. Both rendering strategies, as well as various types of macromolecules (ranging from proteins to ribosomes) underwent this test. Examples of tested protein structures and combinations of parameters are summarized in table in Figure 13. Results correspond to the first level of detail surface – the rough one.

To illustrate the robustness of our algorithm we selected four representatives of molecules containing common to large number of atoms.

To illustrate the importance of input parameters, Figure 14 shows how these parameters influence the resulting voids. Left part of the figure shows the protein structure where cavities in the area of interest defined by a sphere with radius 13.4 Å, where the minimal width (defined by the size of the inscribed sphere) is set to 1.14 Å. Right part of the figure shows a cavity detected in the same area of interest but the minimal width changed to 1.6 Å.

The relevance of computed cavities was tested on many protein structures with well known inner arrangement. To

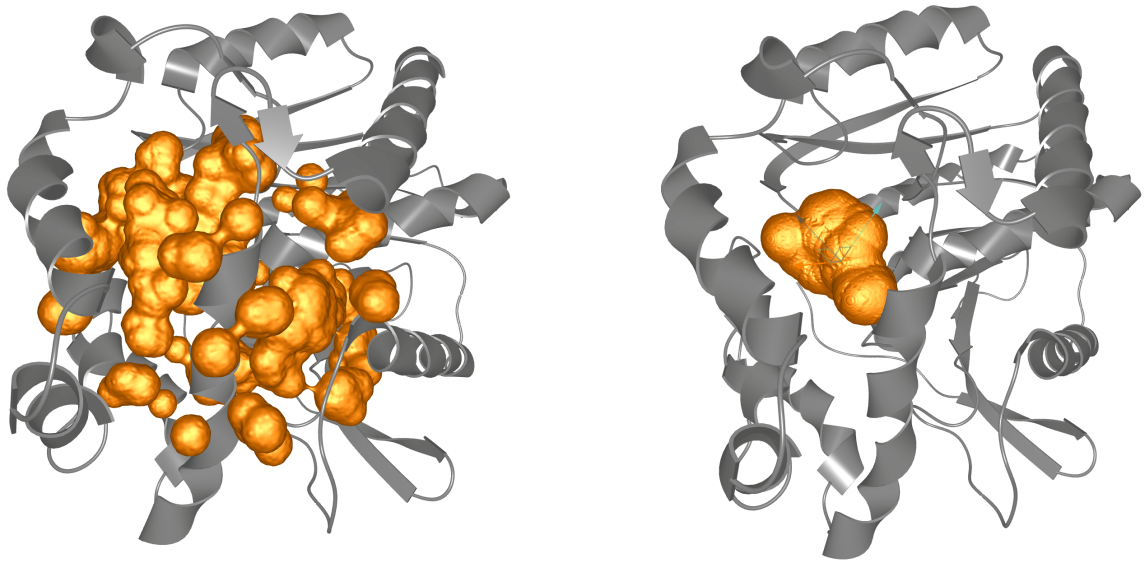


Figure 14. DhaA haloalkane dehalogenase with computed cavities. Left: All cavities detected in the region defined by a sphere with radius 13.4 Å, minimal width (size of the inscribed sphere) is set to 1.14 Å. Right: The largest cavity of minimal width 1.6 Å. This cavity contains the active site - marked with the cross with three axes.

Structure	d (Å)	$w(e)_{min}$ (Å)	spheres (FPS)	surface (FPS)
1cqw 2754 atoms	15	1.4	> 100	78
	max	1.4	> 100	62
1cqz 8218 atoms	15	1.4	> 100	77
	max	1.4	> 100	58
2oau 13580 atoms	15	1.4	> 100	36
	max	1.4	88	19
1aon 58874 atoms	15	1.4	73	16
	max	1.4	28	2

Figure 13. Performance of the algorithm measured in frames per second. Parameter d represents radius of the bounding sphere, $w(e)_{min}$ shows the minimal with of an edge.

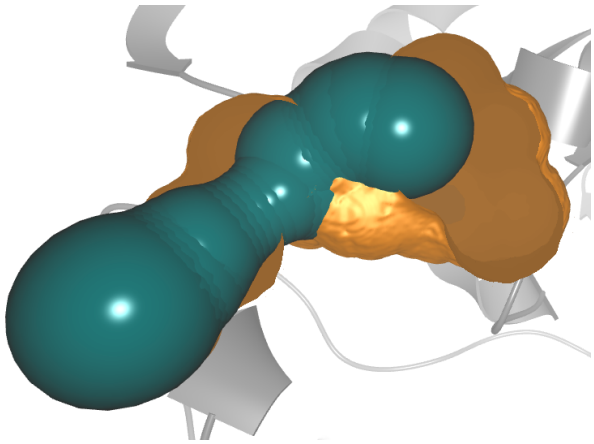


Figure 15. Cavity containing the binding site along with the computed tunnel.

evidence the relevance of empty space detected and visualized using our approach, we performed a comparison with results obtained by the well acknowledged CAVER algorithm. CAVER was designed for the detection of tunnels inside proteins and results were thoroughly tested by the community of protein engineers [26]. Thus, to manifest the relevance of voids detected by our new approach, the computed voids must contain all detected structures such as tunnels, cavities etc. We verified that tunnels detected by the CAVER algorithm lead through the empty space highlighted by our method (see Figure 15).

Figure 16 shows more results of our algorithm – cavities computed on structures with PDB IS's 2OAU and 1M4X.

VI. Future

Further extension of our implementation should lead to the parallelization of the marching cubes algorithm on the modern graphic cards [27]. Such implementation would substantially increase the performance of the rendering phase. Moreover, to obtain even more precise results, the additively weighted Voronoi diagrams should be utilized.

Another step towards the better exploration of empty voids in proteins can involve implementing more variants of shapes of area of interest. Users will not be limited only by a bounding sphere but could create e.g., slices of user-defined width. More importantly, they will be able to explore the local neighborhood of various inner structures, such as tunnels, channels or pores.

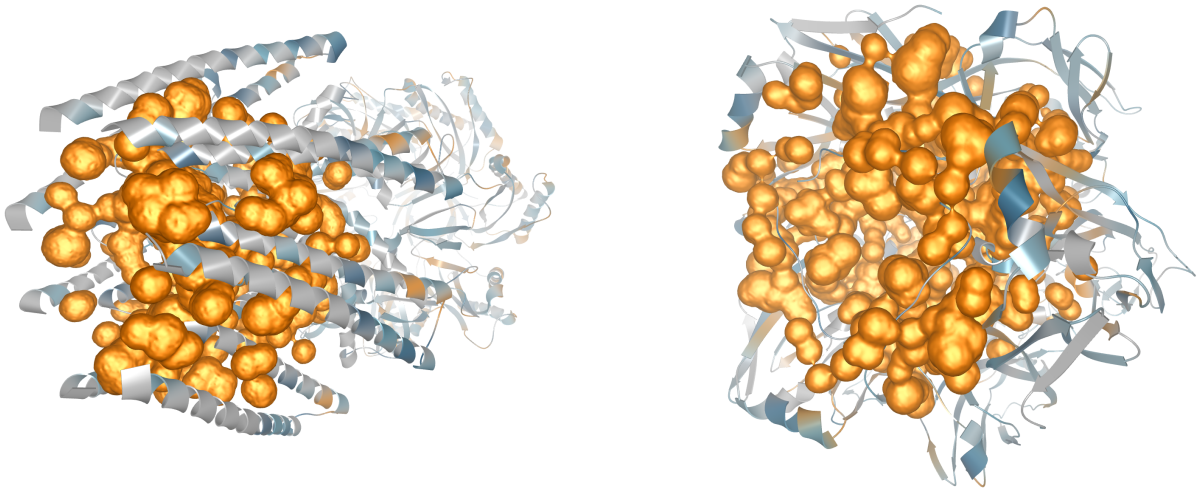


Figure 16. Resulting cavities computed on different structures. Left: Protein structure with PDB ID 2OAU. Right: Structure marked with PDB ID 1M4X.

ACKNOWLEDGMENT

This work was supported by The Grant Agency of the Czech Republic, Contract No. P202/10/1435.

REFERENCES

[1] O.Strnad, V. Šustr, B. Kozlíková, and J. Sochor, "Real-time visualization of protein empty space with varying parameters," *Proceedings of BIOTECHNO, IARIA XPS Press*, pp. 65–70, March 2013.

[2] M. Pavlova, M. Klvana, Z. Prokop, R. Chaloupkova, P. Banas, M. Otyepka *et al.*, "Redesigning dehalogenase access tunnels as a strategy for degrading an anthropogenic substrate," *Nature Chemical Biology*, vol. 5, pp. 727–733, 2009.

[3] T. Koudelakova, R. Chaloupkova, J. Brezovsky, Z. Prokop, E. Sebestova, M. Hesseler *et al.*, "Engineering enzyme stability and resistance to an organic cosolvent by modification of residues in the access tunnel," *Angewandte Chemie International Edition*, vol. 52, pp. 1959–1963, 2013.

[4] T. Kawabata and N. Go, "Detection of pockets on protein surfaces using small and large probe spheres to find putative ligand binding sites," *PROTEINS: Structure, Function, and Bioinformatics*, vol. 68, pp. 516–529, 2007.

[5] M. Pellegrini-Calace, T. Maiwald, and J. M. Thornton, "Porewalker: A novel tool for the identification and characterization of channels in transmembrane proteins from their three-dimensional structure," *PLoS Computational Biology*, vol. 5, no. 7, 2009.

[6] F. C. Bernstein, T. F. Koetzle, G. J. Williams, E. E. M. Jr., M. D. Brice, J. R. Rodgers *et al.*, "The protein data bank: A computer-based archival file for macromolecular structures," *Journal of Molecular Biology*, vol. 112, no. 535, 1977.

[7] C. T. Porter, G. J. Bartlett, and J. M. Thornton, "The catalytic site atlas: a resource of catalytic sites and residues identified in enzymes using structural data," *Nucleic Acids Research*, vol. 32, no. Database-Issue, pp. 129–133, 2004.

[8] The UniProt Consortium, "Reorganizing the protein space at the universal protein resource (uniprot)," *Nucleic Acids Research*, vol. doi: 10.1093/nar/gkr981, 2011.

[9] M. Petřek, M. Otyepka, P. Banáš, P. Košinová, J. Koča, and J. Damborský, "Caver: A new tool to explore routes from protein clefts, pockets and cavities," *BMC Bioinformatics*, vol. 7, p. 316, 2006.

[10] R. G. Coleman and K. A. Sharp, "Finding and characterizing tunnels in macromolecules with application to ion channels and pores," *Biophysical Journal*, vol. 96, no. 2, pp. 632 – 645, 2009.

[11] G. J. Kleywegt and T. A. Jones, "Detection, delineation, measurement and display of cavities in macromolecular structures," *Acta Crystallographica Section D*, vol. 50, no. 2, pp. 178–185, Mar 1994. [Online]. Available: http://dx.doi.org/10.1107/S0907444993011333

[12] F. Aurenhammer, "Voronoi diagrams a survey of a fundamental geometric data structure," *ACM Comput. Surv.*, vol. 23, no. 3, pp. 345–405, sep 1991. [Online]. Available: http://doi.acm.org/10.1145/116873.116880

[13] J. Brezovsky, E. Chovancova, A. Gora, A. Pavelka, L. Biedermannova, and J. Damborsky, "Software tools for identification, visualization and analysis of protein tunnels and channels," *Biotechnology Advances*, vol. 31, pp. 38–49, 2013.

[14] P. Medek, P. Beneš, and J. Sochor, "Computation of tunnels in protein molecules using delaunay triangulation," *Journal of WSCG*, vol. 15(1-3), pp. 107–114, 2007.

[15] E. Yaffe, D. Fishelovitch, H. J. Wolfson, D. Halperin, and R. Nussinov, "MolAxis: efficient and accurate identification of channels in macromolecules." *Proteins*, vol. 73, no. 1, pp. 72–86, oct 2008. [Online]. Available: http://dx.doi.org/10.1002/prot.22052

[16] M. Petřek, P. Košinová, J. Koča, and M. Otyepka, "Mole: a voronoi diagram-based explorer of molecular channels, pores, and tunnels," *Structure*, vol. 15, no. 11, pp. 1357 – 1363, 2007.

[17] J. Liang, H. Edelsbrunner, and C. Woodward, "Anatomy of protein pockets and cavities: Measurement of binding site geometry and implications for ligand design," *Protein science : a publication of the Protein Society*, vol. 7, pp. 1884–1897, sep 1998.

[18] H. Edelsbrunner, *Algorithms in combinatorial geometry*. New York, NY, USA: Springer-Verlag New York, Inc., 1987.

[19] K. Olechnovič, M. Margelevičius, and C. Venclovas, "Voroprot," *Bioinformatics*, vol. 27, no. 5, pp. 723–724, mar 2011. [Online]. Available: http://dx.doi.org/10.1093/bioinformatics/btq720

[20] M. F. Sanner, A. J. Olson, and J. C. Spehner, "Reduced surface: an efficient way to compute molecular surfaces," *Biopolymers*, vol. 38, pp. 305–320, Mar 1996.

[21] H. Edelsbrunner, D. G. Kirkpatrick, and R. Seidel, "On the shape of a set of points in the plane," *IEEE Transactions on Information Theory*, vol. 29, pp. 551–559, jul 1983.

[22] M. Totrov and R. Abagyan, "The contour-buildup algorithm to calculate the analytical molecular surface," *J Struct Biol*, no. 1, pp. 138–43, 1995.

[23] W. E. Lorensen and H. E. Cline, "Marching cubes: A high resolution 3d surface construction algorithm," *SIGGRAPH Comput. Graph.*, vol. 21, no. 4, pp. 163–169, Aug. 1987. [Online]. Available: http://doi.acm.org/10.1145/37402.37422

[24] T. Can, C. I. Chen, and Y. F. Wang, "Efficient molecular surface generation using level-set methods," *Journal of Molecular Graphics & Modelling*, vol. 25, pp. 442–454, 2006.

[25] C. B. Barber, D. P. Dobkin, and H. Huhdanpaa, "The quickhull algorithm for convex hulls," *ACM Trans. Math. Softw.*, vol. 22, no. 4, pp. 469–483, dec 1996. [Online]. Available: http://doi.acm.org/10.1145/235815.235821

[26] E. Chovancová, A. Pavelka, P. Beneš, O. Strnad, J. Brezovský, B. Kozlíková *et al.*, "Caver 3.0: A tool for the analysis of transport pathways in dynamic protein structures," *PLoS Comput Biol*, vol. 8, no. 10, p. e1002708, 2012. [Online]. Available: http://dx.doi.org/10.1371/journal.pcbi.1002708

[27] C. Dyken, G. Ziegler, C. Theobalt, and H. P. Seidel, "High-speed marching cubes using histopyramids," *Computer Graphics Forum*, vol. 27, no. 8, pp. 2028–2039, 2008. [Online]. Available: http://dx.doi.org/10.1111/j.1467-8659.2008.01182.x

One Health Information and Communication Technologies

How digital humanities contribute to public health

Hans C. Ossebaard
RIVM - National Institute for Public Health and the Environment, Bilthoven, The Netherlands;
University of Twente, Enschede, The Netherlands.
hans.ossebaard@rivm.nl

Abstract **- Zoonoses are infectious diseases that are transmitted either directly or indirectly from animals to human beings. The human and economic costs of zoonoses and antimicrobial resistance can hardly be overestimated. Due to ecological and socioeconomic behavioral changes, new and different zoonoses emerge while antibiotics' effectiveness decreases. Early warning and surveillance systems are part of the public health response. However, a more pre-emptive approach is needed. 'One Health' entails a global strategy to advance interdisciplinary collaboration and communications in all aspects of health care for humans, animals and the environment. As of yet, the importance of social sciences and eHealth technologies for infectious disease research and public health policy is underestimated. The digital humanities provide methods and concepts that can set the One Health approach to work.**

Keywords - One Health; information and communication technology; zoonoses; public health; digital humanities

I. INTRODUCTION

The human and the animal worlds have profound and multifarious interactions that lead to a wide range of psychological, social, medical, economical and agricultural benefits. However, there are some severe drawbacks as well, for instance, with regard to the transmission of infectious diseases from vertebrate animals to human beings (zoonoses). The pathogenic micro-organisms - such as bacteria, viruses, fungi or prions - we share with domesticated or wild animals, have caused some of the most significant disease outbreaks in recent years, including HIV, Ebola, avian influenza, Q-fever, H1N1 flu, SARS and more recently MERS-CoV. Over 200 zoonotic diseases, often species-specific, have been identified. They have a serious and rising impact on global public health [1], and accordingly, receive scientific attention [2].

Outbreaks of zoonoses are typically hard to predict because of the complex and ever changing nature of the risk factors involved. At the same time, antimicrobial resistance is increasing and the dawn of the post-antibiotic era is a matter of global concern [3]. The human and economic costs of zoonoses and antimicrobial resistance are substantial. It is estimated that 60% of all human diseases and 75% of all emerging infectious diseases is caused by zoonotic pathogens over the past six decades [4]. The rate at which these have appeared in people has increased over the past 40 years, with at least 43 newly identified outbreaks since 2004. The immediate costs of zoonotic diseases over the last decade have been estimated to be more than $20 billion, with over $200 billion indirect losses to affected economies as a whole [5].

In the Netherlands, at least 19 people died as a consequence of the Q-fever outbreak (2007-2010), while the societal costs have been estimated between EUR 161-366 million [6]. A decade earlier in the UK the 1996 BSE outbreak caused at least 174 deaths (2010) and led to economic costs of at least EUR 1.1-1.4 billion [6]. But, by far the heaviest burden from human-animal diseases is carried by the poorest people. In 2012, the International Livestock Research Institute published a review study analyzing some 1000 disease-surveys covering ten million people and six million animals [7]. The authors conclude that the thirteen most important zoonoses together cause 2.4 billion cases of human illness and 2.2 million deaths each year, mostly in low- and middle-income nations. Approximately 75% of the economic and health damage impacts on only nineteen countries (e.g., Tanzania, India, Togo, Nigeria, Ethiopia) were the density of people and animals create ripe conditions for zoonotic diseases to arise and spread among populations of poor livestock keepers. In Asia and Africa, at least 55000 people recently died of rabies, according to the World Health Organization [8]. Expenses related to the prevention and control of rabies are estimated at US$ 590 million annually on both two continents. Disability-adjusted life years (DALYs) and monetary losses resulting from diseases such as human and livestock cystic echinococcosis (hydatid disease) have been calculated at the global level - assuming substantial under-reporting. The global human burden of echinococcosis may be as high as one million DALYs - or an annual loss of US$ 764 million. A maximum annual livestock production loss is estimated to amount to some US$ 2.2 billion. More figures and estimations on the health and economic impact of zoonoses can be found at the WHO website on 'neglected' zoonotic diseases [8].

Over the last decades the industrial countries have succeeded to eliminate, reduce or control zoonotic diseases

through substantial investments in public health: preventive measures, health education, feed bans, animal vaccination programmes, sanitation, food controls, culling sick animals etc. Nonetheless, the developed world - especially the US, the UK and Australia - now includes the key hot spots for emerging zoonoses (see Fig. 1). In low resource countries, such investments are not yet feasible in operational, legal, cultural or financial terms. Zoonoses present a threat to both human and animal health. Because of the relationship between development and population health the changing course of zoonotic diseases has led to increased attention in public health policy and research.

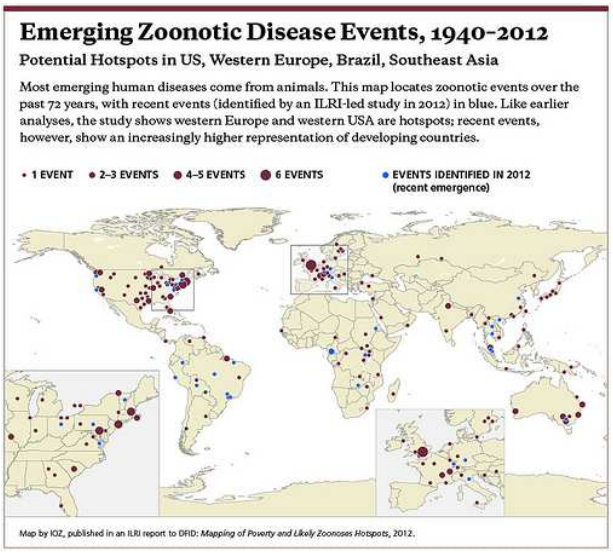


Figure 1. Emerging zoonotic disease events 1940-2012 (source: International Livestock Research Institute, 2012).

It is generally assumed that new and different zoonotic diseases will continue to develop because of

- a worldwide growth of people and animals (population density, cattle density);
- increased international (illicit) traffic of people, animals and products (globalization, eco-tourism);
- cultural changes (urbanization, migration, war, agro-industrial developments, farming new animal species);
- ecological changes (climate, environment, biodiversity).

Quite a few zoonoses are caused by RNA (Ribonucleic acid) viruses, which have high mutation rates and are extremely able to accommodate to changing circumstances. All this has raised global awareness among national, international, intergovernmental bodies and the redirection of resources towards prevention and control of zoonoses. This has led to the establishment of new, international expert networks, and surveillance and early warning systems such as the global network for animal disease research STAR-IDAZ (star-idaz.net), the European Emerging and Major Infectious Diseases of Livestock EMIDA (emida.era.net), the European Network for Diagnostics of 'Imported' Viral Diseases ENIVD (enivd.net/) or Discontools (discontools.eu), a joint initiative of industry and a wide range of stakeholders and several others. In the following sections it is maintained that the 'One Health' strategy to counter the threat of zoonoses can only be effective when the social sciences, the humanities and digital technologies are systematically taken into account.

II. ONE HEALTH

In the course of last century, the age-old shamanist wisdom that human (mental) health, animal health and environmental health are inextricably connected has been revitalized. Long before the well-known 'father' of veterinary epidemiology, the American Calvin W. Schwabe (1927-2006) proposed and promoted the term 'One Medicine' in his textbook "Veterinary Medicine and Human Health" (1984) the German 'founder' of social medicine physician Rudolf Virchow (1821-1902) coined the term 'zoonosis', stating that "... between animal and human medicine there are no dividing lines – nor should there be" [9]. This implies that the scientific foundations between the two do not really differ and that they share the same paradigm. Contemporary insights from human medicine, the natural sciences and veterinary disciplines also indicate that human, plant and animal systems are all part of the shared, planetary eco-system. Biologically speaking, both domesticated animals and wildlife are to be considered as close relatives that possess the same capacities to transfer infectious microorganisms. "We should therefore treat our relationship with other animal species as part of a continuum across which pathogens can emerge and spread, exploiting new niches as we change our interactions, and moving into and out of erstwhile distinct species, regions or communities" [10].

During the last decade, the concept of 'One Medicine One Health' evolved, urged by public concern after the 2003 outbreaks of SARS and Ebola hemorrhagic fever. It strives to advance scientific breakthroughs in an integrative and collaborative way [11]. In 2004 a series of conferences followed, addressing the theme 'One World – One Health' to underscore the health links between people, wildlife and the environment. In 2007/2008 the American Veterinary Medical Association (AVMA) and the American Medical Association (AMA) passed similar 'One Health' policy resolutions and took joint initiatives for action and collaboration. This resulted among others in the One Health commission, in summits and partnerships, targeting "the establishment of closer professional interactions, collaborations, and educational and research opportunities across the health sciences professions, together with their related disciplines, to improve the health of people, animals, plants and our environment" [12].

Today, the integrative, holistic concept of 'One Health' entails a positive, global strategy for expanding interdisciplinary collaborations and communications in all aspects of health care for humans, animals and the environment. It has been disseminated beyond the North Americas. "The synergism achieved will advance health care for the 21st century and beyond by accelerating biomedical research discoveries, enhancing public health efficacy, expeditiously expanding the scientific knowledge base, and improving medical education and clinical care" [13]. The visionary concept inspired many initiatives to improve collaboration between the often-segregated disciplines in policy, education and research. Strengthening the knowledge infrastructure through transdisciplinary cross-fertilization has become an important motivation, e.g., in European public health and policy projects. Since 2008, the EU has promoted the One Health approach, and it has been integrated into certain EU strategy documents [14]. The European Network of Excellence for Zoonoses research (medvetnet.org/) organizes thematic conferences on this subject, international One Health conferences take place (e.g., http://www.onehealthglobal.net/) and many others adopted the concept. One Health is promoted by scientists all over the world and supported by national organizations such as U.S. Centers for Disease Control or the Dutch National Institute for Public Health and the Environment (RIVM) as well as global organizations such as the United Nations (UN), the World Health Organization (WHO), the Food and Agriculture Organization (FAO), the World Organization for Animal Health (OIE), the International Federation for Animal Health and others. WHO, FAO and OIE signed a tripartite collaboration agreement to better coordinate their global activities at the animal-human-ecosystem interfaces [15]. Since 2011 an international open access, peer-reviewed 'One Health Journal' *Infection Ecology & Epidemiology* is published by a Swedish group of scientists to stimulate interdisciplinary collaboration in One Health [16]. In 2012 the World Bank published a report on cost/benefit analyses that accentuate the importance of the One Health approach for prevention and control of infectious diseases and recommended wider implementation [17].

In quite a short time, the concept has drawn massive attention and support. It is "intended to be all-inclusive among the scientific disciplines of human medicine, veterinary medicine and all other related scientific health disciplines" [18]. Therefore, it is remarkable that behavioral sciences, the humanities and technology are most seldomly explicitly mentioned in major documents on 'One Health'.

III. DIGITAL HUMANITIES

'Digital humanities' is a recent term that evolved from an increasing demand to understand and study how digital media interact with human experience and daily life [19]. It denotes an interdisciplinary, academic domain of education, research and practice where digital methods and media are used to study traditional topics in language, history, art, philosophy, communication and cultural studies. Moreover, the 'digital revolution' enabled a range of new behaviors, social contexts and objects, ranging from, e.g., cyber bullying or virtual environments, to web-based recommendations systems or big data, that may be studied from the digital humanities' perspectives. The opportunities of digital media for scholarship, research, education, presentation and cooperation also belong to this area. Like in any new academic field of interest, the subject of defining the scope remains a matter of academic debate. According to the UCLA Center for Digital Humanities "Digital Humanities interprets the cultural and social impact of new media and information technologies - the fundamental components of the new information age - as well as creates and applies these technologies to answer cultural, social, historical, and philological questions, both those traditionally conceived and those only enabled by new technologies" [20]. In line with this definition, we consider the social and behavioral sciences as belonging to the same academic practice and paradigm.

Evidently, the academic exchange between social sciences and computational sciences could benefit all domains of society. Rogers et al. [21] for instance develop methods to study the dynamics of internet censorship through the national web of Iran, or the workings of search engines, query logs and social networks. Others use mapping and advanced visualization techniques to study the role of animals in 19th American century cities [22], plot patterns of intellectual or creative exchange in the early-modern world [23] or operate digital European language repositories [24]. In the biomedical and health sciences the digital humanities gave rise to concepts such as eHealth, Medicine 2.0 or participative healthcare that address the use of information and communication technologies to support and improve health, health care and medical research. The social and participative opportunities enabled by digital media are widely used for the benefit of many [25]. Researchers use the internet as a tool and a source for studying human social behavior. Since 2006, for instance, the use of queries and social media networks have been studied to inform disease surveillance and early warning of infectious diseases [26].

IV. NEW APPROACHES

While the cooperation between the veterinary, botanical, entomological and human domains is now widely encouraged, the involvement of the humanities, i.c., the social and behavioral sciences is conspicuously absent. The ongoing concentration on the pathogen is understandable because of the historical effort - and successes - of the bio-medical professions to curb infectious diseases. It has obviously been the most likely approach. From a contemporary One Health-perspective however, this focus

should be widened as soon as possible. We believe that the humanities and the social sciences could play a decisive role when it comes to operationalizing the high-flown 'One Health' concept on the ground. Especially since they are so happily and prolifically engaged with information and communication technologies in what is sometimes called 'ePublic Health' or 'Public Health 2.0' [27]. At least three reasons illustrate why One Health should be completed with methods and concepts from the digital humanities and the social sciences.

The first is that the threat of (re)emerging zoonoses combined with our limited arsenal to protect public health, simply demand us to combine efforts. Faced with today's challenges we need to better understand what people actually do - not just what microorganisms do. They travel within and between their human hosts who participate in complex social networks. At the end of the day, people's behaviors make the difference when it comes to the transmission of disease-causing microorganisms. People create the conditions for the transmission of novel and re-emerging zoonoses but they also build the conditions that reduce their incidence, and prevent suffering, illness and death. Epidemiological data evidently need a human context to be meaningfully interpreted and put to use. Without the social sciences, the complex interactions of factors and circumstances that determine novel zoonotic disease spill-over get lost because they simply can't be understood. Social behavior is an essential ingredient of R_0, the basic reproduction number of an infectious disease. In Table I the social and behavioral factors associated with all three components of R_0 are summarized, after Janes et al. [28]. This knowledge on the social nature of transmission dynamics and possible modes of transmission informs the design of effective public health interventions and tailored risk communication on zoonoses.

TABLE I. SOCIAL FACTORS AFFECTING ZOONOTIC TRANSMISSION MECHANISMS

		Social and behavioral factors
R_0 basic reproduction number	***Exposure rate***	Social relationships, value system, ethics, rituals, habits, agricultural practice, population density
	Probability of transmission	Poverty, stress, health disparities, level of public health services, density of livestock, housing, sanitation
	Duration of infection	Social inequality, stress, access to care

The second reason to appreciate interdisciplinary collaboration with the humanities and the social sciences is their proven effectiveness in public health practice. Travel vaccination campaigns, personal hygiene, sanitary control, 'test and slaughter', protective clothing, changing life style choices (tobacco, fat, alcohol, sex, drugs), school-based prevention and education are just a few examples of effective social and behavioral measures [29]. Social scientific methodologies could be extended for use in the field of applied infectious disease research. Interesting examples can be found for instance in studies that model the impact of individual behavior on the spread of infectious diseases [30] or in qualitative research on the behavioral defenses and social psychological mechanisms through which people protect themselves against pathogens [31].

Thirdly, the reach and social impact of public health interventions, *when extended and connected with the opportunities offered by information and communication technologies*, is considerably enlarged as compared to biomedical approaches alone [32]. This is demonstrated for instance in the practice of the Antibiotic Stewardship Program of Eursafety Health-net. This cross-border German-Dutch collaboration to prevent and control health care associated infections such as MRSA[1] has been very successful to reach a significant reduction of nosocomial MRSA cases [33]. The use of a web-based tools for both patients and professionals has been important for the protection of patient safety and quality of care. Information and communication technologies [34] potentially empower them to take control over their personal and professional lives. They support for instance adherence to medical guidelines and protocols for hygiene or prudent prescription of antibiotics, which are essential to reduce health care associated infections. Self-management is one of the main promises of Health 2.0 and other innovations in health care [35, 36]. This matches seamlessly with currents attempts to re-define health in terms of peoples' positive capabilities to cope with disease and misfortune [37]. Technology supports this aspiration, which should be extended to the One Health domain. Another example is the design and development of an eHealth intervention to prevent tick bites and Lyme disease infections among green professionals and outdoor people currently undertaken by Beaujean et al. [38]. Although the incidence and prevalence of people presenting to the GP with tick bites or the first signs of Lyme disease has increased substantially, compliance to guidelines for prevention is generally low. A mobile app is developed to address the motivations of people at risk with tailored messages to eventually change this behavior.

The incredible growth in the use, availability, accessibility of these digital technologies - their speed, range and impact - also provides us with new opportunities for disease monitoring, surveillance and research [39]. These allow us

[1] *Methicillin-resistant Staphylococcus aureus* (*MRSA*) is a bacterium resistant against most antibiotics. It belongs globally to the most frequent causes of difficult-to-treat nosocomial infections in humans.

not only to inform a targeted public health response but also to act preemptively. Recently, scholars have called for an investment in fields such as digital humanities and computational social sciences, to using 'big data', the kind of digital information made available by improved data management, advanced mathematical analysis tools and increasing storage capacity [40]. Traditionally social or cultural data were collected via social scientific methods based on data scarcity, demands for formally controlled designs and financial considerations: field studies, user panels, focus groups, interviews, questionnaires and surveys. However today, the social and cultural interactions passing through the internet are considered as fast, valuable and relatively cheap sources of data for social and cultural research [27, 41]. This is obviously of great importance for the early warning, prevention and control of zoonoses. In fact, this new area of study with new methodological issues in risk communication, prevention and big data is the Number One challenge for sustainable and credible One Health solutions.

What is needed to effectively integrate the digital humanities and the social sciences into One Health? A good part of this may simply be translational and relational. Like in art and in music, an open mind is a condition sine qua non for cooperation. Social scientists, natural scientists, and medical practitioners need to be able to confer, to communicate and join forces now the global momentum is there. We hope to co-create one such opportunity at next year's 3rd international One Health Congress in the Netherlands [42].

ACKNOWLEDGMENT

H.C.O. thanks Desirée Beaujean MA, MSc, Head of the Department of Guideline Development and Implementation, National Coordination Centre for Outbreak Management, RIVM Centre for Infectious Disease Control for her support.

REFERENCES

[1] H.C. Ossebaard, "Supporting One Health by Technology" presented at the 2nd International Conference on Global Health, Challenges GLOBAL HEALTH 2013, Nov 17-22, 2013, Lisbon, Portugal. See: http://www.iaria.org/conferences2013/ProgramGLOBALHEALTH13.html.

[2] http://www.thelancet.com/series/zoonoses Accessed June 15, 2013.

[3] USDA/FSIS (U.S. Department of Agriculture/Food Safety and Inspection Service) and U.S. Environmental Protection Agency (EPA), "Microbial Risk assessment Guideline: Pathogenic Organisms with Focus on Food and Water", 2012 FSIS Pub. No. USDA/FSIS/2012-001; EPA Pub. No. EPA/100/J12/001.

[4] Institute of Medicine, "Sustaining Global Surveillance and Response to Emerging Zoonotic Diseases", Washington, DC: National Research Council, 2009.

[5] World Bank: People, Pathogens and Our Plant, Vol 1: Towards a Once Health Approach for Controlling. Zoonotic Diseases Report 50833-GLB, 2010.

[6] C. Tempelman, J. Prins and C. Koopmans, Economic consequences of the Q-fever outbreak [Economische gevolgen van de uitbraak van Q-koorts], SEO-rapport nr. 2011-15. Amsterdam: SEO, 2011.

[7] D. Grace, F. Mutua, P. Ochungo, R. Kruska, K. Jones, L. Brierley, L. Lapar, M. Said, M. Herrero, P.M. Phuc, N.B. Thao, I. Akuku and F. Ogutu, Mapping of poverty and likely zoonoses hotspots. Zoonoses Project 4. Report to the UK Department for International Development. Nairobi, Kenya: ILRI, 2012. Available at http://mahider.ilri.org/handle/10568/21161).

[8] http://www.who.int/neglected_diseases/zoonoses/en/

[9] P.H. Kass, R.H. McCapes, W.R. Pritchard, "In Memoriam. Calvin W. Schwabe. Professor Emeritus of Veterinary Epidemiology. Davis. 1927-2006. http://www.universityofcalifornia.edu/senate/inmemoriam/calvinwschwabe.htm. Accessed Sept 21, 2013.

[10] J. Zinsstag, J.S. Mackenzie, M. Jeggo, D.L. Heymann, J.A. Patz and P. Daszak "Mainstreaming One Health" Ecohealth. June; 9(2) 2012, pp. 107–111, doi: 10.1007/s10393-012-0772-8

[11] B. Kaplan et al. "The brewing storm. Monograph about One Medicine - One Health concept" 2009, http://www.izs.it/vet_italiana/2009/45_1/9.pdf

[12] https://www.onehealthcommission.org/

[13] B. Kaplan and M. Echols, "One Health' - the Rosetta stone for 21st century health and health providers", Vet Ital. 2009 Jul-Sep;45(3), pp. 377-382. PubMed PMID: 20391400.

[14] European Union, "One Health: Addressing health risks at the interface between animals, humans, and their environments," http://eeas.europa.eu/health/pandemic_readiness/index_en.htm Accessed Aug 28, 2013.

[15] http://www.who.int/influenza/resources/documents/tripartite_concept_note_hanoi/en/index.html Accessed 25 June 2013.

[16] http://www.infectionecologyandepidemiology.net Accessed Aug 28, 2013.

[17] The World Bank. "People, Pathogens, and Our Planet. Volume 2: The Economics of One Health." Report no. 69145-GLB. http://www-wds.worldbank.org/external/default/WDSContentServer/WDSP/IB/

2012/06/12/000333038_20120612014653/Rendered/PDF/691450ESW0whit0D0ESW120PPPvol120web.pdf.
Accessed Aug 28, 2013.

[18] B. Kaplan "One Health or…some health?" Vet Ital. 2011 Apr-Jun;47(2), pp. 129-131. PubMed PMID: 21706464.

[19] M.K. Gold (Ed.), Debates in the Digital Humanities. Minneapolis/London: University of Minnesota Press, 2012.

[20] http://www.cdh.ucla.edu/about/what-is.html,
Accessed Aug 28, 2013.

[21] R. Rogers, E. Weltevrede, S. Niederer and E.K. Borra, "National Web Studies: The Case of Iran Online," in A Companion to New Media Dynamics, J. Hartley, A. Bruns, and J. Burgess (Eds), Oxford: Blackwell, 2013, pp. 142-166.

[22] A. Robichaud and E. Steiner, "Working paper; Trail of Blood. The Movement of San Francisco's Butchertown and the Spatial Transformation of Meat Production, 1849-1901", Spatial History Lab: http://www.stanford.edu/group/spatialhistory/cgi-bin/site/pub.php?id=31) Accessed Sep 10, 2013.

[23] http://republicofletters.stanford.edu/
Accessed June 15, 2013.

[24] http://www.clarin.eu/
Accessed June 15, 2013.

[25] H.C. Ossebaard and J.W.E.C. van Gemert-Pijnen, "The future of health care", in Improving eHealth, J.W.E.C. van Gemert-Pijnen, O. Peters and H.C. Ossebaard (Eds.). The Hague: Eleven publications, 2013, pp. 9-33.

[26] T.M. Bernardo, A. Rajic, I. Young, K. Robiadek, M.T. Pham and J.A. Funk, "Scoping Review on Search Queries and Social Media for Disease Surveillance: A Chronology of Innovation" J Med Internet Res 2013;15(7):e147 doi: 10.2196/jmir.2740

[27] H.C. Ossebaard and J.E.W.C. van Gemert-Pijnen "Public Health 2.0 – fresh approaches to old problems" in Public health research methods, Guest G (Ed.) Thousand Oaks CA: SAGE Publications, in press.

[28] C.R. Janes, K.K. Corbett, J.H. Jones and J. Trostle "Emerging infectious diseases: the role of social sciences" The Lancet 2012, 380(9857), pp 1884 – 1886. doi:10.1016/S0140-6736(12)61725-5

[29] D.M. Bell, "Non-pharmaceutical interventions for pandemic influenza, international measures" Emerg. Infect. Dis. 2006;12(1), pp. 81-87.

[30] S. Funk, M. Salathé and V.A. Jansen "Modelling the influence of human behaviour on the spread of infectious diseases: a review", J R Soc Interface. 2010 Sep 6;7(50), pp. 1247-1256. doi: 10.1098/rsif.2010.0142.

[31] M. Schaller, "The behavioural immune system and the psychology of human sociality", Philos Trans R Soc Lond B Biol Sci. 2011 Dec;366(1583), pp. 3418-3426. doi: 10.1098/rstb.2011.0029.

[32] L. Neuhauser and G.L. Kreps, "Ehealth communication and behavior change: Promise and performance" Social Semiotics, 2010, 20(1), pp. 7-24.

[33] A. Jurke, R. Köck, K. Becker, S. Thole, R. Hendrix, J. Rossen, I. Daniels-Haardt and A.W. Friedrich, "Reduction of the nosocomial meticillin-resistant Staphylococcus aureus incidence density by a region-wide search and follow-strategy in forty German hospitals of the EUREGIO, 2009 to 2011", Euro Surveill. 2013;18(36):pii=20579.
http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20579 Accessed Oct 15, 2013.

[34] www.infectionmanager.com
Accessed June 15, 2013.

[35] J.E.W.C. van Gemert-Pijnen, O. Peters and H.C. Ossebaard (Eds.) Improving eHealth. The Hague: Eleven publications, 2013.

[36] E. Murray, "Web-Based Interventions for Behavior Change and Self-Management: Potential, Pitfalls, and Progress", Med. 2.0 2012;1:e3. doi:10.2196/med20.1741.

[37] M. Huber, J.A. Knottnerus, L. Green, H. van der Horst, A.R. Jadad, D. Kromhout, B. Leonard, K. Lorig, M.I. Loureiro, J.W. van der Meer, P. Schnabel, R. Smith, C. van Weel and H. Smid, "How should we define health?" BMJ. 2011 Jul 26, p. 343:d4163. doi: 10.1136/bmj.d4163.

[38] D. Beaujean, L. van Velsen, J.E. van Gemert-Pijnen, A. Maat, J.E. van Steenbergen and R. Crutzen, "Using risk group profiles as a lightweight qualitative approach for intervention development: an example of prevention of tick bites and lyme disease", J Med. Internet Res. - Res Protoc. 2013 Oct 30;2(2):e45.
doi: 10.2196/resprot.2760.

[39] M. Salathé, C.C. Freifeld, R. Sumiko Mekaru, A. F. Tomasulo, and J.S. Brownstein, "Influenza A (H7N9) and the Importance of Digital Epidemiology" N Engl J Med 2013; 369, pp. 401-404. doi: 10.1056/NEJMp1307752

[40] L. Manovich, "Trending: The promises and the challenges of big social data," in Debates in the digital humanities, M.K. Gold (Ed.). Minneapolis: University of Minnesota Press, 2012, pp. 460–475.

[41] R. Rogers, Digital methods. Cambridge: MIT Press, 2013.

[42] The 3rd international One health Congress in Amsterdam, the Netherlands; 15-18 March 2015. See: http://www.iohc2015.com
Accessed Sep 15, 2013.

Telemedicine: Trends and Challenges

Special issue

Guest editors
Lisette van Gemert-Pijnen, University of Twente, The Netherlands
Marike Hettinga, Windesheim University of Applied Sciences, The Netherlands

This special issue is based on selected papers and extended papers of the Fifth International Conference on eHealth, Telemedicine, and Social Medicine, held on February 2013, in Nice, France. The authors of the best papers of this conference were invited to extend their papers to a full journal paper.

The eTelemed conference is a well-visited conference with lively discussions in tracks on technical aspects, societal and financial aspects, as well as on user-oriented research. The theme of the eTelemed conference in 2013 was globalization and the impact on health care, costs and society. In particular, collecting, storing and management of patient data were highlighted to discuss the implications for safety, online access, and information systems to process and store data. The integration of advanced and innovative technologies, like teleradiology, imaging and self-tracking sensing systems, with traditional care is still a challenge. Citizen-centred care through safe, smart and mobile technologies and new information tools requires new visions on accountability, responsibility and managing health and wellbeing.

The best papers of this conference are selected by the International TPC (Technical Program Committee) and the guest-editors on the basis of their contents, specifically for lending themselves to an interesting extended work. TPC consists of international experts who are leading researchers in the field, working in academia and industry. We, as the invited guest-editors of this special issue, both lead research teams in the domain of eHealth.

The papers of this special issue are from researchers all over the world, and from various disciplines. Together, they reflect the interdisciplinary field of eHealth. The papers address topics as: Design and implementation criteria for advanced telemonitoring systems to support chronic care patients and elderly, User centred development approaches to design platforms for telecare and distance learning systems, Multidisciplinary development of devices to support medical decision-making, Critical factors for using social media in a governmental communication setting, and Business modelling for eHealth innovation of small and medium sized enterprises. For example:

- The relevance and significance of the local context when establishing and implementing telemedicine services
- Real time audio-visual telemedicine program providing 24/7 emergency and Intensive Care physician coverage; factors for implementing a successful model.
- Long-term Home Monitoring of COPD Patients via a wearable System
- Design of NOWAPI system for telemonitoring the treatment of obstructive sleep apnea.

- User-centred development of AAL-JP project fearless for elderly. Balanced Scorecard to guide the multi-disciplinary development process.

- Development of an unified platform for Ambient Assisted Living for telecare and communication of elderly

- EMuRgency**;** AI-driven volunteer Notification System for integrating laypersons into Emergency Medical Services

- Improving Online Interactive Modules for e-Learning trough iterative and formative evaluations with stakeholders and designers

- A dynamic interface for Linked Open Drug Data (LODD) to automatically detect the drugs in a prescription, to collect multidimensional data and to supports automatic prevention of possible drug interactions in a prescription.

- Application of Linguistic Approaches to rank the effectiveness of treatment modalities from the most recommended to the contraindicated; in order to support multi-expert decision-making in healthcare

- Multimodal cognitive nonverbal and verbal Interactions: the neurorehabilitation of autistic children via mobile Toy Robots

- A critical perspective on governmental communication through Twitter; what is good practice in communicating through social media channels?

- Introduction of an eHealth Innovation Map, for small and medium-sized enterprises (SMEs)

As the guest editors, we wish the reader an enjoyable and interesting reading of these papers, and above all, we hope to see you at one of the future eTelemed conferences.

Lisette van Gemert-Pijnen is the editor in chief of the International Journal On Advances in Life Sciences and member of the scientific board of eTelemed. She is the head of the Center for eHealth Research & Disease management at the University of Twente, The Netherlands. Her research group's program focuses on persuasive designs (people-driven) and business modelling (value-driven) for interactive eHealth interventions, in particular, on how persuasive designs can increase trust, engagement, and adherence to eHealth interventions aimed at safe-care and self-care. Lisette is one of the Advisory Chairs of the eTelemed series.

Marike Hettinga leads a research team on ICT-innovations in Healthcare at the Windesheim University of Applied Sciences, The Netherlands. The central theme of this team is the structural embedding of ICT-innovations in healthcare leading to a research focus on business modelling, methods for evidence based eHealth and user oriented research. Furthermore, technological research questions are addressed concerning sensor technology and big data. Marike is one of the Advisory Chairs of the eTelemed series.

Analysing the Use of a Telestroke Service

Kari Dyb
Norwegian Centre for Telemedicine,
University hospital of North Norway, Tromsø,
Norway
kari.dyb@telemed.no

Terje Solvoll
Norwegian Centre for Telemedicine,
University hospital of North Norway, Tromsø,
Norway
terje.solvoll@telemed.no

Ellen Rygh
Norwegian Centre for Telemedicine,
University hospital of North Norway, Tromsø,
Norway
ellen.rygh@telemed.no

Tove Sørensen
Norwegian Centre for Telemedicine,
University hospital of North Norway, Tromsø,
Norway
tove.sorensen@telemed.no

***Abstract*—Telestroke can facilitate remote cerebrovascular specialty consults from virtually any location within minutes of attempted contact, adding greater expertise to the care provided to any individual patient. In 2010, a Telestroke service was established between the three hospitals that constitute the Nordland Hospital in Northern Norway. The Telestroke service connected the neurological department at the central hospital with two small, local hospitals. The aim of our study was to implement a Telestroke service and explore if the benefits reported internationally could be achieved in the North Norwegian healthcare context. The study presented herein was based on Telestroke log data, field observations, semi-structured interviews and focus group discussions with the health personnel involved. After more than two years of operation, Telestroke consultations were described as successful. However, the frequency of use was much lower than the estimate before implementation, and the Telestroke service was terminated with the project conclusion. In this article, we ask why the service was terminated, and focus on the significance of the local context when establishing telemedicine services.**

Keywords- Telestroke; video-teleconferencing; consultation; cerebral stroke; local hospitals; local context

I. INTRODUCTION

The volcano ash incident in Iceland in April 2010, gave a reminder of how vulnerable the health services are in Norway without air transport. In May 2010, Northern Norway Regional Health Authority, decided to establish a telemedicine service for diagnosis and treatment of cerebral stroke patients in the Nordland Hospital (NLSH). A Telestroke service was implemented by connecting the two local hospitals, NLSH Vesterålen and NLSH Lofoten, to the Neurology department at the central hospital NLSH Bodø [1]. A pre-study estimate, based on the average frequency of stroke incidents in Norway and the number of inhabitants in the actual area, indicated that the two local hospitals receive an average of two patients with stroke symptoms every week; NLSH Vesterålen (1.25) and NLSH Lofoten (1.6). As the numbers are small, a fully specialized stroke unit with stroke specialists physically present day and night, seven days a week, is expensive and difficult to staff. Therefore, a "VAKe" [2, 3] compatible video-teleconferencing system (VTC) was installed in June 2010. After reaching an agreement on common procedures, revising and on-site training of local health personnel, the Telestroke service was operational from September 2010. It was utilized immediately at one of the local hospitals. At the second local hospital the usage was postponed until a specialized stroke unit was established, including the employment of a specialized stroke nurse and the furnishings of dedicated stroke rooms at the second local hospital.

Telestroke implies that the stroke specialist at the central hospital examines the patient in cooperation with the physicians and nurses at the local hospital, through VTC systems. Radiology images are transmitted using the RIS/PACS systems [4]. The recommended procedure was to make the Telestroke unit ready for transmission and call the central stroke specialist when a patient with stroke symptoms arrived, or was notify to arrive at the local hospitals. The on duty stroke specialist at the central hospital headed to a room dedicated to the Telestroke system.

For more than two years, the Telestroke service was used with success for a few patients with stroke symptoms in Northern Norway. However, the numbers were much lower than the estimate prior to the implementation of the system, and the Telestroke service was terminated with the project conclusion. In this paper, we discuss why the system is no longer in use, and focus on the significance of the local context when establishing Telestroke services.

II. BACKGROUND

Cerebral stroke is the third most frequent cause of death in Norway, and the most common cause of severe disability in adults. The annual incidence is about three per thousand inhabitants, where 85-90% is due to ischemic stroke. The

average treatment cost of one stroke incident is approx. NOK 600,000, adding up to a total annual cost of NOK 7.8 billion. Timely treatment and rehabilitation can reduce disability after stroke, improve quality of life and reduce costs [5, 6]. Hospitalization in a specialized stroke unit leads to a 10 % absolute reduction in mortality in the acute phase [5]. The prognosis for patients with ischemic stroke is further improved by thrombolytic treatment in the acute phase. The main challenge for optimal cerebral stoke treatment is the narrow time window. Thrombolytic treatment should be given as soon as possible, and not later than 4.5 hours after the first symptoms. The national registry of thrombolytic treatment [7] indicates that only 10 % of all stroke patients in Norway received such treatment, whereas 20 % is the recommended frequency by the Norwegian Health Ministry [6]. According to the register, the two local hospitals, in average, used thrombolytic treatment for one patient each during the four month periods reported in 2010 and 2011.

For stroke patients, it is crucial to decide on the cerebral stroke diagnosis and indication for thrombolytic therapy as soon as possible, and no later than 4.5 hours after the first symptoms. Telestroke networks are successful in Europe and the USA [8, 9]. In the Nordic Countries, a Telestroke network between the University Hospital in Helsinki and five local hospitals demonstrates increased frequency of thrombolytic treatment [10].

For cardiac stroke patient, pre-hospital stroke treatment in ambulances is the established procedure in Norway. In Germany, pre-hospital treatment in ambulances has also been tried for cerebral stroke patients [11]. The German study reports that pre-hospital cerebral stroke treatment in specialized ambulances equipped with a Computed Tomography (CT), point-of-care laboratory, and telemedicine connection reduced the median time from alarm to therapy decision substantially [11]. Implementation of CT-ambulances and pre-hospital treatment of cerebral stroke is discussed in Norway, and particular in Northern Norway were dispersed settlement and long travel distances to the nearest hospital is part of everyday life. An internal assessment from the University Hospital of North Norway concludes that the implementation of specialized stroke ambulances with existing mobile CT-equipment will not improve the emergency care for cerebral stroke patient in the northern region. On the contrary, the assessment recommends strengthening the procedures with in-hospital treatment of cerebral stroke patients in this region.

Telestroke can be used to guide and support the local physician remotely [12-16]. According to Audebert, 2006 and Audebert and Schwamm, 2009 [12, 13], the organization of Telestroke networks requires:

- Specialized stroke units in all hospital;
- Comprehensive and continuous education and training of the entire staff in the units;
- Stroke specialists available 24/7 on VTC, combined with teleradiology;
- Centralized organization of patient transfers.

The establishment and use of Telestroke network is consistent with the Norwegian Health Ministry's national guidelines for treatment and rehabilitation of stroke [6]. Telestroke consultations may be useful to decide on the best treatment for the patient, whether it is a conservative treatment, thrombolytic treatment or a more advanced neurological or neurosurgical emergency treatment. Telestroke is also helpful to support quick triaging and transfer to the appropriate unit. Although the decision to give thrombolytic treatment is seen as the end-point of the Telestroke consultation, this is just the beginning of care for the patient [8]. Post-thrombolytic care requires intensive cardiovascular and neurological monitoring, neurosurgical backup, and decision whether to keep the patient or to "drip-and-ship". Tele-consultation may also be useful for follow-up after the acute phase [8].

In Norway, there is limited experience with Telestroke. Some hospitals use remote CT images in combination with phone advice to support acute stroke treatment at small, local hospitals. The single documented service in Norway is between Haukeland University Hospital and the local hospital in Voss, which reports an increase in thrombolytic treatment [17]. Internationally, there is substantial scientific evidence of the medical impacts of Telestroke [12, 13, 18-23]. Several studies of Telestroke solutions including video, versus solutions without video, show that VTC may [24, 25]:

- Reduce the number of wrong diagnoses:
  7.1 % vs. 17.6 %, $p < 0.05$;
- Reduce death rate:
  1.3 % vs. 6.8 %, $p < 0.05$;
- Reduce needs for nursing homes:
  2.6 % vs. 5.4 %, $p = 0.58$.

Telestroke networks, where the experienced stroke specialists performs an evaluation and examination of the patient through a VTC system, and considers the indication for initiation of thrombolytic treatment, is comparable with face-to-face consultations [16]. In our study, the aim was to implement and organize a Telestroke service, and then explore whether the same benefits reported internationally were possible to obtain in a North Norwegian context.

In spite of limited experience with Telestroke in Norway, the nation has long experience with telemedicine, and particular in Northern Norway [26-28]. Unfortunately, in Norway as well as internationally, there is also a long track record of telemedicine services terminating after the pilot project is over [29-30]. The failures do not, in the majority of cases, represent technological problems, but rather human and organizational factors. Numerous studies show that the past failures of technological innovations with respect to improving health outcomes have not necessarily been due to their clinical ineffectiveness, but rather to social, technological and cultural issues relating to their implementation and adoption [31-34].

Hence, the rationale for this paper is to follow the Telestroke service, by studying the social and cultural context of the Telestroke service. The context of the health services has already proven to be important factors to ensure a functional and sustainable telemedicine services in this region [40-41].

III. MATERIALS AND METHODS

A. Research settings

The research was conducted at the three hospitals that constitute NLSH, in Nordland County, Norway. NLSH serves a geographical area with 131,000 inhabitants, and consist of hospitals in the city of Bodø and the archipelagos of Vesterålen and Lofoten. While Bodø is located on the main land, the district of Vesterålen and Lofoten are made up of several islands, see Fig. 1. NLSH Bodø is the largest hospital in the county, and acts as the central hospital of NLSH [3]. The two local hospitals are situated in a geographical area known for its wild and beautiful nature, long distances, small communities, a culture based on earning a living from the sea and extreme rough weather conditions. The intense weather conditions, particularly during the winter, often make it difficult, or even impossible, to use air transport for severely ill patients. Other types of transport from the local hospitals to the central hospital take between five to ten hours.

NLSH Lofoten is placed on the island Vestvågøy, Gravdal (see Fig. 1), which is located in the middle of the Lofoten Islands. The hospital serves approximately 24,000 inhabitants, has a surgical and medical emergency unit, and a maternity ward. NLSH Vesterålen is placed in the community of Stokmarknes (see Fig. 1), covers a geographical area with approximately 30,000 inhabitants. The hospital has a surgical and medical emergency unit and a maternity ward.

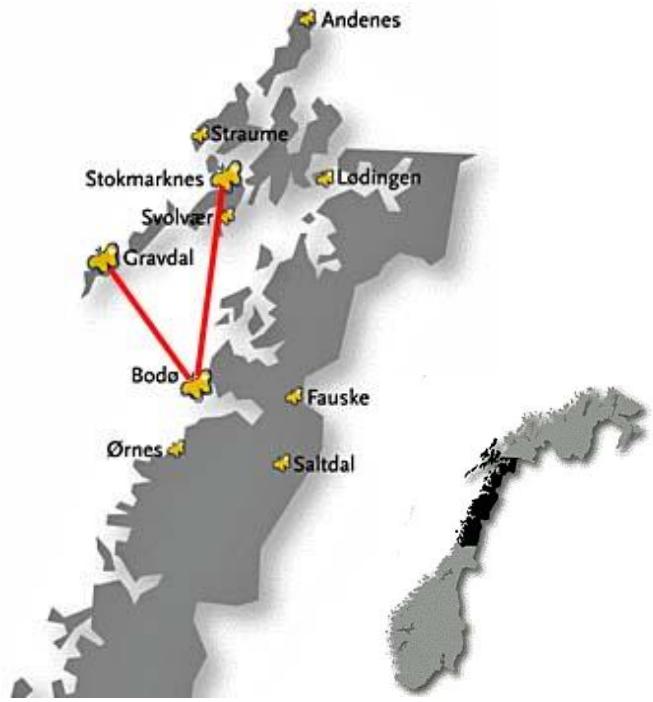


Figure 1. Nordland County, Norway.

In our study the Telestroke service was implemented simultaneously at all three hospitals, and was operative from September 2010. To ensure user confidentiality we will from now on use local hospital A and local hospital B.

At local hospital A, the Telestroke equipment was placed at the intensive care unit (ICU), while the mobile patient centered unit was placed in a small room connected to the emergency room at local hospital B.

To make sure that the technology was operational and worked flawlessly, to maintain the users' expertise and to educate new users, the implementers, Norwegian Centre for Telemedicine (NST), who implemented the service, recommended that each hospital tested the Telestroke solution regularly, and preferably once a week. The local staff agreed to use the system in the initial phase in accordance with these recommendations.

When a potential cerebral stroke patient arrived at one of the two local hospitals, a junior physician could be the first attending. At both hospitals, the procedure was to conduct a CT examination and collect the necessary blood samples as soon as possible. As already mention, the equipment was set up in the intensive care unit at local hospital A. At this hospital the Telestroke system was connected and made ready for a potential session by a nurse, while the patient had the CT examination. At hospital A the Telestroke service was only prepared upon request from a senior physician. This setting was different from local hospital B. Where the system was located next to the emergency unit and part of the nursing procedure was to move the system into position, connected it, and made it ready for use as soon the hospital was notified about an incoming stroke patient.

B. Materials

To implement a Telestroke service VTC equipment must be installed, at least one on each site. In this project, we decided to supply each of the hospitals with the same type of equipment, which reduces possible sources of error and eases user-support. Only minor local adjustments were made due to room configuration. The VTC equipment consisted of a Tandberg Quick set C60 and a Sony full HD (1080p) television, mounted on a mobile rack (see Fig. 2). The Tandberg Quick set C60 has the possibility to connect with medical equipment such as electrocardiography (ECG) together with two full HD (1080p) cameras. The system was also made ready for multipart conferences for communication between more than two locations. The total cost of each unit ready for use was approximately NOK 200,000.

At the central site, the Telestroke system was installed in a dedicated Telestroke studio. The studio had a computer equipped with two 24 inch displays; one display is devoted to the electronic patient record, and one to the CT images. At the two local hospitals, the VTC equipment (see Fig. 2) was mounted on mobile racks, enabling the equipment to be used in different settings such as treatment, education and meetings. The three Telestroke units were connected through the Norwegian Health Net (NHN) [35], which is a dedicated, secured network for health information, see Fig. 3, and were set up to use minimum 2Mbit for this communication.

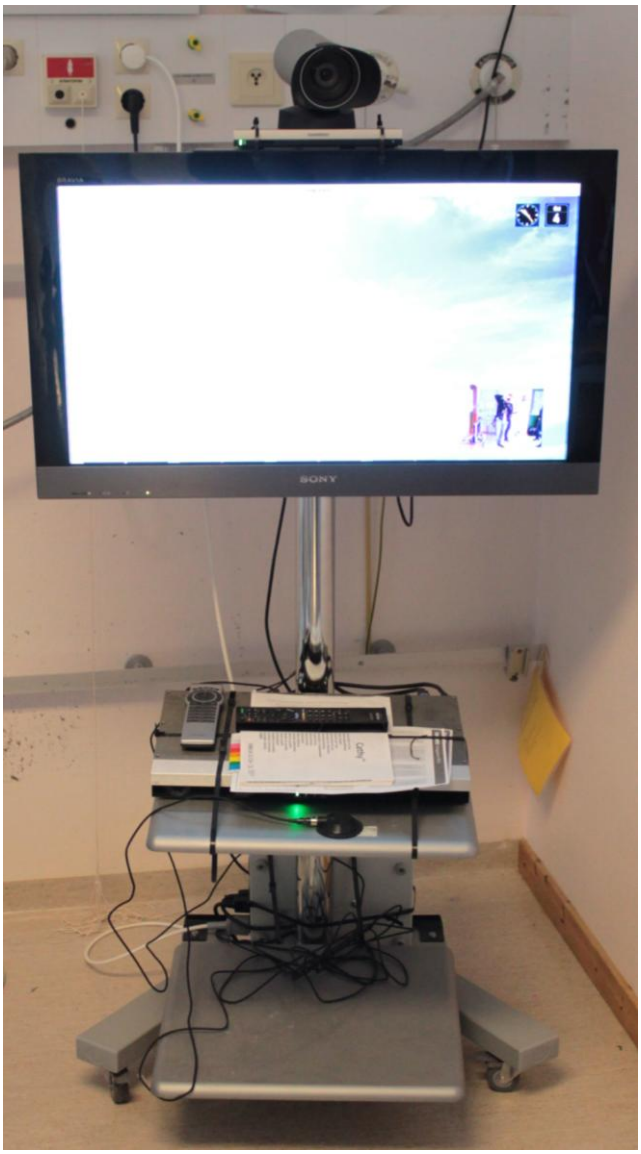


Figure 2. Telestroke equipment.

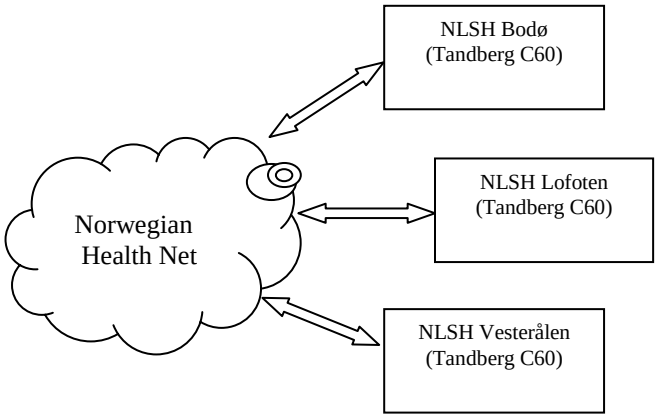


Figure 3. NLSH Telestroke service.

C. Methods

A qualitative, multi-method research approach has been used. Our interdisciplinary analytical approach is inspired by Science Technology Studies (STS) and Actor Network Theory (ANT) [36, 37]. We have followed the technology in use, with a particular focus on how Telestroke is used or not used in the local context. This approach is suitable to avoid both technical and social determinism [38]. Quantitative methods with control-groups were considered, but found non-ethical. It was also considered to include two similar hospitals as control-groups. This was however difficult within the empirical context of Northern Norway, and the framework of our study. The same counts for cost evaluation. As we will show, the numbers in the study are small, and not sufficient for cost analyses. The methods used were:

- Analysis of log data
- Field visits including observations and 15 semi-structured interviews with involved health personnel
- Focus group discussions

Our multi-method approach did not only capture research data, it was also helpful to ensure that the Telestroke system worked according to the plan. It revealed needs for follow-up actions like additional training and revision of procedures.

1) Analyzing Telestroke log data: Exploring and analysing log data is an adequate approach to gather a broad and general understanding of the Telestroke service. The log data indicates the year, date, time, length and frequency of the video and sound communication between the hospitals. Therefore, the study of log data is sufficient to explore the frequency and the length of the connections between the hospitals. It reveals use, as well as lack of activity or absence of Telestroke collaboration. In addition, to visualize the actual Telestroke activity, the study of log data is also a useful approach to map the number and the length of the training sessions, if and how often the hospitals are testing the equipment, if the system is used for other purposes, and the differences in use between the two local hospitals.

2) Field visit and semi-structured interviews: About a year after implementation, autumn 2011, we visited the three hospitals. During the field visits we spent time at the hospitals and in the different local communities. In addition to observations, the data includes informal conversations with health personnel and local inhabitants. Field observations were particularly useful to get an in depth understanding of the stroke treatment at the three hospitals, and to contextualize the Telestroke service in three different hospitals settings. It was also helpful to get an understanding of the visited area and the population.

Semi-structured interviews with the involved health personnel are a constructive approach for gathering the users' satisfaction and experiences with the Telestroke system. During the interviews health personnel reported on specific cases where Telestroke has been used or not used. For instance, why VTC was used or not, the age of the patient, time from onset of symptoms to hospitalization, if

the patient receive thrombolytic treatment or not, etc. The method is also sufficient for gathering information about health personnel experience with the technology and the online collaboration. In addition, semi-structured interviews can reveal potential differences between users, for instance between the remote neurological experts and the onsite local hospital staff. It may also detect technological obstacles or organization challenges with the Telestroke service.

We perfomed a total of fifteen interviews, six were done at the central hospital and nine at the local hospitals. The health personnel have been interviewed at their workplace; in their office, a meeting room, the nurses break room, in the corner of a hospital lounge, or in an empty patient room. Most interviews have been done individually face-to-face, two interviews were in groups, with two and four participants, and one by phone. The interviews were conducted in Norwegian, recorded and transcribed. The quotations used in this paper were translated into English. Throughout the study, we also had e-mail and phone correspondence with selected health personnel.

3) Focus group discussions: Focus group discussions with participants from all three hospitals was an adequate approach to get in contact with the everyday practice of cerebral stroke treatment. Similarities, differences and the dynamic character of the local practices are often revealed in group discussions [39]. Focus group discussions between health personnel from different institutions were also suitable to expose contextual and organizational aspect of the Telestroke service.

4) Ethical considerations: The study has been approved by the North Norwegian Regional Medical Ethics committee (REK). We have not studied patients.

IV. RESULTS

The Telestroke service in Northern Norway was in operation from September 2010 until April 2013. It was only used at one of the two local hospitals, and just for a total of four patients with acute cerebral stroke symptoms. After more than two years of service, the frequency of Telestroke conferences was considerably lower than expected. Telestroke was never integrated in daily practice at the hospital and the service was terminated by the end of our study.

Despite low frequency and lack of use, it is worth noting that the few conferences that were carried out were reported as successful, constructive and valuable. The conference quality was characterized as excellent. The video quality was good enough to detect pupil contraction and eye movements like nystagmus (a rapid, involuntary, oscillatory motion of the eyeball).

During the implementation of Telestroke, the project team strongly recommended to use Telestroke for all patients with acute cerebral stroke symptoms. Neither of the local hospitals followed this advice, but developed local, specific and confined Telestroke procedures. Even if the hospitals procedures varied, both hospitals limited the use of Telestroke for thrombolytic treatment.

The local staff at the hospital that used Telestroke (hospital B) was, in general, more committed to adopt the new service than the involved staff at local hospital A. However, as demonstrated in the four quotes below, physicians from all three hospitals mentioned the potential for using the VTC for additional clinical consultations. This could be other neurological disease, but also entirely different acute or none acute conditions with uncertain diagnoses and treatment.

Informant 1, local hospital:

"Yes, especially in neurology where the examination is particularly important. You know, it is subjective what you see, and how to interpret what you see. If you could get assistance from a neurologist under the examination… very useful!"

Informant 2, local hospital:

"I would like, say a telestudio that can be used for multiply purposes, for instance for interdisciplinary discussion about a cancer patient. I am thinking about the establishment of teams. ….That would have been excellent. I would like an examination and conversation room, where you can talk with the patient and demonstrated in a telestudio."

Informant 3, central hospital:

"I am very skeptical to telemedicine, so I might not be the right person to answer… No, what I would like to use it for is; if you look at YouTube on neurology, it is a lot of involuntary movements and odd symptoms, which is useful to visualize. That is what I would like to use it for. You know; strange movements by the arms, or eye movements that are different. It could be useful, but it is rare conditions."

Informant 4, central hospital:

"No, I think about different neurological conditions, which they (physician at local hospitals) are unsure about and need assistance to diagnose without putting the patient on a plane and fly him over here. I think about the patient. That's who I am concerned about. How to do stuff out there, without flying the patient over here to do the examinations? It must be useful for the patient, avoiding the strain and particularly for older sick people."

A. Hospital A

During the fall of 2011, at the time we did our field studies, a specialized stroke unit with a dedicated stroke nurse, a few beds reserved and adapted for cerebral stroke patients, was opened. Hospital A had one neurologist, five senior physicians, and several junior physicians. The neurologist was usually available during weekdays from 08:00-16:00, while the senior physicians also had second on-call duties, evenings and weekends.

The VTC equipment was mounted on a mobile rack to enable easy movement and use in different settings, though it was usually located in the intensive care unit. The new stroke unit, as well as the intensive care unit, was located on the third floor of the old hospital building. Narrow corridors, lifts and staircases made it impractical, if not impossible, to move the Telestroke equipment, for instance, to the emergency unit at the ground floor, where patients with cerebral stroke symptoms arrived. The laboratory and the CT

were also located at the ground floor. Due to room configuration, the Telestroke equipment could not be placed in the emergency unit. In order to use Telestroke, the patients had to be moved from the emergency unit or CT room at the ground floor, to the third floor and the intensive care unit. The placing of the equipment and the room outline in the old hospital building were described as impractical and time consuming for the use of Telestroke.

All the senior physicians were middle-aged men, who had been working at the hospital for several years. They had broad experience with treatment of cerebral stroke, and most weekdays' also access to local neurologist expertise. Patients with stroke symptoms were usually met by a junior physician in the emergency unit. In the cases where thrombolytic treatment was considered, the junior physicians' routine was to always consult a senior physician prior to initiating thrombolytic treatment or seeking remote cerebrovascular guidance. Remote assistance and using Telestroke were only considered in cases where the local team was uncertain about the best treatment for the patients.

When Telestroke was implemented, remote expertise from senior neurologists were available on weekdays, between 08:00-19:00. The Telestroke service was later expanded to include evening, nights and weekends. During this period, the Telestroke service could be staffed by a junior neurologist, while a senior neurologist was the second on call. Limited "opening hours", or access to a senior neurologist only weekdays and daytime, was a recurrent issue during the interviews. Due to these limitations most of the senior physicians were hesitant and doubtful about the benefits of the Telestroke service. In general, they emphasized that they had local access to equivalent stroke competence that Telestroke could offer from a distance.

At local hospital A, the use of Telestroke was limited to; the intensive care unit; thrombolytic treatment; senior physicians; nights and weekends. The Telestroke service was never used at the hospital.

B. Hospital B

While a specialized stroke unit was newly established at hospital A, such a unit had been in operation for several years at local hospital B. This hospital had trained stroke nurses, two rooms reserved and adjusted for cerebral stroke patients, and a rehabilitation unit for patients and relatives. The hospital had four senior physicians and several junior physicians, but no neurologist. Parts of the hospital buildings were modernized recently, although the major parts were old and not optimal for the work that was carried out. For instance, the emergency unit, which was the first stop for the patients, was located on the second floor. This was close to the stroke unit, still, on a different floor, and in a different part of the hospital than the CT and the laboratory.

Initially, the mobile Telestroke equipment was placed in a small room connected to the emergency unit. Part of the local stoke procedure was to connect the Telestroke system, and make it ready for all patients reported to arrive at the hospital with possible cerebral stroke symptoms. This was done by a nurse. The Telestroke equipment was later moved from the emergency room to a storage facility. The argument was not to occupy space in the emergency unit, as well as to enable alternative usage of the VTC equipment.

The stroke procedure at local hospital B was to use Telestroke to consult a neurologist at the central hospital in all cases where thrombolytic treatment was considered. Telestroke was used four times at the hospital. Prior to Telestroke, the hospitals procedure was to call the central hospital for thrombolytic assistance. Hence, Telestroke replaced the phone as a collaboration tool in cases of thrombolytic treatment. Telestroke, as a substitute for ordinary phone calls, was used on four out of five patients, where the one exception was referred to as a mistake. The hospital did not report delays in thrombolytic treatment when using Telestroke. On the contrary, and as the next quote illustrates, involved health personnel described Telestroke as superior to telephones.

Informant 5, central hospital:

"I think it (Telestroke) was useful because I could talk directly to the patient, which made it easier to assess the aphasia, and also to consider other parameters. He had some additional symptoms, which actually made us decide not to give thrombolytic treatment."

Telestroke can be considered a success at local hospital B. To illustrate the diverse outcome and the clinical value of the service, we have included three cases of use.

- A patient arrived at the local hospital with a diagnosis of possible cerebral stroke. After using the Telestroke service, this diagnose was dismissed and changed. No thrombolytic treatment was given, and the patient was discharged from the local hospital two days later.
- A patient with a heart transplant arrived with cerebral stroke symptoms. After using the Telestroke service, the patient was diagnosed with a possible severe rejection of the transplant. The patient was transferred by air ambulance directly to a specialized hospital in Oslo. No thrombolytic treatment was given.
- A patient with a possible cerebral stroke diagnosis and history of cerebral stroke. The patient had successfully received thrombolytic treatment one and a half year earlier. After using the Telestroke service, the patient received thrombolytic treatment successfully.

However, despite clinical value and positive experience at local hospital B, the frequency of use was considerably lower than expected. From September 2010 - April 2013, the service was only used four times. Throughout the project period, health personnel from all three hospitals were concerned about how limited use could lead to ambiguity and hesitations on how to operate the Telestroke equipment, and how this could be a threat to the long term success of the Telestroke service. After approximately two and a half years of service, this concern became reality. The initial enthusiasm and optimism for Telestroke had faded, and the central hospital could no longer justify spending resources on training and education. Consequential, on one occasion when local hospital B tried to use the Telestroke service, a remote

stroke specialist was no longer available. The Telestroke service in Northern Norway was closed down soon after.

V. DISCUSSION

The three cases of remote stroke collaboration illustrate how a Telestroke service can be useful for diagnosis and treatment of patients with acute cerebral stroke symptoms in Northern Norway. At one of the local hospitals, Telestroke was used for all patients where thrombolytic treatment was considered, except for one. Telestroke became a replacement for phone consultations regarding thrombolytic treatment and was described as constructive, valuable and superior by central stroke specialists and local health personnel. This corresponds with international studies published by Handschu, R. et al. and Meyer, B.C. et al. [24, 25].

The interview data shows that health personnel, in general, were positive to the use of telecommunication services. Nonetheless, in spite of affirmative attitudes and positive feedback, the frequency of use was considerably lower than expected, and the Telestroke service in Northern Norway was only used at one of the local hospitals.

Limited or lack of use raises interesting questions. In the following, we will focus on the local context of the Telestroke service, and start by discussing how we can understand the limited use in Northern Norway. Few or none studies were found on the significance of the local context when implementing Telestroke. This is unfortunate since the context is known to be significant to other telemedicine services in the region, geographical, contextual and organizational environment have proven to be of significance when implementing telemedicine services in the Northern Norwegian region [40, 41].

A Telestroke service in Northern Norway might face challenges that are not reported internationally. Northern Norway, as most parts of Norway, has dispersed settlement, small communities, and relative small hospitals. The numbers of acute stroke incidences in each hospital are low. Long distances and extreme weather conditions, particularly during winter time, might add time to the patients hospital transportation. Clinicians in the region have high turnover, and technical support is not available 24/7.

Successful, but limited use of Telestroke questions the estimated number of potential stroke patients in the region. Was this estimate too high? Local hospital B reported approximately 40 patients with an acute cerebral stoke diagnosis during a 12 month period, from March 2011 to March 2012 (also reported in Nilsen and Solvoll, 2012 [42]). Since our data identified the number of stroke diagnosis, and not the total number of patients that arrived at the local hospital with acute cerebral stroke symptoms, it is difficult to determine whether the initial estimate was too high. On the other hand, interviews with local health personnel confirm a weekly average of one patient with stroke symptoms at the local hospitals. Given that Telestroke was used for almost all patients assessed for thrombolytic treatment at one of the hospitals, limited use might consequently address the low frequency of thrombolytic treatment at the hospital, rather than the actual number of stroke patients in the region.

The main ambition for initiating a Telestroke service in Northern Norway was to improve the stroke treatment and to increase the thrombolytic frequency in the region. This was probably also a key motivation for the Northern regional health authorities funding of this project (HST-1019-11) [43]. The Norwegian health ministry recommends thrombolytic treatment for at least 20 % of all stroke patients [6]. These guidelines specify that treatment must be given as soon as possible, and within 4.5 hours of the debut of the stroke symptoms [6]. However, the Norwegian Registry on thrombolytic treatment (2012) indicates a national average of 10 % [7], and even lower in Northern Norway.

We asked the local physicians if the thrombolytic frequency could be increased in line with the health authority recommendations. A senior physician, informant 2, at one of the local hospitals answered: *"It can be increased to a certain extent, but it's not… We can do something with our procedures internally, but what's more important, or as important, is to catch the public's awareness about cerebral stroke symptoms'. Similar to what has been done with cardiac stroke, where most people know how important it is to get there (to the hospital) quickly."*

Lack of stroke awareness in the population and late incoming patients were highlighted as the main reason for the low thrombolytic frequency at both local hospitals. A rather relaxed attitude towards health and illness might not only be common, but even essential to live in an area where health care services traditionally have not been easily accessible. The two local hospitals are located in a region built on a cultural heritage of earning a living from the sea, known for its remoteness, its closeness to nature and the inhabitants pride in making-do and getting-on. This is a place where it is common to "sleep on it", which means to wait until the morning and see if the situation improves. Contextual aspects like; cultural heritage, closeness to nature, long distances and extreme weather conditions are of significance for how fast stroke patient and their next of kin seek professional health care.

Telestroke was only used to verify thrombolytic treatment. Only a small amount of the stroke patients in the region arrived at the local hospitals within the time frame for thrombolytic treatment. Consequently, the local context was not only of significance for the thrombolytic frequency in the region, it was also crucial for the use of the Telestroke service. In fact, it seems like the strong link between thrombolytic treatment and Telestroke was fatal to the long term success of the Telestroke service in Northern Norway. The low frequency of thrombolytic treatment did not only affect the usage of Telestroke at the local hospitals, it also had consequences for the central hospitals' attitude towards Telestroke. After more than two years of service, the total number of Telestroke conferences was still only four. Due to low usage, the central hospital could not justify spending resources on education and training of new personnel, and the Telestroke service in Northern Norway was closed down by the end of our study.

Local hospital A was a bit larger than hospital B. The hospital had one neurologist and five senior physicians who had been working at the hospital for several years.

Consequently, these physicians knew the hospital, each other, the stroke procedures, the population, and the region well. In addition, most of them questioned the added value of the Telestroke service. That is why, lack of use at local hospital A can be interpreted as an expression for the senior physicians' confidence in their own expertise, and the quality of the stroke treatment at the hospital.

The hospitals' room configurations and the location of the equipment, the newly established stroke unit and clinical practice, were additional aspects that reflected the lack of use.

The air ambulance service in Northern Norway might also have influenced the use of Telestroke. A well-functioning air ambulance service was described as essential for the safety of the population in small rural communities in Northern Norway. Air Ambulances was not exclusive for cerebral stroke patient, it included all emergency patients. Some of the senior physicians were concerned that remote support from medical expertise, like Telestroke, could jeopardize rather than expand the ambulance service.

VI. Conclusions and Future Work

The aim of our study was to implement and organize a Telestroke service to explore if it was possible to obtain the same benefits reported internationally in a North Norwegian context. The results show that usage has been limited to one of the local hospitals, and only for patients where thrombolytic treatment was considered. In total, this was just four cases. However, for all four cases, involved personnel at the local, as well as the central hospital, reported beneficial results for the patient, successful online cooperation for involved personnel, and excellent video and sound quality. Despite positive feedback, the numbers are small, and we cannot demonstrate the same clinical benefits as reported internationally. Our main finding in Northern Norway is low frequency and lack of use.

Our study shows that a Telestroke service depends on a complex interplay of different contextual and local aspects. Lack of stroke awareness in the population, cultural heritage, local in-house procedures, the local hospitals current staff situation, the senior physicians' confidence, the air ambulance situation, and weather conditions, also influences the use of Telestroke.

When implementing and organizing the Telestroke service in Northern Norway, the recommendation was to use Telestroke for all patients with stroke symptoms. This recommendation is in line with international literature, but was not followed. The two local hospitals limited the use of Telestroke for patients considered for thrombolytic treatment. It needs to be questioned if this practice might have terminated the Telestroke service in Northern Norway.

The next step of our research is to compare the Telestroke service in Northern Norway with a similar service in Western Norway. In Western Norway, the local hospital use Telestroke for all patients arriving with cerebral stroke symptoms [17]. It will also be interesting to compare a dedicated service like Telestroke with more general telemedicine services, using similar equipment, in small local hospitals.

Acknowledgment

We would like to thank the Northern Norway Regional Health Authority for funding this research project (HST-1019-11). Special thanks to Professor Rolf Salvesen and the staff at Nordland Hospital. We also want to thank Professor Lars Thomassen at the Haukeland University Hospital, for inspiration and guidance, and Line Lundvoll Warth at the Norwegian centre for telemedicine, for cooperation and discussions.

References

[1] K. Dyb, T. Solvoll, E. Rygh, and T. Sørensen, "Improved Treatment of Cerebral Stroke Patients in Small Hospitals? Reporting from a Telestroke Service in North Norway," in eTELEMED 2012, The Fourth International Conference on eHealth, Telemedicine, and Social Medicine, Valencia, Spain, 2012, pp. 72-77.

[2] S. Bolle, F. Larsen, O. Hagen, and M. Gilbert, "Video conferencing versus telephone calls for team work across hospitals: a qualitative study on simulated emergencies," BMC Emergency Medicine, 2009, vol. 9.

[3] Nordlanssykehuset. 2011 [cited 2013 Nov]; Available from: http://www.nlsh.no/om-oss/category2791.html.

[4] C. Granja, Z. Kokkinogenis, J. Gabriel, and T. Solvoll, "Integration of Healthcare Information Systems: Improving Data Quality in a Diagnostic Imaging Department," in eKNOW 2013, The Fifth International Conference on Information, Process, and Knowledge Management, ed: International Academy, Research and Industry Association (IARIA), 2013, pp. 129-134.

[5] M. Skaset, "Lokalsykehusenes akuttfunksjoner i en samlet behandlingskjede," Oslo, 2007.

[6] Helsedirektoratet, "Nasjonale retningslinjer for behandling og rehabilitering ved hjerneslag. Nasjonale faglige retningslinjer," Oslo, 2010.

[7] Helsedirektoratet. N-016: Trombolysebehandlinger. 2013 [cited 2013 Nov], Available from: http://nesstar2.shdir.no/kvalind/.

[8] J.F. Meschia, "Camera in the Emergency Department: The Evolution of Stroke Telemedicine," Mayo Clinic Proceedings, 2009, vol. 84, no. 1, pp. 3-4.

[9] P. Muller-Barna, L.H. Schwamm, and R.L. Haberl, "Telestroke increases use of acute stroke therapy," Curr Opin Neurol, 2012, vol. 25, no. 1, pp. 5-10.

[10] T. Sairanen, S. Soinila, M. Nikkanen, K. Rantanen, S. Mustanoja, M. Farkkila, I. Pieninkeroinen, H. Numminen, P. Baumann, J. Valpas, T. Kuha, M. Kaste, T. Tatlisumak, and F. Finnish Telestroke Task, "Two years of Finnish Telestroke: thrombolysis at spokes equal to that at the hub," Neurology, 2011, vol. 76, pp. 1145-1152.

[11] S. Walter, P. Kostopoulos, A. Haass, I. Keller, M. Lesmeister, T. Schlechtriemen, C. Roth, P. Papanagiotou, I. Grunwald, H. Schumacher, S. Helwig, J. Viera, H. Korner, M. Alexandrou, U. Yilmaz, K. Ziegler, K. Schmidt, R. Dabew, D. Kubulus, Y. Liu, T. Volk, K. Kronfeld, C. Ruckes, T. Bertsch, W. Reith, and K. Fassbender, "Diagnosis and treatment of patients with stroke in a mobile stroke unit versus in hospital: a randomised controlled trial," Lancet Neurol, 2012, vol. 11, pp. 397-404.

[12] H. Audebert, "Telestroke: effective networking," Lancet Neurol, 2006, vol. 5, no. 3, pp. 279-282.

[13] H.J. Audebert and L. Schwamm, "Telestroke: scientific results," Cerebrovasc Dis, 2009, vol. 27 no. 4, pp. 15-20.

[14] T. Johansson and C. Wild, "Telemedicine in acute stroke management: systematic review," Int J Technol Assess Health Care, 2010, vol. 26, no. 2, pp. 149-155.

[15] M. P. LaMonte, M. N. Bahouth, P. Hu, M. Y. Pathan, K. L. Yarbrough, R. Gunawardane, P. Crarey, and W. Page, "Telemedicine for acute stroke: triumphs and pitfalls," Stroke, 2003, vol. 34, pp. 725-728.

[16] L. H. Schwamm, R. G. Holloway, P. Amarenco, H. J. Audebert, T. Bakas, N. R. Chumbler, R. Handschu, E. C. Jauch, W. A. Knight, S. R. Levine, M. Mayberg, B. C. Meyer, P. M. Meyers, E. Skalabrin, L. R. Wechsler, o. b. o. t. A. H. A. S. Council, and t. I. C. o. P. V. Disease, "A Review of the Evidence for the Use of Telemedicine Within Stroke Systems of Care," Stroke, 2009, vol. 40, pp. 2616-2634.

[17] L. Thommasen, "Behandling av akutt hjerneinfarkt," Tidsskr Nor Lægeforen, 2007, vol. 127, pp. 1060-1063.

[18] K. Khan, A. Shuaib, T. Whittaker, M. Saqqur, T. Jeerakathil, K. Butcher, and P. Crumley, "Telestroke in Northern Alberta: a two year experience with remote hospitals," Can J Neurol Sci, 2010, vol. 37, pp. 808-813.

[19] B.M. Demaerschalk, and T.E.J. Kiernan, "Vascular neurology nurse practitioner provision of telemedicine consultations," International journal of telemedicine and applications, 2010.

[20] B.M. Demaerschalk, "Telemedicine or telephone consultation in patients with acute stroke," Current neurology and neuroscience reports, 2011, vol. 11, no. 1, pp. 42-51.

[21] G. S. Silva, S. Farrell, E. Shandra, A. Viswanathan, and L. H. Schwamm, "The status of Telestroke in the United States: a survey of currently active stroke telemedicine programs," Stroke, 2012, vol. 43, pp. 2078-2085.

[22] S. Wang, S. B. Lee, C. Pardue, D. Ramsingh, J. Waller, H. Gross, F. T. Nichols, 3rd, D. C. Hess, and R. J. Adams, "Remote evaluation of acute ischemic stroke: reliability of National Institutes of Health Stroke Scale via telestroke," Stroke, 2003, vol. 34, pp. e188-1891.

[23] S. F. Zaidi, M. A. Jumma, X. N. Urra, M. Hammer, L. Massaro, V. Reddy, T. Jovin, R. Lin, and L. R. Wechsler, "Telestroke-guided intravenous tissue-type plasminogen activator treatment achieves a similar clinical outcome as thrombolysis at a comprehensive stroke center," Stroke, 2011, vol. 42, pp. 3291-3293.

[24] R. Handschu, M. Scibor, B. Willaczek, M. Nuckel, J. G. Heckmann, D. Asshoff, D. Belohlavek, F. Erbguth, and S. Schwab, "Telemedicine in acute stroke: remote video-examination compared to simple telephone consultation," J Neurol, 2008, vol. 255, pp. 1792-1797.

[25] B. C. Meyer, R. Raman, T. Hemmen, R. Obler, J. A. Zivin, R. Rao, R. G. Thomas, and P. D. Lyden, "Efficacy of site-independent telemedicine in the STRokE DOC trial: a randomised, blinded, prospective study," Lancet Neurol, 2008, vol. 7, pp. 787-795.

[26] D. R. Elford, "Telemedicine in northern Norway," J Telemed Telecare, 1997, vol. 3, pp. 1-22.

[27] D. Gammon, S. Bergvik, T. Bergmo, and S. Pedersen, "Videoconferencing in psychiatry: a survey of use in northern Norway," 1996, J Telemed Telecare, vol. 2, pp. 192-8.

[28] E. Johnsen, Breivik, E., Myrvang, R., & Olsen, F. , "Gevinster av norsk telemedisin," NST, Tromsø, 2006.

[29] B. Andre, G. I. Ringdal, J. H. Loge, T. Rannestad, and S. Kaasa, "The importance of key personnel and active management for successful implementation of computer-based technology in palliative care: results from a qualitative study," Comput Inform Nurs, 2008, vol. 26, pp. 183-9.

[30] A. Obstfelder, K. H. Engeseth, and R. Wynn, "Characteristics of successfully implemented telemedical applications," Implement Sci, 2007, vol. 2, no. 25.

[31] J. Car, A. Black, C. Anandan, K. Cresswell, C. Pagliari, B. McKinstry, R. Procter, A. Majeed, and A. Sheikh, "The impact of eHealth on the quality and safety of health care: a Systematic Overview and Synthesis of the Litterature," Report for the NHS Connecting for Health Evaluation Programme, 2008.

[32] PG. Neumann, "The Risks Digest" [Cited 2013 Nov.]; Available from: http://catless.ncl.ac.uk/risks

[33] G. I. Rochlin, "Trapped in the Net: The Unanticipated Consequences of Computerization," Princeton University Press, 1997

[34] E. Tenner, "Why things bite back: technology and the revenge of unintended consequences," Vintage Books, 1997.

[35] Norsk Helsenett 2011 [cited 2013 Nov]; Available from: http://www.nhn.no/om-oss.

[36] B. Latour, "Science in action: How to follow scientists and engineers through society," 1987: Harvard university press.

[37] J. Law and J. Hassard, "Actor network theory and after," 1999.

[38] M. Akrich, "The de-scription of technical objects," Shaping technology/building society, 1992, pp. 205-224.

[39] A. Fontana and J. Frey, "The interview: from structured questions to negotiated text," Collecting and Interpreting Qualitative Materials, N.K. Denzin and Y.A. Lincoln, Editors, Sage, pp. 61-106, 2003.

[40] H. K. Andreassen and K. Dyb, "Differences and inequalities in health," Information, Communication & Society, 2010, vol. 13, pp. 956-975.

[41] K. Dyb and S. Halford, "Placing Globalizing Technologies: Telemedicine and the Making of Difference," Sociology, 2009, vol. 43, pp. 232-249.

[42] L. L. Nilsen and T. Solvoll, "Traditional Practice vs. New Tools and Routines in Stroke Treatment," Future Internet, 2012, vol. 4, pp. 688-699.

[43] T. Sørensen, T. Solvoll, E. Rygh, R. Salvesen, L. Rye, and K. Dyb, "Utprøving av teleslag ved Nordlandssykehuset," NST, 2011.

Innovation Routes and Evidence Guidelines for eHealth Small and Medium-sized Enterprises

Towards Feasible yet Convincing Evidence

Ruud Janssen, Marike Hettinga, Sikke Visser, Robbert Menko,
Hilco Prins, Irene Krediet, Timber Haaker, and Lianne Bodenstaff
Research group ICT Innovations in Health Care
Windesheim University of Applied Sciences
Zwolle, The Netherlands
{TJWM.Janssen}{M.Hettinga}{S.Visser}{RA.Menko}{H.Prins}{I.Krediet}@windesheim.nl
Timber.Haaker@novay.nl, L.Bodenstaff@bizzdesign.nl

Abstract **- eHealth applications hold many promises, for instance to improve the quality of health care, to increase its accessibility, or to reduce its cost. Yet, many eHealth innovations never reach the stage where they get embedded into routine health care. This is due in part to a lack of evidence that these innovations indeed deliver what they promise. For small and medium-sized enterprises (SMEs) in particular, collecting convincing evidence for eHealth innovations proves to be a challenge as the available time, resources and expertise to do so are often limited. In response to this challenge, the research group *ICT Innovations in Health Care* at the Windesheim University of Applied Sciences initiated the project *Successful Entrepreneurship in eHealth*. The project is a cooperation between 28 parties in The Netherlands: health care providers, patient organizations, health insurance companies, national health care authorities, and 15 eHealth SMEs. Its aim is to speed up eHealth innovation by providing eHealth SMEs with guidelines for collecting feasible yet convincing evidence. In this paper the project's approach is introduced and the main results and lessons learned are discussed. The evaluation of eHealth innovations was found to be highly context-driven, with parties involved each employing their own criteria. Among the project's main results is an *eHealth innovation map*, a diagram showing eHealth SMEs which parties in the Dutch health care system to involve, their roles and their mutual relations, their interests in eHealth innovation, and the kinds of evidence that may convince them of the added value of an eHealth innovation. A set of corresponding fact sheets was developed to provide eHealth SMEs with concise yet easily accessible information for choosing an innovation route and for determining what evidence to collect for relevant stakeholders.**

Keywords - eHealth; innovation route; evidence guidelines; health care system; stakeholder

I. INTRODUCTION

Getting an eHealth innovation embedded into routine health care often turns out to be a challenge. Several causes can be identified, including a narrow focus on technological aspects of the innovation, too little involvement from key stakeholders during design and implementation, or lack of a good underlying business model [2]. The research group *ICT Innovations in Health Care* at the Windesheim University of Applied Sciences (Zwolle, The Netherlands) has dedicated itself to study these issues and to support small and medium-sized enterprises (SMEs) in overcoming them. For instance, the research group recently published the eHealth Innovation Matrix [3]; an online assessment and library that offers eHealth SMEs[1] guidance in developing and evaluating a business model for their eHealth innovations.

A. Problems collecting evidence for eHealth innovations

In 2011 the research group organized a series of workshops for eHealth SMEs and organizations in health care. During these workshops an inventory was made of the problems encountered when getting eHealth innovations embedded in routine health care. Among the list of problems, collecting *evidence* for an innovation came out first. To get their innovation accepted by patients and care providers, reimbursed by health insurance companies, endorsed by patient organizations, or approved by national health care authorities, innovators often need to show evidence for the innovation's effectiveness, for instance to improve treatment quality or reduce the cost of delivering health care.

For a typical eHealth SME it is often unclear what kind of evidence is expected and by whom, and according to which standards this evidence should be collected. In other cases, the standard may be clear (e.g., a randomized controlled trial) yet practically unfeasible for an SME due to a lack of available time, (financial) resources, or expertise. To complicate matters further, care providers, insurance companies and care authorities offer no clear guidelines for eHealth innovators. They recognize that this discourages eHealth adoption and that it impedes innovation within the Dutch health care system [4].

Other researchers have also identified this barrier to eHealth implementation, albeit not specifically for SMEs. For instance, Mair et al. [5, 6] conclude in a meta-review of eHealth implementation studies that lack of validation and

[1] eHealth SMEs are defined here as small and medium-sized enterprises offering eHealth products and services to patients, health care providers, and the general public. All eHealth SMEs participating in the project had less than 10 employees.

evaluation is frequently presented as a barrier to eHealth implementation: *"Without strong data demonstrating that a system works, improves standards of care, can be used efficiently and easily, and is cost-effective to implement, it is unlikely to win the confidence of policy makers and users."* [5, p. 23].

B. Towards feasible yet convincing evidence

The project described here, *Successful Entrepreneurship in eHealth* [1], was initiated by the research group to address these challenges. The project constitutes a cooperation between 28 eHealth SMEs, health care providers, patient organizations, health insurance companies, and national health care authorities in The Netherlands. The project's aim is to establish guidelines for collecting evidence in such a way that (i) it is practically feasible for eHealth SMEs to do so and (ii) the resulting evidence is acceptable and potentially convincing for care providers, health insurers, or care authorities. Hence, the project's motto: *towards feasible yet convincing evidence*.

To achieve its aim, the project set out to address the following research questions:

1. What kinds of evidence for eHealth innovations are generally recognized? Are there any commonly accepted evaluation frameworks?
2. What are relevant outcome indicators and methods to collect specific kinds of evidence? How do these compare in terms of methodological quality and practical feasibility?
3. Which parties in the Dutch care system (patients, care providers, health insurance companies, national care authorities, others) will need to be convinced of the effectiveness of an eHealth innovation before it can be embedded into routine practice?
4. How do these parties value the kinds of evidence mentioned earlier? What typically constitutes "convincing evidence" for these parties?

By generating answers to these questions the project aims to offer guidance to eHealth SMEs: which parties will need to be convinced of the effectiveness of an innovation, what evidence will be required, and how to collect this evidence in a feasible yet acceptable way.

The structure of the remainder of this paper is as follows. In Section II the approach followed will be introduced, including the four phases in which the project was structured. Next, Sections III to VI will discuss the main results for each of the project's phases (inventory, case studies, guidelines & best practices, and consolidation & tool development). Finally, Section VII summarizes the main conclusions.

II. APPROACH

The project *Successful Entrepreneurship in eHealth* started at the beginning of 2012 and will conclude at the end of 2013. At the outset the project was structured into four phases. These phases are briefly outlined in this section.

A. Phase 1: Inventory

During this phase an inventory was made of generally recognized types of evidence. This was done by means of a literature review, consultation of online documentation, an expert session with representatives of Dutch health care providers, insurers, patient organizations, and national health care authorities, and a series of follow-up interviews with these experts. Questions to be answered included: Which parties are involved when getting an eHealth innovation embedded in routine health care? What kind of evidence is generally needed, and how should it be collected? How do parties value various kinds of evidence? And what criteria are typically used?

The expert session was recorded and transcribed, and of each follow-up interview a report was made and sent to the participant for verification. The results of the expert session and the follow-up interviews were then summarized in a joint interpretation session by the research team, using so-called affinity diagramming [7]. Thus, the main results of this phase were:

- An overview of (scientifically founded) frameworks for the evaluation of eHealth innovations;
- A detailed list of outcome indicators and methods, clustered into three identified themes (effectiveness, cost efficiency, and labor savings);
- A comprehensive description of the Dutch health care system, including the roles of the parties involved and their interests in eHealth innovation;
- An outline of four main strategies ("innovation routes") to get eHealth innovation embedded into routine health care;

B. Phase 2: Case studies

Whereas the analysis during the inventory phase was top-down, the analysis during the case studies was deliberately bottom-up – to involve the SMEs and to enrich the analysis with examples of concrete situations, dilemmas and obstacles encountered. To this end, cases from the participating eHealth SMEs were subjected to a detailed study by means of in-depth, semi-structured interviews and an analysis of available documentation. Questions included: How are SMEs trying to get their innovations embedded into routine care? Which stakeholders do they identify and involve? What kinds of evidence do these stakeholders require? What evidence did the SMEs collect so far, and in what ways? How did stakeholders evaluate the evidence, against what criteria?

During each interview, the path followed by the SME to get its eHealth innovation embedded into routine care was reconstructed. Particular attention was paid to the stakeholders that had been identified and involved, and (if applicable) the evidence that had been collected. Where available, underlying documentation was used to analyze the collected evidence, in particular the outcome indicators and methodology used, the conclusions drawn, and, if applicable, how these conclusions were translated into a business case for stakeholders. Of each interview a report was made and sent to the participant for verification. Here as well, the

results were summarized in a follow-up interpretation session using the affinity diagramming technique [7]. The main results included:

- Detailed, in-depth descriptions of successful and less successful strategies followed by SMEs to get their eHealth innovations embedded in routine health care;
- Specific examples of evidence that was collected and, if applicable, how it was evaluated by health care providers, insurers, patient organizations, or national health care authorities.

C. Phase 3: Guidelines and best practices

In this phase, the insights gained from the inventory and the case studies were combined. Best practices for embedding eHealth innovations in routine health care were identified, and guidelines for collecting required evidence were developed. Best practices and guidelines were then combined into a systematic approach for collecting evidence for eHealth innovations: the "eHealth innovation map". This was done in a series of joint interpretation sessions by the research team. At several moments during this process, intermediate results were presented to and discussed with experts and project partners to collect feedback and suggestions for improvements.

To validate the newly developed approach it is currently being applied and evaluated in a second series of case studies. Validation has also been performed during a series of workshops, both within the project (as part of the regular project meetings) and outside of the project (e.g., at national and regional eHealth-related conferences and symposia). In these workshops the approach was applied to a range of different cases at hand (usually provided by workshop participants) and evaluation happened afterwards by means of questionnaires and discussions with participants. Thus, the results of this phase included:

- A systematic approach, consisting of an "eHealth innovation map" showing which parties to involve, their roles and mutual relations, their interests in eHealth innovation, and the kinds of evidence that may convince them;
- A set of corresponding fact sheets providing concise yet detailed information for choosing an innovation route and for determining what evidence to collect for relevant stakeholders;
- Validation of the approach, including an inventory of practical issues and points for improvement (partly this is still work in progress).

D. Phase 4: Consolidation and tool development

In this final project phase, the systematic approach described above is being consolidated in a workshop protocol and a web-based tool, and documented in a booklet:

- The workshop protocol and web-based tool both provide guidance in choosing the most promising innovation route, identifying relevant stakeholders, and determining which kinds of evidence they may require.
- The booklet documents the systematic approach in a concise and accessible way, and aims to disseminate the project's results to the wider audience of eHealth SMEs in The Netherlands.

E. Ongoing dialogue

Next to the activities in the above four phases, regular project meetings were organized to stimulate an ongoing dialogue between the participating organizations. During these meetings, SMEs introduced their cases, representatives of health care organizations discussed procedures or criteria used to evaluate eHealth innovations, and the research team presented the project's latest results. To collect feedback from the project's participants, mini-workshops were organized to evaluate the usefulness and correctness of the developed tools (such as the eHealth innovation map and fact sheets), typically by applying these tools to cases at hand.

This approach resulted in several collaborations among the project's participants. For instance, during one of the meetings a representative of a health insurance company called upon the participating SMEs to enroll their innovations at the company's health innovation desk. Three SMEs did, and the progress of the three enrollments was then monitored closely by the research team to learn about the procedure and the criteria being used by this health insurance company to evaluate the three innovations. During this process both sides (SMEs and insurance company) were regularly heard by means of structured telephone interviews or brief questionnaires sent and answered by email. Furthermore, and as part of this process, the insurance company provided a list of the indicators used by its innovation desk to select promising eHealth innovations.

III. RESULTS: INVENTORY

This section highlights the results and lessons learned from the first phase, the inventory.

A. Frameworks for evaluating eHealth

During the literature study more than a few reports and scientific papers offering proposals for eHealth evaluation frameworks were found, most of them containing guidelines for setting up a proper evaluation study, lists of outcome indicators and measures for various aspects of eHealth's impact, or descriptions of methods and instruments to collect data. We will describe four representative examples here. The extent to which these frameworks are actually being adopted and used, could not be established from the literature.

1) NTOIP

The Canadian *National Telehealth Outcome Indicators Project* [8] is a comprehensive framework covering four dimensions of evaluating eHealth: quality, access, acceptability, and cost. It is based on a hierarchy consisting of categories (for instance, "health status"), themes ("quality"), indicators ("quality of life"), measures ("morbidity"), and tools ("SF-12"). In total, 12 outcome indicators have been defined for the dimension quality, 6 for access, 15 for acceptability, and 11 for cost. For each dimension a top-3 of most important indicators has also been

selected. Per outcome indicator, detailed information is provided on 16 elements, including purpose, characteristics, definition, rationale, potential uses, outcome measures and tools. NTOIP was designed to improve the scientific quality of evaluations of eHealth applications by providing guidance on specific outcome indicators. It is based on the results of an extensive literature review followed by a national experts workshop, and is aimed primarily at academic researchers.

2) MAST

In the European Union, *Model for the Assessment of Telemedicine Applications* [9] follows a similar approach. The model contains three elements: preceding considerations (to determine whether it is relevant to carry out an assessment), a multidisciplinary assessment (to describe and assess different outcomes of an eHealth application), and a transferability assessment (to assess the transferability of study results from one setting to another). In the multidisciplinary assessment, eHealth applications are evaluated in terms of seven domains, ranging from safety, clinical effectiveness, and patient perspective to economic, organizational, and socio-cultural aspects. Each domain is defined, and issues to consider within each domain are listed. No detailed guidelines are given with regard to study designs, methods, and outcome measures, although a manual is provided with examples of outcome measures for each domain, and methods for data collection. The development of MAST was initiated by the European Commission and carried out by a consortium of academic institutions using consensus building workshops with experts and decision makers. Its aim is to provide guidelines for a consistent assessment of eHealth outcomes, primarily for academic researchers performing evaluation studies to inform decision makers in health care.

3) CADTH Economic Guidelines

Economic evaluation (i.e., "value for money" analyses) of eHealth applications is an important area and specific frameworks have been developed for this purpose. An example of a rigorous framework is *Guidelines for the Economic Evaluation of Health Technologies* [10] by the Canadian Agency for Drugs and Technologies in Health. The third edition of the Economic Guidelines contains 61 guideline statements (do's and don'ts) on 14 topics: study questions, types of evaluations, target population, comparators (i.e., alternatives likely to be replaced), perspective (e.g., society), effectiveness, time horizon, modeling, valuing outcomes (e.g., in terms of QALYs: quality-adjusted life-years), resource use and costs, discounting ("translating" future costs and outcomes to present-day values), variability and uncertainty, equity ("fairness"), generalizability, and reporting. Five types of economic evaluation are discussed (cost-utility analysis, cost-effectiveness analysis, cost-minimization analysis, cost-benefit analysis, and cost-consequence analysis), selection of which depends on the research question, the health condition of interest, and the availability of data on outcomes. The framework is primarily aimed at experts performing economic evaluations to inform decision makers in health care.

4) Health IT Evaluation Toolkit

In contrast to the above frameworks, the *Health Information Technology Evaluation Toolkit* [11] by the U.S. Agency for Healthcare Research and Quality is primarily aimed at the non-expert. The toolkit provides step-by-step guidance for project teams who are developing evaluation plans for health IT projects. It does so by means of a comprehensive template guiding the team through the process of determining the goals of a project, what is important to its stakeholders, what needs to be measured to satisfy stakeholders, what items are realistic and feasible to measure, and how to measure these items. To this end, each step is accompanied by explanations and checklists. Furthermore, examples are provided of eHealth evaluation projects with suggested evaluation methodologies, as well as a detailed list of evaluation measures (with suggested data sources, cost considerations, potential risks, general notes, and references to the literature). These measures include clinical outcome measures, clinical process measures, provider adoption and attitudes measures, patient adoption, knowledge, and attitudes measures, workflow impact measures, and financial impact measures.

The Health IT Evaluation Toolkit was developed as part of a remote mentorship initiative by the AHRQ, to address common challenges experienced by health IT project teams at nonacademic institutions [12]. These challenges include: leaving evaluation as something to be determined "later", overscoped and unrealistic evaluation plans, a mismatch between the health IT being implemented and the evaluation measures chosen, searching for relatively rare events without the required statistical power, incorrect initial assumptions about data quality and feasibility of data collection, using an improper comparison group, insufficient attention to the details on data collection and analysis, and a lack of consideration of qualitative methods. It is very likely that the same or similar challenges are also faced by eHealth SMEs when they set out to collect evidence.

B. No generally accepted evaluation practice

Other frameworks were found in the literature as well. Like the examples discussed above, most are comprehensive and detailed. With a few exceptions, they are developed for academic researchers or experts informing decision makers in health care. They offer guidance – at least, once the aim of evaluating an eHealth innovation has become relatively clear – and a collection of outcome indicators, measures, methods and tools with clarifications to make informed choices. However, none of the frameworks found provide the same detailed and comprehensive guidance with regard to identifying the various stakeholders involved in embedding an eHealth innovation into routine care, their interests in the innovation, and subsequently the aim of an evaluation and the kinds of evidence that may be required. Possibly, these choices are too simply dependent on the specific case at hand to establish generic guidelines.

Furthermore, there seems to be a general consensus in the literature that there are currently no commonly accepted standards for collecting evidence for eHealth applications [e.g., 13, 14]. Some researchers have argued that the

assumptions, methods, and study designs of experimental science may altogether be less suited for application in the socio-political context in which eHealth evaluations usually take place, and that alternative approaches that view evaluation as social practice rather than scientific testing need to be considered [15]. Others argue that the tendency to focus on "hard" evidence as provided by randomized controlled trials may result in a disregard for the interests and experiences of the individual patient [16]. The Council for Public Health and Health Care in the Netherlands, which advises the Dutch government on health care policy, concurs [17]: *"The advance of evidence-based medicine has brought about a lot of good things.* [...] *However, there are disadvantages as well. The emphasis on scientific evidence may lead to a diminished appreciation for types of care where acquiring such evidence is impracticable, and to a neglect of elements such as personal attention, trust and 'presence'. The ethical question of what constitutes good (and reimbursable) care is being reduced to the question of what has been proven effective.* [...] *We must ask ourselves the question: what kind of evidence fits what kind of care? Hard evidence where it is attainable, but for other types of care where it is not realistic, different requirements need to be imposed. In short, we need to look for suitable evidence."*

Some researchers argue for a contextualized approach in which all relevant stakeholders are actively involved in the definition of the outcome indicators that will be used for evaluation [18, 19]. These findings were confirmed later on during workshops and interviews with representatives of the participating organizations; they have become a cornerstone in the approach that has been developed.

C. Three main themes for evidence

During the expert session with representatives from health care providers, insurers, patient organizations, and national health care authorities, three dominant themes were recognized by the participants within the larger concept of evidence: *effectiveness* ("did health care get any better?"), *cost efficiency* ("did it get any cheaper?") and *labor savings* ("did it get any less labor intensive?"). Below we briefly describe each theme, including a few relevant issues mentioned by the participants.

1) Effectiveness

This kind of evidence relates to clinical effectiveness, quality of care, safety, accessibility, timeliness, and patient satisfaction. However, eHealth's primary purpose may not always be patient recovery; frequently, eHealth is directed at retaining autonomy, strengthening the involvement from relatives, maintaining social participation, or improving a patient's wellbeing. Although these aspects are hard to measure, they are important from the patient's perspective and also valued by care professionals and society as a whole.

2) Cost efficiency

This includes evidence with regard to cost savings, cost control, and efficiency in terms of time, money, and other resources. eHealth applications have traditionally been considered as a promising way to reduce the cost of delivering health care. With the growing emphasis on budget control in health care, evidence for eHealth's cost efficiency is becoming increasingly relevant for decision makers. The current Dutch health care policy, for instance, is directed at stimulating cost-efficient eHealth applications that are replacing (instead of supplementing) traditional forms of care [4].

3) Labor savings

This relates to evidence that the same number of patients can be treated with the same quality, but with fewer hours worked by health care professionals. Although labor savings might be considered a special case of cost efficiency, the predicted labor shortage in the Dutch health care system justifies this kind of evidence to be considered separately. Labor savings also occur when an eHealth application reduces the complexity of a particular task, allowing highly schooled professionals to delegate part of their work to less skilled staff, or when applications directed at self-management stimulate or allow patients and their informal caretakers to assume an active role in a treatment.

Various outcome indicators and methods relating to the above themes of evidence have been identified during the session and also from the literature. They have been compiled into three detailed overviews, i.e., one for each theme. The results of this are reported elsewhere [20].

D. Putting evidence into perspective

During the same expert session with representatives of health care organizations it became clear that strong forms of evidence (obtained using, for instance, randomized controlled trials) are certainly not always necessary to facilitate the uptake of eHealth applications. The participants agreed that randomized controlled trials are not always useful, necessary, or practically feasible. Furthermore, care providers and health care insurers indicated that they will still rely on their own patient data to support any decisions they make about embedding eHealth applications.

National care authorities, on the other hand, hold the view that eHealth applications typically only change the way in which health care is being delivered. As long as there are no indications that safety or clinical effectiveness are at stake, and within the limits defined by regulations governing the provision of health care, care providers and health care insurers are free to negotiate and decide about the use (and reimbursement) of eHealth applications.

E. "Innovation routes" for eHealth innovations

One topic which arose very prominently during the expert session, is that it is not straightforward which path an SME should follow within the Dutch care system to get an eHealth innovation embedded into routine care. In part this is due to the wide variety of applications that fall under the common denominator of eHealth, but it is also due to the complexity of the Dutch care system, which is highly regulated and in which various authorities and other parties each play a distinct role. An SME should consider very carefully which "innovation route" to follow, as the chosen route will determine which stakeholders to address and involve. Stakeholders will have their own roles, responsibilities and interests, and hence will need their own arguments to get convinced of an eHealth application's

added value. It is, therefore, the chosen innovation route that determines the context in which evidence will be collected, the purpose for which it is collected, and the requirements that it should satisfy.

Based on the above findings, a review of online documentation pertaining to innovation in the Dutch health care system took place (e.g., [21-25]), and follow-up interviews with representatives of the participating health care organizations were organized. These efforts resulted in a comprehensive description of the Dutch health care system, including the roles of the parties involved, their interests in eHealth innovation, and criteria they use to evaluate eHealth innovations. Four main innovation routes were identified and described, including the specifics of each route and criteria for when to choose which route:

- The *consumer route* where an eHealth application is offered to and paid by patients/consumers. For example, a medical translation app that can be used when visiting a doctor abroad.
- The *provider route* where an application is offered to and paid by health care providers. For instance, an online treatment plan which allows clients to consult their plan and report about their progress.
- The *insurer route* where an application becomes part of an existing treatment that is offered by a care provider and reimbursed by a health insurance company. For example, a real-time medication monitoring service to improve the medication adherence of a diabetes patient. (In this case, the medication is the existing treatment and real-time monitoring becomes part of it.)
- The *government route* where an application leads to a new treatment not yet offered by care providers or reimbursed by health insurance companies, and where health care authorities need to decide whether it should be admitted to publicly insured care. Here, an example might be the introduction of telemonitoring of epilepsy patients in the home environment, to respond quickly in the event of a major seizure.

More details about the innovation routes are presented in Section V of this paper, and elsewhere [20].

IV. RESULTS: CASE STUDIES

This section highlights the results and lessons learned from the second phase, the case studies. During this phase, eight cases submitted by seven SMEs were selected for in-depth, semi-structured interviews. Table I lists relevant details per case.

The selected cases represented a variety of eHealth applications, including telemonitoring services, electronic health records, telemedicine, and electronic consultation. Although the SMEs would usually mention a combination of intended impacts, the primary intended impact was most frequently on effectiveness or cost efficiency; in one case it was on labor savings. Nearly all applications were designed to be used within the cure or care domains; one was aimed at prevention. None of the cases concerned applications that would lead to new treatments; instead, all were intended to impact the way in which care is being organized or existing treatments are being provided. With one exception, all were in the latest stages of development: pilot and roll-out.

In the following sections we discuss the main findings; detailed results per case are reported elsewhere [26].

A. The paths followed

The interviewed SMEs followed paths which can, in terms of the above innovation routes, be classified as either "provider routes" (where an application is offered to and paid by the care provider) or "insurer routes" (where an application becomes part of a treatment already offered by care providers and reimbursed by insurers). In the first case, the SMEs had indeed identified and involved the care provider as the main stakeholder (i.e., the party using the application and paying for it), in the second case the main stakeholders identified and involved were the care provider (the party using the application) and the health insurance company (the party reimbursing it). Sometimes other stakeholders were identified as well, for instance a patient organization (for endorsement or financial support) or a professional association (for approval).

In two cases the provider route had been followed, and in both cases with success. However, these cases had more in common than just the route followed: both innovations concerned electronic health records and were intended to organize the provision of care in a more efficient or client-centered way, the involved entrepreneurs had themselves a background in health care, they had developed their

TABLE I: RELEVANT DETAILS PER CASE.

Case	Type	Phase	Domain	Route	Stakeholders	Impact	Evidence
1	telemonitoring	roll-out	care	insurer route	care provider, insurer	cost efficiency	trial, business case
2	electr. health record	roll-out	care	provider route	care provider	cost efficiency	
3	electr. consultation	pilot	cure	insurer route	care provider, insurer, professional association	effectiveness	
4	telemonitoring	pilot	care	insurer route	care provider, insurer, patient association	effectiveness	trials (ongoing)
5	electr. health record	roll-out	cure, care	provider route	care provider	effectiveness	
6	telemedicine	roll-out	cure	insurer route	care provider, insurer	labor savings	trial, business case
7	telemonitoring	inventory	prevention	insurer route	care provider, insurer	cost efficiency	
8	telemonitoring	roll-out	cure	insurer route	care provider	effectiveness	trial

applications to address trends and challenges readily recognized by care providers, and they had done so in close cooperation with health care professionals. Hence, they needed little or no evidence to convince care providers: an appealing vision on trends in health care or challenges faced by the care provider, or a concise business case identifying the main costs and benefits, combined with the option to let professionals gain hands-on experience with the application (using a demonstrator setup, or by means of trial licenses) sufficed.

Compared to the paths classified as "provider routes", the paths classified as "insurer routes" tended to be less clear and more challenging. The key issues encountered by SMEs following this route are:

- Entrepreneurs with little or no experience in the health care sector often had difficulties in identifying a successful innovation route. The paths they followed were frequently based on trial and error, during which they steadily built up a better understanding of how the health care system works.
- The role of health insurance companies in the health care system, their interests in health care innovations, and the criteria by which they evaluate eHealth innovations were often unclear to the SMEs.
- SMEs tended to involve health insurance companies too early, when strong support among care providers, endorsements from patient organizations, or approvals from professional associations were still lacking. Insurance companies, on the other hand, used these as principal criteria for the selection of promising innovations.
- Health care providers and health insurance companies often had partly conflicting interests, making it difficult to come up with a business case which was compelling to both parties at the same time.
- Within this route, clinical trials were often essential to build up evidence for an innovation's effectiveness. Generally, SMEs lacked the expertise and financial resources to carry out a proper trial, forcing them to involve experts and to find sponsoring. Furthermore, it was not always clear exactly what evidence was required.

At the moment the interviews took place, none of the six SMEs following the insurer route had yet achieved success. In three cases this was because plans for pilots were still being made or trials were still ongoing. However, in three cases evidence had been collected in trials – yet in all three cases it failed to convince important stakeholders.

B. The evidence collected

In all three cases where evidence had been collected in trials, this had been done using randomized controlled trials – the "golden standard" for clinical trials [27]. Two of these cases concerned the application of real-time medication monitoring (RTMM) to improve medication adherence, albeit for different patient groups and in different contexts: in one case (no. 1) for patients with diabetes type 2 [28] and in the other (no. 8) for patients with refractory epilepsy [29]. The third case (no. 6) concerned the application of a computer-assisted therapy for patients with knee or hip replacements [30]. This application had been developed by a large German firm, and the clinical trial had already been performed in Germany; the participating SME was now trying to convince care providers and health insurers in The Netherlands of its added value. Table II provides more details about the designs, methods and outcome indicators used in each of these cases.

In all three cases, the trials were designed and performed by (or in close cooperation with) experts from scientific institutes. These experts assumed responsibility for deciding which evidence was to be collected and how this should be done. As said, in all cases (including case no. 4, where evidence was still being collected) they opted for randomized controlled trials in which a combination of objective measures and subjective judgments were being used. However, with the exception of case 1 (where the innovation department of a health insurance company was closely involved) it remained unclear to which extent external stakeholders – such as decision makers in care providers or insurance companies – had been consulted before these choices were made.

In cases 1 and 6 the results from the trial were further developed by the SME into a business case for stakeholders; in case 1 this was again done in close cooperation with a health insurance company. Table II includes information on how trial results were translated into a business case for stakeholders. The approach followed in case 1 is illustrative: the effect that was found on the intermediary outcome measure used in the trial (an increase in medication adherence) was first translated into an effect on a relevant end measure (a reduction in health related costs) using the results of a systematic review found in the scientific literature. The reduction in health related costs that was calculated, was then translated into a reduction in insurance claims for the health insurance company using the results of an internal study performed by the health insurance company. In this way, the clinical trial could focus on an intermediary outcome measure where effects could be measured on a much shorter time scale.

In none of the cases were the obtained results sufficient to convince the main identified stakeholders. The reasons for this varied. In the case 1 the trial had shown satisfactory results; however, the health insurer's purchasing department required that an additional economic analysis be carried out before it could support a nationwide adoption and reimbursement of the innovation. Unfortunately, the parties involved were unable to reach a consensus about this. In case 8 the care provider was convinced of the benefits for the treatment of its patients, yet the insurance company (having been requested to reimburse the use of the innovation) was unconvinced by the effects found during the trial. In case 6, the implementation model proposed by the SME (which entailed the establishment of new clinics where care provision would be centered on computer assisted therapy) did not satisfy the insurer's purchasing policy and so the lack of success had little to do with the evidence collected.

TABLE II: DETAILS ABOUT THE THREE CASES WHERE EVIDENCE HAD ALREADY BEEN COLLECTED.

Case	Patient group	Design	Outcome indicators	Conclusions	Business case
1	161 diabetes type-2 patients, randomly divided over a control group (no RTMM) and two intervention groups (RTMM with SMS-reminders; RTMM without SMS-reminders)	1. pre-measurement 2. intervention 3. post-measurement 4. follow-up duration of the intervention: 6 months	*medication adherence:* • refill adherence (determined from pharmacists' delivery registrations); • days without dosing, percentage missed doses, and percentage of doses taken within standardized time intervals (determined by real-time registration of medication intake) *experiences of patients:* • judgments on medication adherence (determined using surveys) • judgments on satisfaction (determined using surveys)	*medication adherence:* • significantly higher refill adherence (+26,5% for RTMM with SMS-reminders, +15,3% for RTMM without SMS-reminders, +10,5% for control group); • patients receiving SMS-reminders took significantly more doses within predefined time windows than patients receiving no reminders *experiences of patients:* • positive experiences of patients with the use of RTMM and with SMS-reminders	translation into financial consequences for a health insurance company, in two steps: 1. increased medication adherence (in %) was translated into a reduction in health related costs (in %), based on a published systematic literature review into this relation; 2. the reduction in health related costs (in %) was translated into a reduction of insurance claims (in €), using an internal study performed by the insurer into the costs of hospitalizations of diabetes patients
6	274 rehabilitating patients with hip or knee replacements, randomly divided over a control group (conventional therapy) and an intervention group (computer assisted therapy)	1. pre-measurement 2. post-measurement duration of the intervention: 6 months	*effectiveness:* • judgments determined using standardized surveys (Harris Hip Score, Hospital for Special Surgery Score, FIM instrument, and Hanover Functional Ability Questionnaire) *acceptance by patients:* • determined using surveys	*effectiveness:* • no statistically significant difference in effectiveness between conventional therapy and computer assisted therapy *acceptance by patients:* • positive acceptance by patients of computer assisted therapy	translation into labor savings for care providers: the intensity of the therapy can be increased by 20% with the same number of hours worked by therapists (further details are unknown)
8	28 epilepsy patients, randomly divided over two intervention groups (RTMM with SMS-reminders; RTMM first without and then with SMS-reminders)	1. pre-measurement 2. intervention 3. post-measurement duration of the intervention: ranging from 2 months to 2 years	*medication adherence:* • percentage missed doses and percentage of doses taken within standardized time intervals (determined by real-time registration of medication intake) *experiences of patients:* • judgments on medication adherence, seizure frequency and intensity (determined using surveys); • judgments on quality of life (determined using the QOLIE-31 survey); • seizure frequency (determined using diaries)	*medication adherence:* • significantly higher medication adherence for RTMM with SMS-reminders compared to RTMM without SMS-reminders: +14,4% correct intakes, and -10,8% missed doses *experiences of patients:* • positive experiences of patients (higher medication adherence, reduced seizure frequency or intensity; • no significant effect on quality of life except for higher overall health; • diaries showed no clear reduction of seizure frequency	translation into benefit for the care provider: increased effectiveness of the treatments offered to patients (no detailed business case had been developed yet)

The few cases described above are clearly insufficient to draw any firm conclusions. Yet, judging from these cases, it does seem necessary that decision makers (for instance in health insurance companies, but also in other stakeholders) are more closely involved when an evaluation is being planned. In this way, the criteria that play a role in the decision process can be clarified early on, when they can still

be taken into account in the development of evaluation plans or business cases, or in the design of clinical trials.

C. The criteria used by the insurance company

As mentioned before in Section II.E, three participating SMEs (i.e., cases 3, 4 and 6) accepted the open invitation to enroll their innovations at the health innovation desk of one of the participating health insurers. These three SMEs had all been following the insurer route.

Over the course of the ensuing procedures it became clear that three criteria are essential for the insurer: (i) is there sufficient support for the innovation among care providers (for instance, does it address any evident needs or demands), (ii) does the innovation fit into existing health care processes, and (iii) will it be able to substitute for existing forms of care. Other important criteria were: (iv) is the innovation fully developed, (v) is it fully interoperable with existing systems (for instance, systems in use by general practitioners), and (vi) is the potential for a nationwide adoption clear. Evidence for an innovation's effectiveness was among the criteria listed by the insurer, yet in these three cases it did not seem to have played a central role. However, a detailed business case supported by relevant figures was a clear requirement used within all three cases.

Of the three enrollments, one (i.e., case 4) ultimately led to a follow-up in which the involved SME, a care provider, and the insurer are currently planning a joint evaluation study of the innovation. In this particular case:

- The SME had gained clear support from a care provider, who eventually took over the negotiations with the insurance company;
- The application was fully developed and interoperable, and had already been used on a small scale in several pilots;
- The SME and the care provider were able to come up with a detailed business case showing how the innovation would be implemented in existing care processes, and where it would lead to substitution of existing care.

Summarizing, most of the criteria used by the insurer seem to be driven by a concern to identify early on which innovations will most likely be successfully implemented. However, the principal criterion is cost reduction by means of substitution: an eHealth innovation should either lead to the replacement of an existing form of care by a more cost efficient one; or, by being more effective, it should contribute to a reduced health care consumption in the near future. To convincingly show this to the health insurer, a detailed quantitative business case will be required.

V. RESULTS: GUIDELINES AND BEST PRACTICES

This section highlights the results and lessons learned from the third phase, guidelines and best practices. In the preceding phases of the project it had become clear that, when evaluation plans or clinical trials are being planned, relevant stakeholders should be identified and there interests taken into account. This is especially important because, ultimately, the evidence that is collected will be constituting the foundation beneath a business case in which all relevant stakeholders and their interests are accounted for. Preferably,

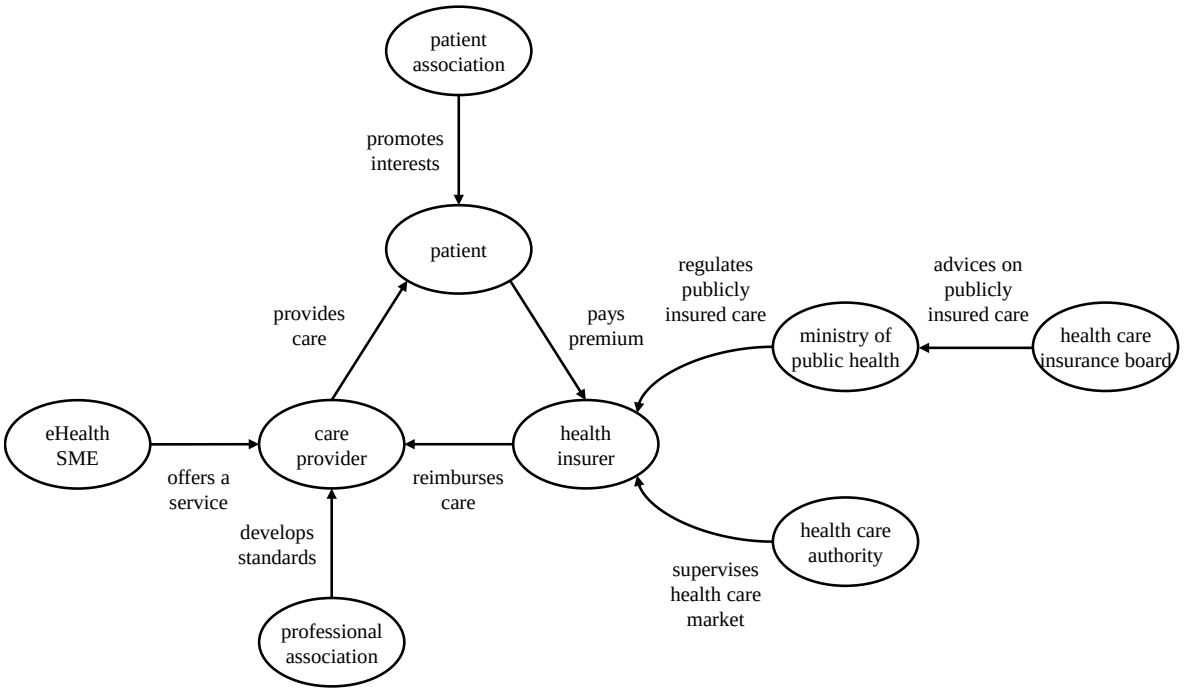


Figure 1: Elementary version of the eHealth innovation map, showing the main parties in the Dutch health care system. Each of these parties and their interests in eHealth innovations are further described in accompanying fact sheets.

principal stakeholders should be involved as early as possible, and the required evidence defined and collected in a cooperative effort.

To facilitate this, eHealth SMEs required a "map": to find the most promising innovation routes within the Dutch care system, and to identify relevant stakeholders and their interests. Creating such a map, and complementing the map with "fact sheets" (detailed yet concise and accessible information on innovation routes, relevant stakeholders and their interests, and types of evidence required) became the project's highest priority.

A. The eHealth innovation map

The starting point when developing the innovation map was that it should provide concise yet accessible information for SMEs on (i) the Dutch health care system, (ii) the roles of the main parties within it, (iii) the interests these parties have in eHealth innovations, and (iv) examples of applicable evidence to convince them. Furthermore, the map should visualize the four innovation routes and so facilitate the identification of relevant stakeholders. The map should provide only an overview; detailed information with guidelines and best practices was to be provided in sets of accompanying fact sheets (of one page each): a set on the innovation routes, a set on the stakeholders involved, and a set on applicable evidence. The following paragraphs briefly discuss each of these elements.

1) The innovation map

Figure 1 shows the innovation map in its elementary version, displaying only the main parties in the Dutch health care system and the relations among them. Care has been taken to streamline the map without oversimplifying it. Three thematic versions of the innovation map display additional information: one shows the various stakeholders within each party, one the interests that stakeholders may have in eHealth innovations, and one the kinds of evidence (or other applicable forms of proof) that may be used to convince them. For practical reasons these three thematic versions are not shown here; instead the information has been summarized in Table III. Last, there are four thematic versions displaying the identified innovation routes; these versions are shown in Figure 2. Each version is accompanied by a brief description of what is shown. In this way, SMEs are provided with "at a glance" information which acts as an index to the accompanying sets of fact sheets.

2) Fact sheets on stakeholders

Each party is described in more detail in its own fact sheet. These fact sheets contain concise information on (i) the role of this party in the health care system, (ii) relevant stakeholders within this party that may play a role in decision making, (iii) their interest (or interests) in eHealth innovations, and (iv) general guidelines on how (and by what means) this party can be convinced. For instance, the fact sheet on the insurer mentions that the insurer will reimburse care that has been provided to patients and not the use of eHealth applications as such. It mentions the roles and interests of its innovation, investment, purchasing and commercial departments, for instance, the partly conflicting interests of innovation department (interested in the potential of new developments) and the purchasing department (interested in the efficiency of concluding large contracts). It also mentions the main criteria by which these parties will evaluate an eHealth application, listing critical success factors such as clear support among care providers, and a detailed business case showing the potential for substitution or cost reduction. Table IV shows a representative example of a stakeholder fact sheet.

3) Fact sheets on innovation routes

The four innovation routes are also described in their own fact sheets. These fact sheets contain information on (i) situations where a particular route is applicable, (ii) matters to take into account when following a route, (iii) special circumstances or regulations that may play a role, (iv) the main stakeholders that need to be involved, and (v) the main anticipated risks (pitfalls). The fact sheet on the insurer route, for example, mentions that this route is appropriate when an eHealth application leads to a new way of organizing care, or a when it will be used for providing a treatment that is already being offered by care providers and reimbursed by insurers. It mentions the importance of first

TABLE III: SUMMARY OF THE INFORMATION DISPLAYED IN THE THREE THEMATIC VERSIONS OF THE INNOVATION MAP.

Party	Stakeholders within party	Interests in eHealth innovations	Evidence or other applicable forms of proof
care provider	nurses, specialists, ict, board	"will it improve the quality or efficiency of my work?"	a strong vision, a sound business case, experiences of care professionals
patient	partner, parents, children, informal carers	"will it aid in my recovery or improve my well-being?"	advice of care professionals, experiences of other patients
health insurer	innovation, purchasing, commerce, investment	"is it effective care for a reasonable price?"	a business case supported by the results of a clinical trial
professional association	medical specialists	"is it safe and well-founded?"	results of clinical trials, experiences of care professionals
patient association	patients, advisors	"will it improve the quality of care for our patient group?"	results of clinical trials, experiences of the patient group
health care authority	advisors	"is there support for it among care providers and insurers?"	joint innovation requests from care providers and insurers
health care insurance board	advisors	"has it been proven effective?"	viewpoints of professional associations, scientific research results of the highest possible quality
ministry of public health	policy makers	"does it contribute to affordable care?"	positive advice by the health care insurance board

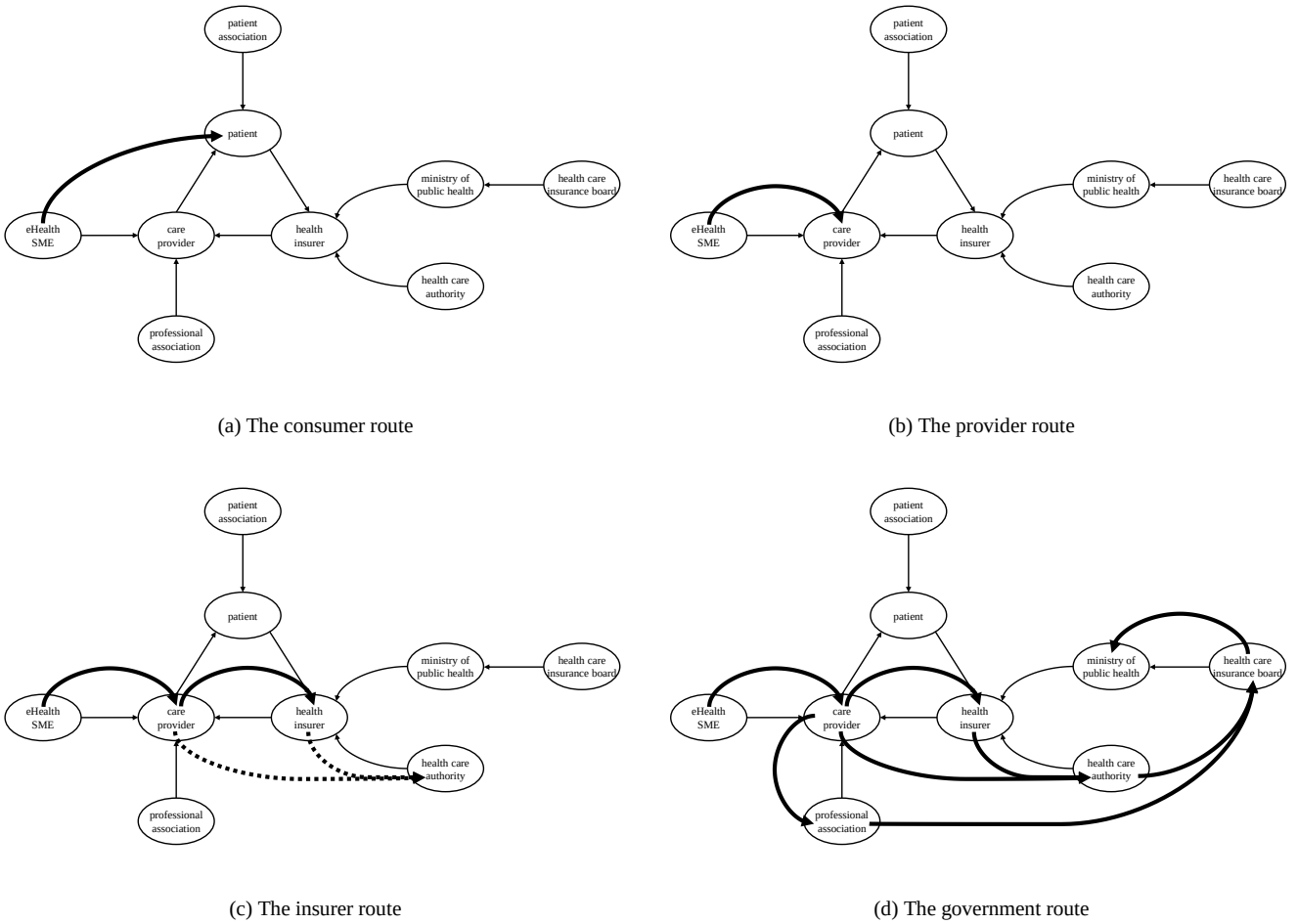
gathering sufficient support among care providers, professional associations and patient associations before turning to the insurer, and it mentions the special circumstance where a care provider and an insurer should jointly file an innovation request to the Dutch health care authority before a pilot can be started. Table V shows a representative example of an innovation route fact sheet.

4) Fact sheets on evidence

The third set of fact sheets concerns the evidence that will be required to convince the main stakeholders along each of the four innovation routes. The information provided in these fact sheets is necessarily generic; details on exactly which evidence to collect will depend on the specific situation (e.g., the type of eHealth application, where it is being used and to what effect, and the specific interests of relevant stakeholders). The fact sheets therefore contain (i) a concise description of the kinds of effects that need to be demonstrated for the main stakeholders, (ii) examples of the kinds of evidence that may be applicable, (iii) a few generic guidelines and best practices on how to collect evidence, and (iv) references to relevant sources of information, such as the frameworks discussed earlier in Section III.A. For instance, in the insurer route cost savings are mentioned as a principal effect to be established for the insurer. To this end, it should be shown how the eHealth innovation leads to the replacement of an existing form of care by a more cost efficient one, or how the application contributes to reduced health care consumption in the near future. Methods to establish these effects are mentioned such as, in the first case, a process analysis comparing care processes before and after introduction of the application or, in the second case, a clinical trial followed an elaboration of the effects found in a detailed, quantitative business case. Given the lessons learned during the project, the main best practice mentioned in this particular example is to closely involve key decision makers in the care provider and the insurer in order to clarify the criteria that will play a role in the decision process. Table VI shows a representative example of an evidence fact sheet.

B. Validation of the innovation map

Validation of the eHealth innovation map and the corresponding fact sheets has been (and is currently being)



(a) The consumer route

(b) The provider route

(c) The insurer route

(d) The government route

Figure 2: Thematic versions of the innovation map showing the four innovation routes. Thick arrows represent subsequent steps that should be undertaken by the SME or other involved stakeholders. Each version is accompanied by a descriptive fact sheet.

performed along four different lines:

- First, experts from the participating health care providers, patient associations, and government organizations have been asked to carefully check the map and the fact sheets for correctness and completeness of the provided information. Several corrections and suggestions have been made by them, which have subsequently been incorporated into the materials.
- Second, the usability and usefulness of the map and fact sheets have been evaluated with representatives from eHealth SMEs during a series of workshops where the eHealth innovation map was applied to a range of different cases at hand. In this way, a substantial amount of valuable feedback was collected and used to improve the materials.
- Third, validation of the map is currently being performed by means of "action research", where the research team is getting actively involved in a few selected cases (i.e., cases 4 and 8 from Table I, and one additional case contributed by a medical research center) with the aim to evaluate and extend the current insights. At the time of writing this research is still ongoing.
- Fourth, a number of successful cases are currently being analyzed by means of desk research and interviews with parties involved, to assess the innovation routes that have been followed and the evidence that has been collected. This, too, is at the time of writing still ongoing.

Due to the "mixed methods" approach followed during the project, with ample involvement of both experts and SMEs, a substantial amount of feedback and support has been collected for the innovation map. Overall, the responses given by these parties have been very favorable. Judging from the feedback that was given, the innovation map does indeed manage to provide a concise and accessible overview of the various ways in which eHealth innovations can be embedded in routine health care. The innovation route and stakeholder fact sheets are no "recipes for success", instead they are considered by both parties as highly useful scenarios

TABLE IV: EXAMPLE OF A STAKEHOLDER FACT SHEET. THIS ONE DISCUSSES THE INSURER. OTHERS DISCUSS THE CARE PROVIDER, THE PATIENT, THE PATIENT ASSOCIATION, THE PROFESSIONAL ASSOCIATION, AND THE GOVERNMENT ORGANIZATIONS. (ORIGINAL VERSION IN DUTCH. REFERENCES TO SOURCES HAVE BEEN OMITTED.)

Fact sheet healthcare insurer

Role

The healthcare insurer is the party paying for the care being provided to patients with the eHealth application. Keep in mind that there will be various stakeholders within the insurer, all with particular interests with regard to the eHealth application:

- The innovation department, where potential eHealth applications are selected and evaluated.
- The investment fund, which backs the development of eHealth applications financially.
- The purchasing department, which negotiates with care providers and purchases large quantities of healthcare (as efficiently as possible). Therefore, the role of eHealth applications it often limited.
- The commercial department, which sets up additional insurance packages for private parties and collective insurances for organizations and which sees eHealth as a distinguishing feature.

Keep in mind that any enthusiasm in the innovation department is not necessarily shared by the other stakeholders!

Interests

As far as healthcare insurers are concerned, what is most important is high-quality care at low cost, which translates into the following demands being made regarding eHealth applications:

- The application needs to have sufficient support among care providers and patients (through co-creation).
- The application must deliver healthcare gains (better quality care or higher quality of life).
- The application has to reduce healthcare costs (through increased independence on the part of the patient or reduced burden on the healthcare provider).
- The application has to lead to substitution (no extra care but substitution of existing care).
- The application has to lead to reduced health-related absence (prevention or quicker recovery).
- The application has to be in line with national agreements and purchasing policies.

Healthcare insurers do business with care providers, who they see as interlocutor, which means it is important to make sure that the application is suggested to the healthcare insurer by an enthusiastic care provider (rather than by the entrepreneur).

Persuasion

Healthcare insurers have medical advisers who will assess the added value of an application on the basis of their expertise. Generally speaking, they will demand to see a business case, based on financial estimates and supported by research results (for instance a clinical trial or pilot project).

A business case can be created in stages, for instance by translating the effects that have been detected in a pilot study into financial consequences for the healthcare insurer. Always determine the design of a pilot study or clinical trial (what is being measured, and how) together with the care provider and healthcare insurer.

to be explored by an SME and used as a means to create a stakeholder inventory. The evidence fact sheets are highly useful to start the discussion with stakeholders and experts when drafting plans for an evaluation.

At the time of writing there is a strong interest in the map. It has, for instance, been made accessible to a large audience via the website of the Netherlands Organization for Health Research and Development [23] and a well-known website maintained by a joint initiative of four government organizations (the Dutch Healthcare Insurance Board, the Dutch Healthcare Authority, the Ministry of Health, Welfare and Sports, and the Netherlands Organization for Health Research and Development) [24].

VI. Results: Consolidation and tool development

The fourth and last phase of the project, consolidation and tool development, is currently nearing completion. Based on the eHealth innovation map a workshop protocol has been developed, and the innovation map and the fact sheets have been incorporated into an interactive, web-based tool [31]. The workshop protocol and the web-based tool both aim to provide guidance to SMEs in finding a promising innovation route, in identifying relevant stakeholders to involve, and in determining which evidence they may require.

Last, the project's results have been documented in an accessible and illustrated booklet for SMEs [32]. The booklet summarizes all the information contained within the innovation map and the fact sheets, such as the descriptions of the main parties in the Dutch health care system, the identified innovation routes, the interests of various parties in eHealth innovations, and various kinds of evidence that may be required. It is hoped that in this way, the project's results will be well consolidated and accessible for all interested eHealth SMEs in The Netherlands.

VII. Concluding remarks

The main conclusion to be drawn from the research presented here, is that evidence constitutes the foundation underneath a business case in which all relevant stakeholders and their interests are accounted for. Preferably, principal stakeholders should be involved as early as possible when

Table V: Example of an innovation route fact sheet. This one discusses the insurer route. Others discuss the consumer route, the provider route, and the government route. (Original version in Dutch. References to sources have been omitted.)

Fact sheet insurer route

When does this route apply?
An eHealth application is integrated into care that is already being provided or reimbursed. The application does not alter the care being provided, only the form in which it is delivered. As a result, for example, the care becomes more accessible or it can be provided more efficiently.

Examples
- An online nutrition diary that is used as part of diet advice by a dietician and promotes the patient's self-management.
- A pillbox that alerts patients when they forget to take their medication. This takes place on doctor's order and promotes patient discipline.

Points of interest
Make sure there is sufficient support! It is important for care providers, patients and patient organizations to be enthusiastic about the application, which is why it is crucial to involve them at an early stage in the development (co-creation). The specialists' professional association plays an important role in nationwide up-scaling, because they determine the guidelines for good and safe care.

If an application leads to cheaper or less labor-intensive care, while the quality of the provided care remains the same at least, this is interesting for the care provider and it may not be necessary to involve the insurer. If, on the other hand, the application makes the care being provided more expensive, it has to be demonstrated that the quality of the care has improved and a larger support base is needed. Do not approach the insurer yourself, but let the enthusiastic care provider do the negotiations.

As far as insurers are concerned, it is crucial for the application to lead to a replacement of existing care (for instance through substitution or self-management) and, ultimately, to a reduction in reimbursements. It is important to demonstrate this in a detailed business case.

Special details
If an application does not match the existing care descriptions defined by the Dutch Healthcare Authority (for example due to restrictions in the description or rate), the care provider and insurer together can submit an application at the Dutch Healthcare Authority. The Dutch Healthcare Authority can modify an existing care description or create a temporary one, giving the application time to "prove" itself.

The main stakeholders
- Care provider and professional association
- Patients and patient association
- Care insurer
- Dutch Healthcare Authority (if a care description needs to be modified or a temporary one created)

Pitfalls
Creating insufficient support (among patients, care providers, patient associations and professional associations). Approaching the insurer yourself without the backing of at least one care provider. Paying insufficient attention to the substitution of the existing care.

TABLE VI: EXAMPLE OF AN EVIDENCE FACT SHEET. THIS ONE DISCUSSES EVIDENCE FOR THE INSURER ROUTE. OTHERS DISCUSS EVIDENCE FOR THE CONSUMER ROUTE, THE PROVIDER ROUTE, AND THE GOVERNMENT ROUTE. (ORIGINAL VERSION IN DUTCH. REFERENCES TO SOURCES HAVE BEEN OMITTED.)

Fact sheet evidence within the insurer route

What needs to be demonstrated?
A business case needs to be developed in which the interests of the care provider (see the provider route) and the health insurer are combined. Ultimately, healthcare insurers want to see a reduction in healthcare costs (through substitution or self-management), but they also focus on support among providers, scalability and compatibility with existing care processes. See the ZonMw website for a list of relevant criteria.

Which evidence is suitable?
Demonstrating a reduction in healthcare costs can be done in two ways:

1. By replacing expensive forms of care by less expensive ones ("substitution"). This leads to "definite", short-term cost reductions. Make clear to the insurer how the current care process will change and how this will lead to labor savings, process optimization, or lower costs. Pay attention to the aspects that will be included in the business case, and how this will be measured in a pilot or trial. Insurers will want to know how substitution is actually accomplished.
2. More effective care will lead to a reduction in care consumption in the long term, but the cost reduction is surrounded by uncertainty. Note that insurers will want to see a return on investment within three years. Reduced healthcare consumption will need to be demonstrated with methodologically sound research, for instance using this three-stage process: (1) a clinical trial aimed at measuring a process measure or intermediary measure, (2) translation of the effects found on the process or intermediary measure into an effect on a relevant end measure, based on the best available scientific evidence on the relation between these two, (3) calculation of the potential cost reduction based on insurer data. The Achmea Health Database is a good source of information to do this.

Some eHealth applications may be attractive for health insurer for commercial or marketing purposes (e.g., to attract or maintain subscribers). In such cases, contact the commercial department, which is responsible for additional insurances for consumers and collective insurances for organizations. In the latter case, it should be clear how the application can lead to fitter employees or reduced sick leave.

Things to keep in mind:

- In the case of improved efficiency, there has to be a clear (clinically relevant) improvement, which has to be demonstrated through scientifically sound research.
- Be careful about making assumptions, for instance in translating an intermediary measure (for instance, medication adherence) to an end measure (reduction or delay of complications). Do not add assumptions to assumptions.
- "Pick your battle": using a certain application may prove more beneficial with some syndromes compared to others. Think about this carefully.
- "Hard" data (which can be determined objectively) have more weight than "soft" data (opinions or experiences of patients and other people involved), no matter how they are collected. "Hard" data can also be obtained through routine registrations of care suppliers.

Important:

- Discuss as early as possible with the insurer and the care provider what evidence will be required.
- Involve important stakeholders, such as decision-makers, when working out the appropriate research approach.
- Consult experts when methodologically strong research is needed, but keep stakeholders involved.

planning an evaluation study or a (clinical) trial. In this way, the criteria that will play a role later on in the decision process can be clarified early on, when they can still be taken into account.

This insight has become the corner stone of the approach developed in the project "Successful Entrepreneurship in eHealth". Following this approach, the chosen innovation route, the identified stakeholders, and their interests in the eHealth innovation at hand eventually determine which kinds of evidence will be needed and how they should be collected. The developed eHealth innovation map, the workshop protocol, and the web-based tool were all developed to provide guidance to eHealth SMEs, allowing them to make better, more informed decisions. The design, implementation and analysis of clinical trials will nevertheless remain the domain of academic experts or highly trained staff members working at care providers; the level of expertise that is required makes this simply unavoidable.

The implications of this research are threefold. First, further research is needed to deepen and validate the insights gained so far, preferably by consistently applying the developed approach along the full development cycles of a number of eHealth services, and by evaluating the results. Since this is a process that can take several years to complete, this could not be done within the current project. Furthermore, academic experts involved in the development of frameworks to evaluate eHealth should place more emphasis an stakeholders' varying roles and interests, and incorporate these factors into their designs. Second, policy makers in both government and health organizations could use the identified innovation routes to identify any unwanted obstacles (for instance, conflicting requirements imposed by parties involved), take measures to alleviate barriers along a route, and initiate any required coordination between stakeholders, thus streamlining each route. Third, eHealth SMEs would do well to study the workings of the health care

system, for instance by using the proposed innovation map as a scenario building instrument and by applying it early on for stakeholder inventory and analysis.

Now that the project is nearing completion, the question arises how unique the Dutch situation really is. Can the eHealth innovation map be generalized to other countries? When an early concept of the innovation map was presented at an international eHealth conference [1] it seemed from the responses given by the international audience that certain basic principles, such as the roles and interests of the care provider and the insurer, are certainly generalizable. Other aspects, such as the government legislation pertaining to the health care system, will vary. Nevertheless, judging by this first impression it seems that the proposed approach may be fruitful for parties in other countries as well.

ACKNOWLEDGMENT

The authors wish to thank Stichting Innovatie Alliantie for financially supporting the project *Successful Entrepreneurship in eHealth*. Furthermore, they would like to thank the eHealth SMEs and all other organizations who participated in the project: Achmea divisie Zorg en Gezondheid, ActiZ, Be4Care, BeagleBoxx, Bonstato, College voor Zorgverzekeringen, Consuming Knowledge, DiaDerma, Dutch IT Consultants, Eusamed, Evalan, EvoCare, Focus Cura, Health Valley, Installatie Techniek Masseling, Isala Klinieken, MS Nederland, Nederlandse Zorgautoriteit, Novay, Nederlandse Patiënten Consumenten Federatie, Nederlandse Vereniging voor eHealth, SalesSpirit, SymaX, Syntens, Valetudo Interpres, Vos Projectadvies Gezondheidszorg, VRC Telecom, and Zorgverzekeraar VGZ.

REFERENCES

[1] R. Janssen, M. Hettinga, H. Prins, S. Visser, R. Menko, I. Krediet, T. Haaker, and L. Bodenstaff, "Developing evidence guidelines for eHealth small and medium-sized enterprises," Proc. eTELEMED 2013, Feb. 2013, pp. 92-95.

[2] J.E. van Gemert-Pijnen, N. Nijland, M. van Limburg, H.C. Ossebaard, S.M. Kelders, G. Eysenbach, and E.R. Seydel, "A holistic framework to improve the uptake and impact of eHealth technologies," J. Med. Internet Res., 2011, 13(4):e111.

[3] R.A. Menko, S. Visser, R. Janssen, M. Hettinga, and T. Haaker, "Applying the STOF business model framework in eHealth innovations," Proc. eTELEMED 2013, Feb. 2013, pp. 108-113.

[4] Koninklijke Nederlandsche Maatschappij tot bevordering der Geneeskunst, Nederlandse Patiënten Consumenten Federatie, and Zorgverzekeraars Nederland, "Nationale Implementatie Agenda (NIA) eHealth," Utrecht: Koninklijke Nederlandsche Maatschappij tot bevordering der Geneeskunst, 2012. ["National eHealth implementation agenda". In Dutch.]

[5] F. Mair, C. May, E. Murray, T. Finch, G. Anderson, C. O'Donnell, P. Wallace, and F. Sullivan, "Understanding the implementation and integration of e-health services," London: NHS Service and Delivery R and D Organisation (NCCSDO), 2009.

[6] F. Mair, C. May, C. O'Donnell, T. Finch, F. Sullivan, and E. Murray, "Factors that promote or inhibit the implementation of e-health systems: an explanatory systematic review," Bull. World Health Organ., 2012, 90:357-364.

[7] J. Hackos and J. Redish, "User and task analysis for interface design," New York: Wiley, 1998.

[8] R.E. Scott, F.G. McCarthy, P.A. Jennett, T. Perverseff, D. Lorenzetti, A. Saeed, B. Rush, and M. Yeo, "Telehealth outcomes: a synthesis of the literature and recommendations for outcome indicators," J. Telemed. Telecare, 2007, 13 Suppl 2:1-38.

[9] K. Kidholm, A.G. Ekeland, L.K. Jensen, J. Rasmussen, C.D. Pedersen, A. Bowes, S.A. Flottorp, and M. Bech, "A model for assessment of telemedicine applications: MAST", Int. J. Technol. Assess. Health Care, 2012, 28(1):44-51.

[10] Canadian Agency for Drugs and Technologies in Health, "Guidelines for the economic evaluation of health technologies: Canada [3rd edition]," Ottawa: Canadian Agency for Drugs and Technologies in Health, 2006.

[11] C.M. Cusack, C. Byrne, J.M. Hook, J. McGowan, E.G. Poon, and A. Zafar, "Health information technology evaluation toolkit: 2009 update," Rockville, MD: Agency for Healthcare Research and Quality, 2009.

[12] E.G. Poon, C.M. Cusack, and J.J. McGowan, "Evaluating healthcare information technology outside of academia: observations from the National Resource Center for Healthcare Information Technology at the Agency for Healthcare Research and Quality," J. Am. Med. Inform. Assoc., 2009, 16(5):631-6.

[13] A.G. Ekeland, A. Bowes, and S. Flottorp, "Methodologies for assessing telemedicine: a systematic review of reviews," Int. J. Med. Inform., 2012, 81(1):1-11.

[14] A.D. Black, J. Car, C. Pagliari, C. Anandan, K. Cresswell, T. Bokun, B. McKinstry, R. Procter, A. Majeed, and A. Sheikh, "The impact of eHealth on the quality and safety of health care: a systematic overview," PLoS Med., 2011, 8(1):e1000387.

[15] T. Greenhalgh and J. Rusell, "Why do evaluations of eHealth programs fail? An alternative set of guiding principles," PLoS Med., 2010, 7(11):e1000360.

[16] E. Cummings and P. Turner, "Patients at the centre: methodological considerations for evaluating evidence from health interventions involving patients use of web-based information systems," Open Med. Inform. J., 2010, 4:188-94.

[17] D.W. Willems, R. Vos, G. Palmboom, and R. Lips, "Passend bewijs. Ethische vragen bij het gebruik van evidence in het zorgbeleid," Den Haag: Raad voor de Volksgezondheid en Zorg, 2007. ["Suitable evidence. Ethical questions regarding the use of evidence in health care policy." In Dutch.]

[18] M.P. Gagnon and R.E. Scott, "Striving for evidence in e-health evaluation: lessons from health technology assessment," J. Telemed. Telecare, 2005, 11, Suppl 2:S34-6.

[19] H. Hyppönen, E. Ammenwerth, and N. de Keizer, "Exploring a methodology for eHealth indicator development," Stud. Health Technol. Inform., 2012, 180:338-42.

[20] T. Haaker, L. Bodenstaff, R. Menko, S. Visser, H. Prins, I. Krediet, M. Hettinga, and R. Janssen, "Inventarisatie eHealth innovatieroutes en evidence," Zwolle: Hogeschool Windesheim, 2013. ["Inventory of eHealth innovation routes and evidence." In Dutch.]

[21] Website Nederlandse Zorgautoriteit, www.nza.nl. [Website of the Dutch Healthcare Authority. In Dutch.] Last accessed: December 6, 2013.

[22] Website College voor Zorgverzekeringen, www.cvz.nl. [Website of the Dutch Healthcare Insurance Board. In Dutch.] Last accessed: December 6, 2013.

[23] Website ZonMw, www.zonmw.nl. [Website of the Netherlands Organization for Health Research and Development. In Dutch.] Last accessed: December 6, 2013.

[24] Website Zorg voor Innoveren, www.zorgvoorinnoveren.nl. [Website by a joint initiative of four government organizations: the Dutch Healthcare Insurance Board, the Dutch Healthcare Authority, the Ministry of Health, Welfare and Sports, and the Netherlands Organization for Health Research and Development. In Dutch.] Last accessed: December 6, 2013.

[25] Website Nictiz, www.nictiz.nl. [Website of the National IT Institute for Healthcare in the Netherlands. In Dutch.] Last accessed: December 6, 2013.

[26] R. Janssen, W. de Haan, T. Haaker, L. Bodenstaff, S. Visser, R. Menko, H. Prins, I. Krediet, and M. Hettinga, "Rapportage casestudies," Zwolle: Hogeschool Windesheim, 2013. ["Report about the case studies." In Dutch.]

[27] P.M. Ho, P.N. Peterson, and F.A. Masoudi, "Evaluating the evidence. Is there a rigid hierarchy?" Circulation, 2008, 118(16):1675-84.

[28] M. Vervloet, L. van Dijk, J. Santen-Reestman, B. van Vlijmen, P. van Wingerden, M.L. Bouvy, and D.H. de Bakker, "SMS reminders improve adherence to oral medication in type 2 diabetes patients who are real time electronically monitored," Int. J. Med. Inform., 2012, 81(9):594-604.

[29] M. Hettinga, T. Gutter, and J.T. te Gussinklo, "Real time medication monitoring with customized SMS reminders for people with refractory epilepsy," Proc. eTELEMED 2013, Feb. 2013, pp. 162-167.

[30] U. Eisermann, I. Haase, and B. Kladny, "Computer-aided multimedia training in orthopedic rehabilitation," Am. J. Phys. Med. Rehabil., 2004, 83(9):670-80.

[31] Website Innovatieroutes in de Zorg, www.innovatieroutesindezorg.nl. [A website and interactive assessment about four eHealth innovation routes, maintained by the Research group ICT Innovations in Health Care, Windesheim University of Applied Sciences. In Dutch.] Last accessed: December 6, 2013.

[32] R. Janssen, L. Bodenstaff, E. Gyaltsen-Lohuis, T. haaker, W. de Haan, I. Krediet, R. Menko, H. Prins, S. Visser, and M. Hettinga, "Succesvol ondernemen met eHealth: Innovatieroutes in de zorg," Zwolle: Hogeschool Windesheim, 2013. ["Succesful entrepreneurship in eHealth: innovation routes in health care." In Dutch.]

A WYSIWYM Interface for Semantic Enrichment of E-Prescriptions using Linked Open Drug Data

Ali Khalili* and Bita Sedaghati†
*Institute of Informatics, University of Leipzig, Germany
khalili@informatik.uni-leipzig.de
†Institute of Pharmacy, University of Leipzig, Germany
bita.sedaghati@uni-leipzig.de

***Abstract*—In this paper, we present an approach to enrich electronic prescriptions using linked open drug data. The proposed approach employs WYSIWYM (What-You-See-Is-What-You-Mean) interface for integrated authoring, visualization and browsing of semantic data in medical prescriptions. The generated semantic medical prescriptions serve as intelligent e-prescription documents enriched by drug-related meta-data thereby know about their content and the possible interactions. In an e-health system, semantic prescriptions provide an interoperable interface which helps patients, physicians, pharmacists, researchers and pharma companies to collaboratively improve the quality of pharmaceutical services by facilitating the process of shared decision making. In order to showcase semantic prescription we develop a mobile/web application called Pharmer. Pharmer provides different views for the different personas involved in the process of e-prescribing. It employs datasets in Linked Open Drug Data (LODD) such as DBpedia, DrugBank, DailyMed and RxNorm to automatically detect the drugs in a prescription and to collect multidimensional data on them. It also supports automatic prevention of possible drug interactions in a prescription.**

***Keywords*—Semantic prescription, e-prescription, WYSIWYM, semantic annotation, e-health.**

I. INTRODUCTION

In this article, we elaborate on the basic requirements of semantic prescriptions and will extend our previously published work [1] to address novel user interfaces as well as novel shareholders to take advantage out of the semantic medical prescriptions.

As reported in MedicineNet [2], *medication errors* are the most common type of medical errors in health care. Errors such as improper dose of medicine, adverse drug interactions, food interactions, etc. often stem from invalid prescriptions and unawareness of the patients. Medication-oriented errors are usually the result of failures during the medication process [3]. Electronic prescriptions which are recently gaining attention in the e-health domain, are one of the solutions proposed to solve these type of errors. In an e-prescription system, prescriber electronically sends an accurate, error-free prescription directly to a pharmacy from the point-of-care.

During the recent years, the adoption of e-prescriptions has been spreading relatively rapidly. In the US, the so called *Electronic Prescribing Incentive Program* is a reporting program that uses a combination of incentive payments and payment adjustments to encourage electronic prescribing by eligible professionals [4]. As recently published by [5], hospitals' use of computerized prescriptions prevented 17 million drug errors in a single year in the United States. The *Canadian Medical Association* (CMA) and the *Canadian Pharmacists Association* (CPhA) have approved a joint statement on the future of e-prescribing that aims to have all prescriptions for Canadians created, signed and transmitted electronically by 2015. The Australian government removed commonwealth legislative barriers to electronic prescribing started from 2007 [6]. A system called *epSOS* [7], which performs the use of e-prescriptions all around Europe, is currently passing the extensive practical testing phase.

However, one of the main challenges in current e-prescription systems is dealing with the heterogeneity of available information sources. There exist already different sources of information addressing different aspects of pharmaceutical research. Information about chemical, pharmacological and pharmaceutical drug data, clinical trials, approved prescription drugs, drugs activity against drug targets such as proteins, gene-disease-drug associations, adverse effects of marketed drugs, etc. are some examples of these diverse information. Managing these dynamic pieces of information within current e-prescription systems without blurring the border of the existing pharmaceutical information islands is a cumbersome task. On the other hand, *Linked Open Data* as an effort to interlink and integrate these isolated sources of information is obtaining more attention in the domain of pharmaceutical, medical and life sciences.

Combining the best practices from Linked Open Data together with e-prescription systems can provide an opportunity for patients, researchers as well as practitioners to collaborate together in a synergetic way. A consequence of introducing linked data in health care sector is that it significantly changes the daily duties of the employees of the health care sector. Therefore, the most challenging aspect will not be the technology but rather changing the mind-set of the employees and the training of the new technology [8]. Furthermore, the information generated via that approach can be employed as a data source for researchers. Drug companies are also able then to take the advantage of considering these informative statistical data.

Semantic prescriptions are a proposed approach to utilize semantic web technologies in e-prescription systems. As intel-

ligent prescriptions, they can automatically handle the medical errors occurring in prescriptions and increase the awareness of the patients about the prescribed drugs and drug consumption in general. Semantic prescriptions also enable the creation of more efficient and effective search approaches for drug discovery and consumption. We created a tool called *Pharmer* as a showcase application to facilitate the process of semantic prescription generation. Pharmer adopts a bottom-up approach for enriching the normal e-prescriptions with semantic annotations using a set of predefined datasets from Linked Open Data.

The remainder of this article is structured as follows: Section II, Section III and Section IV provide a background on the basic concepts such as Linked Open Data, Semantic Content Authoring and WYSIWYM (What-You-See-Is-What-You-Mean) interface employed in this paper. In Section V, we describe the process of semantic e-prescribing and will showcase Pharmer as a WYSIWYM UI to effectively create semantic prescriptions. Then we discuss the possible use cases of Pharmer in Section VI. To better clarify the use cases of the Pharmer system, an example scenario including Pharmer stakeholders is drawn in Section VI-C. Finally, Section VII concludes with an outlook on future work.

II. LODD Applications

In computing, *Linked Data* describes a method of publishing structured data so that it can be interlinked and become more useful. It builds upon standard Web technologies such as HTTP and URIs, but rather than using them to serve web pages for human readers, it extends them to share information in a way that can be read automatically by computers. This enables data from different sources to be connected and queried [9]. Tim Berners-Lee, the inventor of the Web and Linked Data initiator, suggested a 5 star deployment scheme for *Linked Open Data*: 1) make your stuff available on the Web (whatever format) under an open license, 2) make it available as structured data (e.g., Excel instead of image scan of a table), 3) use non-proprietary formats (e.g., CSV instead of Excel), 4) use URIs to identify things, so that people can point at your stuff, 5) link your data to other data to provide context.

A. Linked Open Drug Data (LODD)

Particularly in the areas of health care and life sciences with the wealth of available data, large scale integration projects like Bio2RDF [10], Chem2Bio2RDF [11], and the W3C HCLS's (Health Care and Life Sciences) Linked Open Drug Data (LODD) [12] have not only significantly contributed to the development of the Linked Open Data effort, but have also made social and technical contributions towards data integration, knowledge management, and knowledge discovery.

There are already many interesting information on pharmaceutical research available on the Web. The sources of data range from drugs general information, interactions and impacts of the drugs on gene expression, through to the results of clinical trials. LODD [13] has surveyed publicly available data about drugs, created Linked Data representations of the data sets, and identified interesting scientific and business questions

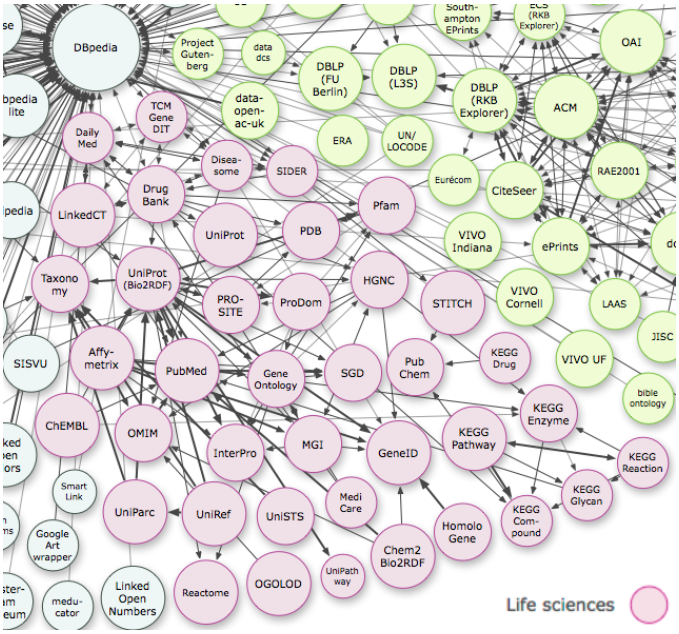


Fig. 1. Available datasets related to life sciences and pharmaceutical research.

that can be answered once the data sets are connected (cf. Figure 1).

B. LODD Applications in Medical Domain

There exists few approaches that address the medical and pharmaceutical applications using LODD. *TripleMap* (http://www.triplemap.com) is a project connecting widespread distribution of journal articles, patents and numerous databases in pharmaceutics research. TripleMap as a web-based application provides a dynamic visual interface to integrate RDF datasets such as the LODD. Showing an unexpected associations between entities related to researcher's interest is main advantage of TripleMap inspired by the broad interconnected data available in the LODD data sets. The goal of the TripleMap project is to deliver and sustain an 'open pharmacological space' by using and enhancing the state-of-the-art semantic web standards and technologies [14].

Another related project is the Open Pharmacological Space (OPS), *Open PHACTS* (Pharmacological Concept Triple Store http://www.openphacts.org) project under the European Innovative Medicines Initiative (IMI http://www.imi.europa.eu/). The goal of this project is integration of chemical and biological data using Linked Data standards to support drug discovery [15].

Linked Cancer Genome Atlas Database [16] as another Linked Data project aims to create an atlas of genetic mutations responsible for cancer. The project provides an infrastructure for making the cancer related data publicly accessible and to enable cancer researchers anywhere around the world to make and validate important discoveries.

Although these projects address the backend side of creating LODD applications, there has been a clear lack of applications with user-friendly, efficient and effective interfaces to make Linked Data resources accessible to end-users outside the biomedical community. One of the use cases of LODD datasets addressed in this paper is authoring of *Semantic Prescriptions* which are prescriptions enriched by Linked Open Data.

III. Semantic Content Authoring

A *Semantic Document* is an intelligent document (with explicit semantic structure) which "knows about" its own content so that it can be automatically processed in unforeseen ways. Semantic documents facilitate a number of important aspects of information management [17]. For *search and retrieval*, they provide more efficient and effective search interfaces, such as faceted search [18] or question answering [19]. In *information presentation*, they support more sophisticated ways of flexibly visualizing information, such as by means of semantic overlays as described in [20]. In *information integration*, they provide unified views on heterogeneous data stored in different applications by creating composite applications such as semantic mashups [21]. For *personalization*, they provide customized and context-specific information which better fits user needs and will result in delivering customized applications such as personalized semantic portals [22]. For *reusability* and *interoperability*, they facilitate exchanging content between disparate systems and enabling applications such as executable papers [23].

The above benefits, however, come at the cost of increased authoring effort. A *Semantic Authoring User Interface* is a human accessible interface with capabilities for writing and modifying semantic documents which are either. *Semantic Content Authoring* (SCA) is a tool-supported manual composition process aiming at the creation of semantic documents which are either:

- fully semantic in the sense that their original data model uses a semantic knowledge representation formalism (such as RDF, RDF-Schema or OWL) or
- based on a non-semantic representation form (e.g., text or hypertext), which is enriched with semantic representations during the authoring process.

With an ontology and a user interface appropriate for the type of content, semantic authoring can be easier than traditional composition of content and the resulting content can be of higher quality [24].

IV. WYSIWYM User Interface

The term *WYSIWYG* as an acronym for What-You-See-Is-What-You-Get is used in computing to describe a system in which content (text and graphics) displayed on-screen during editing appears in a form closely corresponding to its appearance when printed or displayed as a finished product. WYSIWYG text authoring is meanwhile ubiquitous on the Web and part of most content creation and management workflows. It is part of content management systems (CMS), weblogs, wikis, fora, product data management systems and online shops, just to mention a few. However, the WYSIWYG model has been criticized, primarily for the verbosity, poor support of semantics and low quality of the generated code and there have been voices advocating a change towards a *WYSIWYM* (What-You-See-Is-What-You-Mean) model [25], [26].

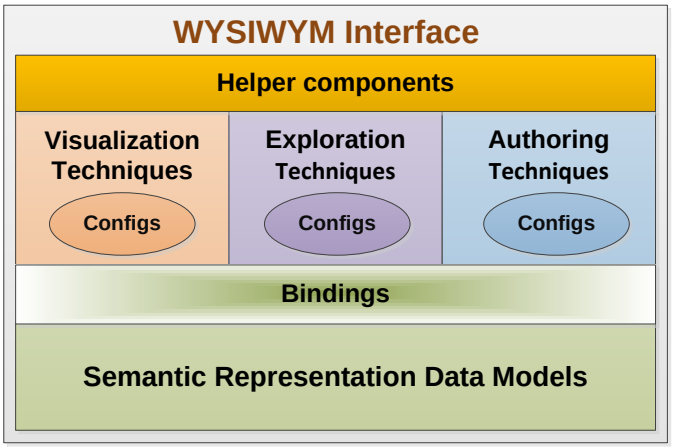


Fig. 2. WYSIWYM conceptual view.

Similar to the WYSIWYG UI, the WYSIWYM term as defined in [27] targets the novel aspect of integrated visualization, exploration and authoring of unstructured and semantic content. The rationale of our WYSIWYM concept is to enrich the existing WYSIWYG presentational view of the content with UI components revealing the *semantics* embedded in the content and enable the exploration and authoring of semantic content. Instead of separating presentation, content and meaning, our WYSIWYM approach aims to integrate these aspects to facilitate the process of *Semantic Content Authoring*.

WYSIWYM model as depicted in Figure 2, is defined by:

- a set of semantic representation data models (e.g., RDF), where each one has an associated set of data model elements;
- a set of visualization techniques (e.g., framing using borders and background) and a set of possible configurations for them;
- a set of exploration techniques (e.g., faceted browsing based on the type of entities) and a set of possible configurations for them;
- a set of authoring techniques (e.g., form editing, inline edit) and a set of possible configurations for them;
- a set of bindings to map each element of a semantic representation model to a visualization/exploration/authoring technique;
- a set of helper components (e.g., automation, real-time tagging, recommendation).

Medical prescriptions are a good candidate to be enriched by the WYSIWYM semantic annotations interface. Semantically

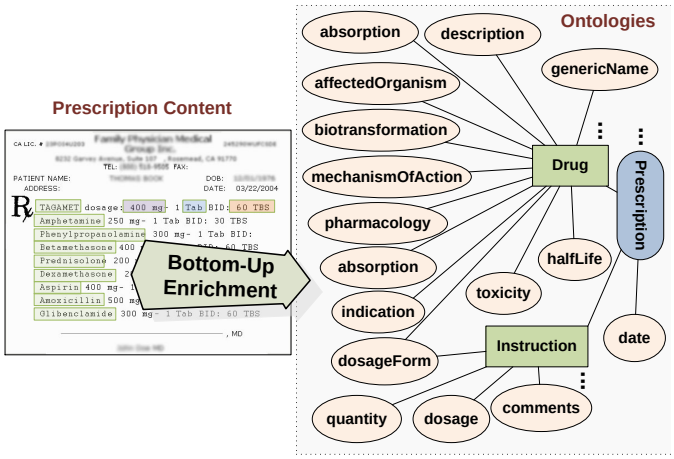


Fig. 3. Bottom-up semantic enrichment of prescriptions.

enriched prescriptions enable the traditionally written prescriptions to be utilized in novel ways as discussed above. In the following sections, we first describe the e-prescriptions and then discuss how they can be enriched as semantic documents.

V. SEMANTIC E-PRESCRIBING

E-health has evolved and emerged recently in many forms. E-prescription is one of those forms and defined as a computer-generated prescription utilized by health-care providers. E-prescribing as it is commonly called, is the use of an automated data entry system to generate a prescription that is then transmitted through a special network to a pharmacy in such a way that the data goes directly into the pharmacy's computer system. It plays an important role in improving the quality of patient care. For the prescriber, e-prescribing happens when a physician uses a computer or handheld device with a software that allows him or her to (with the patient's consent) electronically access information regarding a patient's drug benefit coverage and medication history; electronically transmit the prescription to the patient's pharmacy of choice; and, when the patient runs out of refills, his or her pharmacist can also electronically send a renewal request to the physician's office for approval.

One of the main challenges of the current e-prescription systems is the heterogeneity and evolving nature of available information sources. There exist already different sources of information addressing different aspects of pharmaceutical research. Recruiting available e-prescription systems in order to connect the existing dynamic pharmaceutical information is a challenging task. Linked Open Data when combined with existing e-prescribing systems, proposes a solution to tackle this challenge.

We define *Semantic Medical Prescriptions* as intelligent e-prescription documents enriched by dynamic drug-related meta-data thereby know about their content and the possible interactions. As depicted in Figure 3, semantic prescriptions are created based on a bottom-up process [28] in which normal e-prescriptions (unstructured or semi-structured with lower

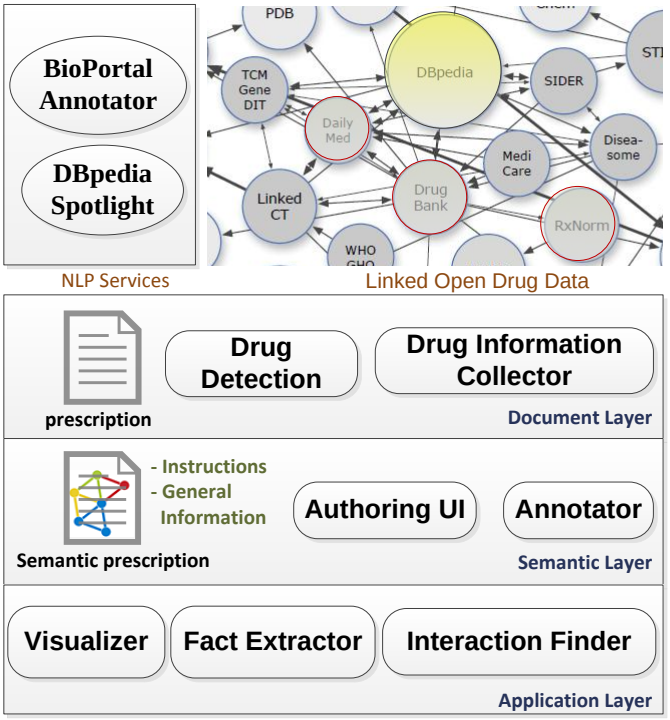


Fig. 4. Architecture of the Pharmer system.

level of expressiveness) are enriched with semantic metadata coming from a set of predefined ontologies (with upper level of expressiveness).

A. Architecture

The Pharmer system architecture is depicted in Figure 4 and consists of three layers:

a) Document Layer: This layer includes the traditional e-prescription document plus two components as *Drug Detection* and *Drug Information Collector*. Drug detection component performs the natural language processing (NLP) of the e-prescription document to detect the terms referring to a drug in the prescription. The component uses *DBpedia Spotlight* [29] and *BioPortal annotator* [30] NLP services to parse and analyze the text looking for known drugs. DBpedia Spotlight is a tool for automatically annotating mentions of DBpedia resources in text (i.e., Named Entity Recognition). BioPortal annotator is an ontology-based Web service that annotates public datasets with biomedical ontology concepts based on their textual metadata.

Automatic drug detection component is configurable so that users can easily add other existing NLP services for drug detection. When user is writing the prescription, this component asynchronously performs the drug recognition and adds the related annotations as real-time semantic tagging.

Another component in this layer is drug information collector which grabs all the information regarding a specific drug from Linked Open Data. To pursue this, it utilizes

datasets such as DrugBank, DailyMed and RxNorm (available at [12]) by sending federated SPARQL queries.

b) Semantic Layer: There are two main components in this layer namely *Annotator* and *Authoring UI*. The *annotator* component handles the automatic annotation and embeds the general information of the drugs as meta-data into the e-prescription. Annotator adopts the RDFa format. *RDFa* (Resource Description Framework in attributes) is a W3C Recommendation that adds a set of attribute level extensions to XHTML for embedding RDF metadata within web documents. RDFa fulfills the principles of interoperable metadata such as publisher independence, data reuse, self containment, schema modularity and evolvability.

The *authoring UI* component provides users with a set of input forms to manually embed the meta-data related to prescription instructions into the prescription document.

c) Application Layer: This layer provides a set of applications on top of the generated semantic prescriptions. *Interaction Finder* checks the possible interactions between the prescribed drugs by querying the relevant dataset in LODD and warns the prescriber about them. *Visualizer* is responsible for graphically representing the embedded semantics of a prescription (e.g., as depicted in Figure 7). The *Fact Extractor* generates the RDF/Turtle representation of the semantic prescriptions.

B. Features

The main features of Pharmer can be summarized as:

- *WYSIWYM User Interface*. Pharmer employs the WYSIWYM concept for integrated visualization, exploration and authoring of un-structured and semantic content. In Pharmer, users are able to directly manipulate the conventional e-prescriptions in order to enrich them with semantics. The generated annotations can be viewed by different sets of user interfaces with are configurable by users. For example, users can select specific border/background colors to distinguish the annotated drugs in a prescription.

- *Providing Different Semantic Views*. Semantic views allow the generation of different views on the same metadata schema and aggregations of the knowledge base based on the roles, personal preferences, and local policies of the intended users. Pharmer suggests two types of views: generic and domain specific views. Generic views provide visual representations of drug information (e.g., as information view depicted in Figure 6 or graph view in Figure 7). Domain specific views address the requirements of a particular domain user (e.g., a researcher need specific views for visualizing the atomic structure of chemical compounds).

- *Real-time Drug Tagging.* Real-time tagging means creating drug annotations while the user is typing. This will significantly increase the annotation speed [31]. Users are not distracted since they do not have to interrupt their current authoring task. Pharmer has a client-side component which interacts with the server asynchronously to make real-time tagging possible.

- *Drug Suggestion.* When searching for a drug, Pharmer suggests the similar drugs by taking into account the history of search terms and by sending SPARQL queries to the relevant datasets.

- *Automatic Drug Annotation.* Automatic annotation means the provision of facilities for automatic mark-up of prescriptions. The automatic process of annotating in Pharmer is composed basically of finding drug terms in prescription using an NLP service, mapping them against an ontology (i.e., DBpedia), and disambiguating common terms.

VI. USE CASES

A. Pharmer as a Ubiquitous Computing Platform for Semantic E-Prescribing

Mobile and ubiquitous computing devices are increasingly present and prevalent in the health contexts. This trend brings a number of possibilities of *mobile health* (m-health) to address critical aspects of health care and health system needs, by virtue of these devices' ubiquity, simplicity, and cost-efficiency [32]. In particular, in the process of semantic e-prescribing, having a mobile application will facilitate the creation of semantic medical prescriptions using any device and in any location.

Pharmer mobile application as shown in Figure 5 provides a mobile user interface for authoring of semantic prescriptions as well as accessing multi-dimensional data on medical prescriptions. Current ubiquitous devices are programmable and come with a growing set of facilities including multi-touch screens and cheap powerful embedded sensors, such as an accelerometer, digital compass, gyroscope, GPS, microphone, camera and other type of sensors. Utilizing these rich set of facilities in the context of medical prescriptions will enrich the patient medical prescription with sensor data thereby improves the quality of e-health services. For example, the location of user and some indicators like blood pressure or hear rate can be received from sensors by which Pharmer can specify the suitable drugs located in pharmacies close to the user.

B. Pharmer as a Professional Social Network for Health-care Service Providers

Pharmer offers couple of advantages over current e-prescribing systems. The main benefit of using semantic prescriptions is the persistent connection to up-to-date drug information coming from multiple dynamic data sources. So, when the information about a drug is updated occurs (e.g., change in its effects or interactions), the semantic prescription automatically adopts to this new change.

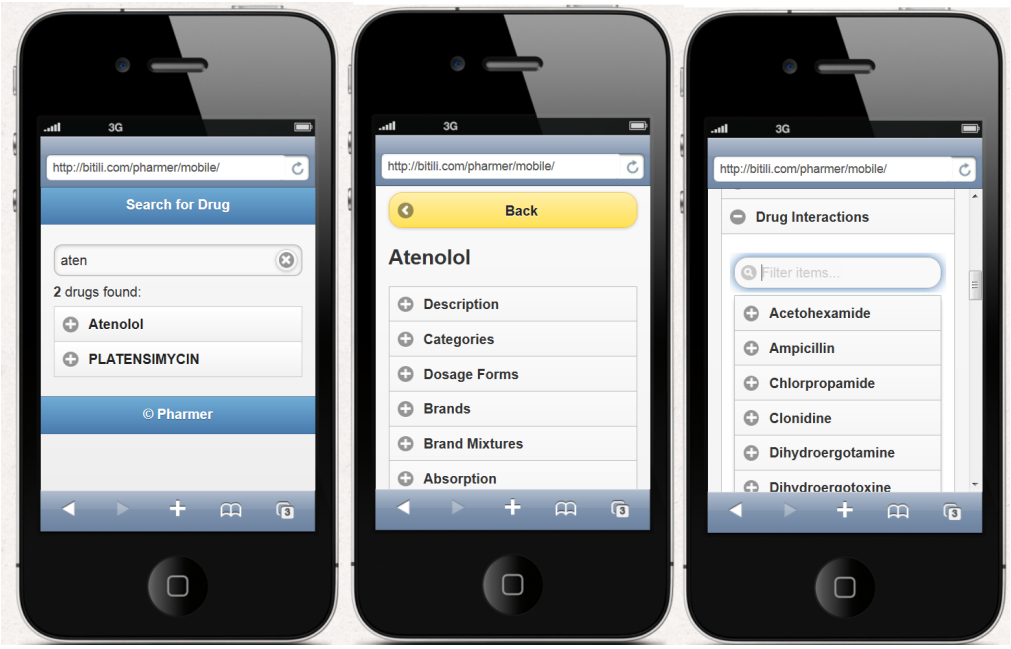


Fig. 5. Screenshot of Pharmer Mobile Application (available at http://bitili.com/pharmer/mobile).

Once writing a prescription it is very critical to consider drug interactions. Drug interactions are divided to three categories namely *food-drug*, *drug-drug* and *drug-plant* interactions. Coadministration can either be synergistic or antagonistic which respectively increase or decrease the drugs effect. The interactions may sometimes lead to change in the drug effect. By applying semantic prescriptions, all types of drug interactions are prevented and the probability of errors in prescriptions are reduced to a great extend.

A semantic prescription is a self-contained document, which is aware of its content and is connected to the linked open data. In contrast to database-oriented e-prescriptions, semantic prescriptions can easily be exchanged among other e-health systems without need to changing their related infrastructure hence enabling a connection between physicians, pharmacists, patients, pharmaceutical researchers, insurance and drug companies.

Pharmer as a prescribing tool is able to be incorporated in a health care social network. Such a network composed of health care professionals and patients who collaboratively write, correct and modify prescriptions in a semantically enriched environment. This social health care network provides patients and health care providers with services. It further facilitates relations between patients and health care professionals in order to improve shared decision making (SDM).

As information source, the network accesses LODD, where diagnostic and prescribing data has been located as well. Accessing such pieces of information, Pharmer is able to be used as a helper tool for facilitating the diagnosis and prescribing in assistance to physicians. Privacy of that network is also a critical point worth considerations.

1) Shared Decision Making: The traditional model of medical decision-making, in which doctors make decisions on treatment has no longer used in updated health care. The role of the patient, instead, in the consultation has been highlighted, mainly through introducing'patient-centred' strategies. Therefore, nowadays the models promoting patients active involvement in the decision-making procedure becoming developed.

A model introduced by Charles et al. [33] defines shared decision making only under the following four key characteristics. These keys are:

- both the patient and the doctor are involved
- both parties share information
- both parties take steps to build a consensus about the preferred treatment
- an agreement is reached on the treatment to implement

Pharmer as social network facilitates shared decision making through the connection amongst patient and physician on one hand and pharmacist on the other hand. According to Charles et al. model, Pharmer not only connects patients and physicians but also pharmacist as third party has an supervisory role on medication choice.

2) Fast Diagnostic Tool: Free access to LODD enables Pharmer to not only linked to e-prescribing systems but also to further assist physicians in diagnosis and treatment. Pharmer with direct connection to up-to-date information enables physicians to reconfirm their diagnosis and help them in finding proper treatment approaches. Physician, after general examination, enters the observed symptoms in Pharmer system and there, with the wealth of data available, Pharmer assists in diagnosis followed by therapies.

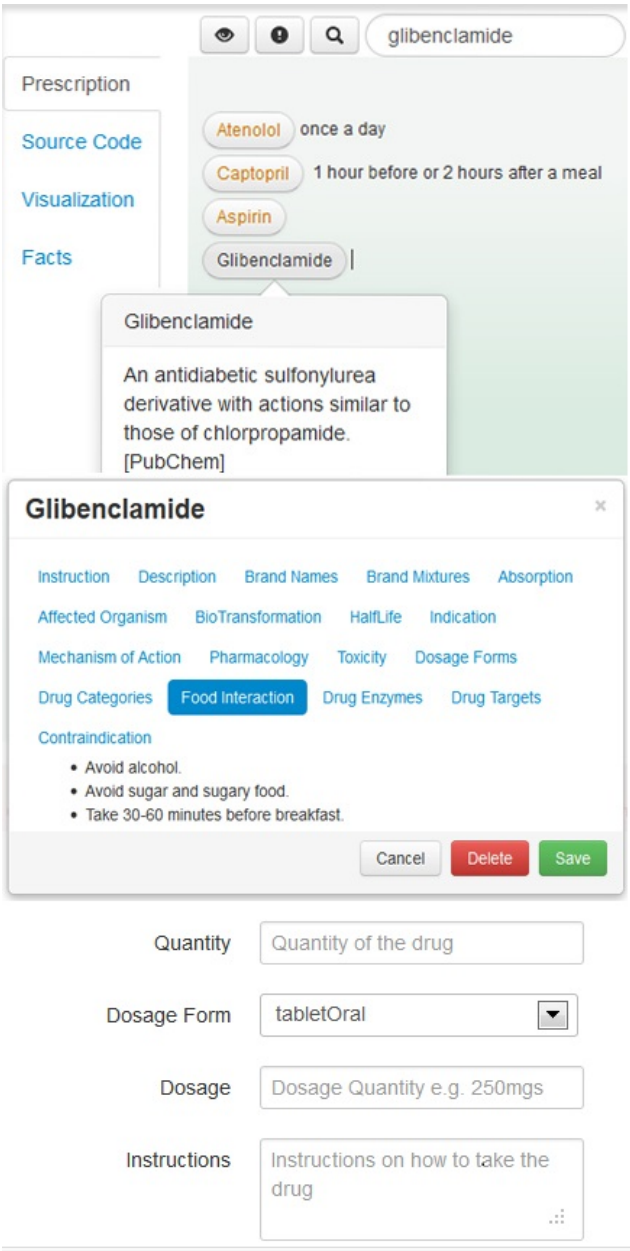


Fig. 6. Screenshot of the Pharmer application (general view, drug information view and prescription authoring view).

3) Privacy: Systems containing patient treatment history profiles, such as Pharmer, are required to be considered about privacy issues. This issue is solved by providing different users, with different views which is password protected. In such a protection, patient has only access to his profile while physician has access to all data of his patients, the same holds true for pharmacists and insurance companies. Other organisations(e.g., research institutes or pharma companies) can access to patients information as statistical data and only if the patient agrees. This protection helps Pharmer users to ensure data privacy.

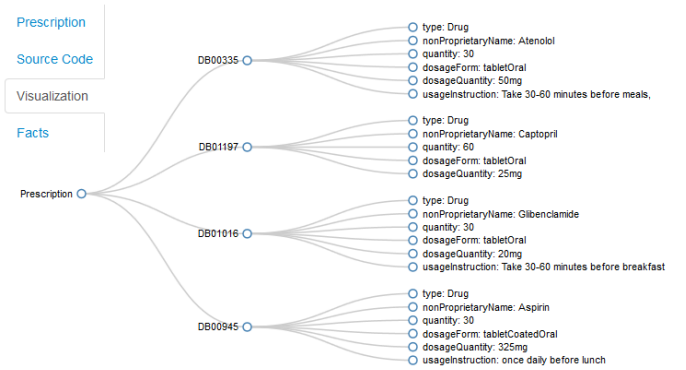


Fig. 7. Graph view in Pharmer.

C. Example Scenario

As depicted in Figure 8, Pharmer approach is very versatile and can be applied in a vast number of use cases by different stakeholders. The arrows in the figure can be summarized as the following:

1) The physician diagnoses the disease and writes the corresponding semantic prescription using the Pharmer, where patient's medication history is available.
2) The patient accesses to drug information, food interactions and adverse drug reactions via Pharmer.
3) The pharmacist verifies the prescription and considers alternative options suggested by Pharmer.
4) Pharma companies utilize the Pharmer data store in order to balance their production and distribution according to the market taste and demand
5) The Researchesrs easily access to the abundant data source and prescription statistical data.
6) Pharmer informs insurance companies to perform fair coverage plans according to covered drugs and patient's medication history.

All the above stakeholders utilize Linked Open Data as their integrated information source.

As a scenario, a 63 year old man with the history of MI (Myocardial Infarction) and type 2 diabetes visits a heart and coronary specialist complaining about frequent headaches and heavy head feeling. The specialist, after general inspection and monitoring vital signs, asks for a blood test. He then considers symptoms including high blood pressure (sys/dias:158/95 mmHg) and high Fasting Blood Sugar (150 mg/dl). He diagnoses high blood pressure and severe type 2 diabetes. Thereby, The patient profile is defined in Pharmer by patient's information besides diagnosis. "no weight loss" is mentioned as a preference in the patient's profile. Regardless of the patient's preferences, the physician would prescribe `Metformin` as a drug of choice. However, since the major side effect of `Metformin` is weight loss, the physician replaces `Metformin` with `Rosiglitazone`.

Considering the medication that the patient took before (`Glibenclamide` only), The specialist dispenses a new

Fig. 8. Pharmer ecosystem.

semantic prescription by entering the following drugs:

- **Rosiglitazone** 4 mg Oral Tablet once daily
- **Glibenclamide** 5 mg Oral Tablet bid
- **Atenolol** 50 mg Oral Tablet once daily

He then checks for the possible drug interactions by clicking the attributed button in the Pharmer software. As the Pharmer is connected to Linked Open Drug Data, it is capable of recognizing the most recent updated drug interactions (available at Drugbank dataset). He finds out that Sulfunyl Urea class drugs (here `Glibenclamide`) are not compatible to be coadministrated with beta-blockers (here `Atenolol`). So, he needs to replace it with another drug. Using the Pharmer and its connection to Linked Open Data, the physician can find the possible alternatives. Then he decides to choose `Captopril` as replacement.

The semantic prescription is then sent to the patient's pharmacy of choice. There, the pharmacist is able to review the semantic prescription and comments on that directly in the system so that the physician is also aware of the corresponding changes. The pharmacist comments may cause minor or major modifications in the semantic prescription. For instance, using the Pharmer she is able to check the appropriate dose of each medicine or suggest cheaper alternatives (if possible). In this case, as the `Rosiglitazone` elevates cardiovascular risks, the pharmacist suggests `Rosiglitazone` to be replaced by `Pioglitazone`. This change happens as a realization of the shared decision making between physision, pharmacist and patient. Thereafter, the patient who was referred to the pharmacy takes the prescribed drugs.

Before he starts taking the tablets, he enters in Pharmer system with his ID as patient. There, he is able to observe drug information embedded in the error free semantic prescription besides the preferred time and drug intake instructions. He is also informed about the possible food interactions. The patient's profile completes as he visits physicians or ask for refills. Furthermore, he is followed up by the physician and the pharmacist via the Pharmer. After 2 months the patient visits another specialist for his recurrent symptoms of diabetes. The specialist via the Pharmer accesses to the patient's medical profile and increases the anti-diabetic drug dose.

A researcher in an academy research institution investigates `Captopril` (as an Angiotansion II antagonist) effect on preventing diabetes recurrence. Having the data from the aforementioned patient follow up along with other similar patients allows investigator to lead her goal. In this case, for example, the `Captopril` along with anti-diabetic drugs led to diabetes recurrence. Observing all the corresponding patient profiles will either confirm or reject the research assumption.

A pharma company manager requires to determine the compliance rate of `Captopril` in the market in order to balance the production based on market demand. Applying the Pharmer allows him to simply access to these data and decide how to go on with this product. He is also able to collect the evidence which may prevent further dispense of `Captopril`

by physicians or consumption among patients. Pharmer allows insurance companies to customize and individualize their services based on patient's medical records. Recruiting Pharmer which contains information on insured drugs, the physician can choose the drugs accordingly. In the scenario, insurance company checks the dispensed medication with the disease and patient's insurance status therefore decides to refund the patient.

VII. Conclusion

Providing a consistent connection between patients, physicians, pharmacists, pharmaceutical researchers and pharma companies is a crucial step towards enhancing the quality of knowledge management and thereby e-health services in the pharmaceutical domain. With Pharmer, we presented in this article an approach for user-friendly implementation of *Semantic Prescriptions* as intelligent medical prescriptions to improve the integration and interoperability of e-prescribing systems with other e-health services.

We see the work presented in this article as an initial step in a larger research agenda aiming at promoting the authoring and annotation of semantically enriched medical documents. Regarding future work, we envision to extend the Pharmer application towards different modalities, such that the annotation of images and other medical objects is supported. Furthermore, we aim to integrate the other existing linked open datasets (e.g., related to publications, laboratories or insurance documents) into the Pharmer to extend its stakeholders.

Acknowledgments

We would like to thank our colleagues from the AKSW research group (http://aksw.org) for their helpful comments and inspiring discussions during the development of the WYSIWYM model for authoring of semantic medical prescriptions. Authors would also like to acknowledge the German Academic Exchange Service (DAAD) for the financial support of this work.

References

[1] A. Khalili and B. Sedaghati, "Pharmer – towards semantic medical prescriptions," in *eTELEMED 2013, The Fifth International Conference on eHealth, Telemedicine, and Social Medicine*. IARIA, 2013, pp. 9–14.

[2] M. Melissa Conrad Stoppler. Last visited date: 28/11/2012. [Online]. Available: http://www.medicinenet.com/script/main/art.asp?articlekey=55234

[3] A. R. González, Á. García-Crespo, R. C. Palacios, J. M. G. Berbís, and E. Jiménez-Domingo, "Using ontologies in drug prescription: The semmed approach," *IJKBO*, vol. 1, no. 4, pp. 1–15, 2011.

[4] Electronic prescribing (erx) incentive program. Last visited date: 28/11/2012. [Online]. Available: http://www.cms.gov/erxincentive

[5] W. Galanter, S. Falck, M. Burns, M. Laragh, and B. L. Lambert, "Indication-based prescribing prevents wrong-patient medication errors in computerized provider order entry (cpoe)," *Journal of the American Medical Informatics Association*, vol. 20, pp. 477–481, 2013.

[6] Medicare. Last visited date: 28/11/2012. [Online]. Available: http://www.medicareaustralia.gov.au/

[7] epsos : the european ehealth project. Last visited date: 28/11/2012. [Online]. Available: http://www.epsos.eu/

[8] J. Puustjärvi and L. Puustjärvi, "The challenges of electronic prescription systems based on semantic web technologies," in *ECEH*, 2006, pp. 251–261.

[9] C. Bizer, T. Heath, and T. Berners-Lee, "Linked Data - The Story So Far," *International Journal on Semantic Web and Information Systems (IJSWIS)*, vol. 5, no. 3, pp. 1–22, 2009. [Online]. Available: http://dx.doi.org/10.4018/jswis.2009081901

[10] Bio2rdf. Last visited date: 28/11/2012. [Online]. Available: http://bio2rdf.org/

[11] Semantic web in systems chemical biology. Last visited date: 28/11/2012. [Online]. Available: http://chem2bio2rdf.wikispaces.com/

[12] Linking open drug data (lodd). Last visited date: 28/11/2012. [Online]. Available: http://www.w3.org/wiki/HCLSIG/LODD

[13] M. Samwald, A. Jentzsch, C. Bouton, C. Kallesoe, E. Willighagen, J. Hajagos, M. Marshall, E. Prud'hommeaux, O. Hassanzadeh, E. Pichler, and S. Stephens, "Linked open drug data for pharmaceutical research and development," *Journal of Cheminformatics*, vol. 3, no. 1, 2011.

[14] M. Samwald, A. Jentzsch, C. Bouton, C. S. Kallesøe, E. Willighagen, J. Hajagos, M. S. Marshall, E. Prud'hommeaux, O. Hassenzadeh, E. Pichler, and S. Stephens, "Linked open drug data for pharmaceutical research and development," *Journal of Cheminformatics*, vol. 3, no. 19, 2011.

[15] A. J. Williams, L. Harland, P. Groth, S. Pettifer, C. Chichester, E. L. Willighagen, C. T. Evelo, N. Blomberg, G. Ecker, C. Goble, and B. Mons, "Open phacts: semantic interoperability for drug discovery," *Drug Discovery Today*, vol. 17, no. 21–22, pp. 1188 – 1198, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1359644612001936

[16] M. Saleem, S. S. Padmanabhuni, A.-C. Ngonga Ngomo, J. S. Almeida, S. Decker, and H. F. Deus, "Linked cancer genome atlas database," in *Proceedings of I-Semantics*, 2013.

[17] A. Khalili, S. Auer, and D. Hladky, "The rdfa content editor – from wysiwyg to wysiwym," in *COMPSAC 2012*, 2012, pp. 531–540.

[18] D. Tunkelang, *Faceted Search (Synthesis Lectures on Information Concepts, Retrieval, and Services)*. Morgan and Claypool Publishers, Jun. 2009.

[19] V. Lopez, V. Uren, M. Sabou, and E. Motta, "Is question answering fit for the semantic web? a survey," *Semantic Web ? Interoperability, Usability, Applicability*, vol. 2, no. 2, pp. 125–155, September 2011.

[20] G. Burel, A. E. Cano1, and V. Lanfranchi, "Ozone browser: Augmenting the web with semantic overlays," ser. CEUR WS Proceedings, vol. 449, June 2009.

[21] A. Ankolekar, M. Krötzsch, T. Tran, and D. Vrandecic, "The two cultures: mashing up web 2.0 and the semantic web," in *WWW 2007*, pp. 825–834.

[22] M. Sah, W. Hall, N. M. Gibbins, and D. C. D. Roure, "Semport - a personalized semantic portal," in *18th ACM Conf. on Hypertext and Hypermedia*, 2007, pp. 31–32.

[23] W. Muller, I. Rojas, A. Eberhart, P. Haase, and M. Schmidt, "A-r-e: The author-review-execute environment," *Procedia Computer Science*, vol. 4, pp. 627 – 636, 2011, iCCS 2011.

[24] K. Hasida, "Semantic authoring and semantic computing," in *Proceedings of the 2003 and 2004 international conference on New frontiers in artificial intelligence*, ser. JSAI'03/JSAI04. Berlin, Heidelberg: Springer-Verlag, 2007, pp. 137–149. [Online]. Available: http://dl.acm.org/citation.cfm?id=1899169.1899184

[25] J. Spiesser and L. Kitchen, "Optimization of html automatically generated by wysiwyg programs," in *WWW 2004*, pp. 355–364. [Online]. Available: http://doi.acm.org/10.1145/988672.988720

[26] C. Sauer, "What you see is wiki – questioning WYSIWYG in the Internet age," in *Proceedings of Wikimania 2006*, 2006. [Online]. Available: http://wikimania2006.wikimedia.org/wiki/Proceedings:CS1

[27] A. Khalili and S. Auer:, "Wysiwym authoring of structured content

based on schema.org," in *The 14th International Conference on Web Information System Engineering (WISE 2013)*, 2013.

[28] A. Khalili and S. Auer, "User interfaces for semantic content authoring: A systematic literature review," 2012. [Online]. Available: http://svn.aksw.org/papers/2011/journal_SemanticContentAuthoring/public.pdf

[29] Dbpedia spotlight. Last visited date: 28/11/2012. [Online]. Available: http://spotlight.dbpedia.org/

[30] Bioportal annotator. Last visited date: 28/11/2012. [Online]. Available: http://bioportal.bioontology.org/annotator

[31] "One Click Annotation," ser. CEUR Workshop Proceedings, G. T. W. G. A. Grimnes, Ed., vol. 699, February 2010. [Online]. Available: http://CEUR-WS.org/Vol-699/Paper4.pdf

[32] P. Petrucka, S. Bassendowski, H. Roberts, and T. James, "mhealth: A vital link for ubiquitous health," *Online Journal of Nursing Informatics (OJNI),*, vol. 17, p. 2675, 2013.

[33] C. Charles, A. Gafni, and T. Whelan, "Shared decision-making in the medical encounter: What does it mean? (or it takes at least two to tango)," *Social Science & Medicine*, vol. 44, no. 5, pp. 681 – 692, 1997. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0277953696002213

Multimodal Cognitive Nonverbal and Verbal Interactions: the Neurorehabilitation of Autistic Children Via Mobile Toy Robots

Irini Giannopulu
Cognitive Neuroscience
Pierre & Marie Curie University
email: igiannopulu@psycho-prat.fr

Abstract—**Multimodal cognitive nonverbal processes could be thought as a building block from which emotional and verbal expressions could emerge. With the intention to explore this hypothesis, we studied the interaction between autistic children and mobile toy robots during free spontaneous game play both quantitatively and qualitatively. Cognitive nonverbal criteria (eye contact, touch, manipulation, and posture) were analyzed, firstly in a dyadic interaction and secondly in a triadic interaction. The frequency of nouns and verbs including those which express positive emotion was figured out only in dyadic interaction. Once the cognitive nonverbal state between the child and the robot established, the child interacts with a third person displaying positive emotion. A positive correlation exists between multimodal cognitive nonverbal processes and verbal expression when the free game play with the robot is possible. This data suggests that in free spontaneous game play (i.e., ecological situation) the mobile toy robots could be used as a neural orthesis to enhance severe, middle and moderate autistic children's brain multimodal activity. The findings allow us to infer that this neural orthesis could pave the way for the development of synergistic dialogues between autistic children and human environment.**

Keywords-multimodal verbal and nonverbal interactions; autism; mobile toy robots; free game play; positive emotion; neural orthesis.

I. INTRODUCTION

Robots are utilized in training, and education of autistic children. The studies we develop aim to analyze the multimodal cognitive nonverbal and verbal interactions of autistic children, during free game play with mobile toy robots [1].

Autism spectrum disorder is a complex and heterogeneous neurological disorder that affects cognitive functioning but also emotional, social behavior and language development [2]. Language problems appear early and persist. Severe autistic children do not develop expressive language. However, when the children do acquire expressive language, it is often lacking any depth, it is echolalic and it is characterized by an absence of imagination [3]. Genetic studies have highlighted the complexity of the genetic architecture underlying autism. They consider autism as a complex multifactor disorder involving many genes [4], [5], [6]. These studies have given rise to new insights into neuronal circuits relevant to autism disorders. Post-modern analysis had demonstrated evidence of altered brain development, which strongly affects the formation of a multimodal neural network. The analysis shown that the neural substrate underlying cognitive, social, emotional and linguistic impairment involves multimodal areas such as the exterior superior temporal sulcus [6], the interior temporal lobe, amygdala included [7], as well as the ventral part of the prefrontal cortex, i.e., orbitofrontal cortex [8]. The autistic brain is also characterized by aberrant brain connectivity and disruption of white matter tracts between temporal regions [9], which disrupt verbal and nonverbal acquisition, consolidation as well as social interaction [10], [11], [12], [13]. These functional studies provide the basis for concluding that in autism the more impaired cortical areas are those that are involved in complex cognitive functions such as perception, social interaction, emotion and language. Such complex expression of autism necessitates a more generic consideration of this disorder at the multimodal neural level.

Developmentally speaking, the most widely accepted hypothesis in autism is the theory-of-mind deficit [14]. Even if this theory cannot account for the whole spectrum of autistic disorders, it raises many issues that not only involve mental representation of others but also social skills such as posture [15], eye contact [16], touching [9] and manipulation [17] that express social interaction [18].

Game play is a very important feature of early childhood and is of particular importance for children with autism. Play in children with autism is more like "learned routine" rather than "spontaneous" [19]. Autistic children show difficulty in their play activities, which could be associated with their deficit in cognitive, and emotional development, i.e., multimodal cognitive nonverbal and verbal interactions. Free game play characterized by spontaneity could allow children with autism the possibility to express themselves and engage in satisfying social activity, which in turn, could lead to development of their cognitive nonverbal and verbal skills [20], [21], [22].

Different approaches are being utilized to better understand the capacity of autistic children to interact with a robot [23]. The Aurora's project aim was to create a tool based on an autonomous robot (e.g., Labo-1, Kaspar, Robota doll) that convinces autistic children to engage in a process of interaction [24], [25], [26] [27]. A sensitive robot named Tito was employed in social interaction [28], [29]. Keepon, a very small fixed robot, can capture and maintain visual contact with the child [30]. Roboto uses the form of an animated face (mouth, eyebrows, eyes) to cause behavior imitation [31]. The dinausor Pleo seems

reinforce social behavior [32]. All these studies have shown that animated robots, humanoid or not, using different stimulation encourage interaction in autistic children. Even if quantitative metrics of social response for autism diagnosis including robots were developed [33]; only one study has used a quantitative technique for analyzing dyadic (child-robot) interaction for autism therapy [34]. With the exception of Labo-1 in the Aurora project, and Roball in Michaud's project so far, only fixed robots have been utilized reducing the child's spontaneity and self– expression in game play.

We used a mobile toy robot named "GIPY 1" (Figure 1), which incites the child to engage in interaction. On the hypothesis that autistic children will be in quasi-constant interaction with the robot, the cognitive behavior of severe autistic children in interactive activities with a robot, i.e., dyadic interaction, during spontaneous game play using multimodal cognitive nonverbal criteria was analyzed. In addition, we hypothesized that once dyadic interaction is established, the child could use the robot as a mediator to initiate the interaction with the third person, an adult, and express emotion, i.e., triadic interaction. This cognitive and emotional interaction of the autistic child with a third person was investigated, once again, in spontaneous, free game play by means of a multimodal approach.

Under the hypothesis that multimodal cognitive nonverbal interactions could be thought of as the building block from which expressive language could emerge, we used a new mobile toy robot named "POL" (Figure 4), which incites the middle and moderate autistic children to engage in dyadic (i.e., child-robot) interaction and express language. The relationship between multimodal cognitive nonverbal criteria (visual, tactile, manipulation and posture) with verbal behavior (including positive or negative emotion) was analyzed "with" and "without" free game play.

The present studies are part of our project actually in progress concerning multimodal interactions in typically and atypically developing children using natural and/or artificial environments.

Beginning with the design of the studies, we will continue with the analysis of the results of both multimodal cognitive nonverbal interactions in dyadic and triadic situation. Then, we will describe the correlation between multimodal cognitive nonverbal and verbal interaction in dyadic interaction before discussing the embodiment of multimodal information during free spontaneous game play between mobile toy robots and autistic children.

II. METHOD

A. Dyadic and triadic nonverbal interaction

1. Participants

• Dyadic nonverbal interaction

Four severe autistic children (3 boys and 1 girl) participated in this study. Their chronological ages ranged from 7 to 9 years old (mean 8.3 years). Their developmental age ranged from 2 to 4 years old. The children were diagnosed according to the D.S.M. IV-TR criteria of autism [35]. The C.A.R.S [36] had been administered at the age of 6 years by an experienced clinical psychologist. The C.D.I [37] was used to estimate intellectual disability (Table I).

TABLE I. GENERAL CHARACTERISTICS OF POPULATION A) CHILDHOOD AUTISM RATING SCALE B) INTERNATIONAL CLASSIFICATION OF DISEASES

Children	Developmental age	Sex	C.A.R.S (a)	C.D.I (b)
1	4	M	46.5	20 to 34
2	3.6	M	35.5	35 to 49
3	2.1	F	31.4	20 to 34
4	3.7	M	43.5	20 to 34

• Triadic nonverbal & emotional interaction: Case Study

"A" is a right-handed young boy. He exhibits mental retardation as per the C.D.I. [37]. His chronological age is 8 years old and his developmental age is 2 years old. The child was diagnosed with autism when he was 3 years old and still displays all characteristics of autism according to the D.S.M IV-TR [35]. In addition, the C.A.R.S. [36] has shown severe autism with a score of 43 points. "A" has deficits in reciprocal social interactions and communication (speech and language), stereotyped behavior and restricted interests and activities.

At the time of the experiment all of the children were attending special education classes of autism for both studies. The study was conducted in a day hospital outside of Paris. The experiment took place in a familiar room. The study was approved by the local ethics committee and was in accordance with the Helsinki convention. All the parents were formally informed and agreed to the participation of their children in this study. Anonymity was guaranteed.

2. Material

• Room

The room was 4.56 m by 3.34 m. A chair, a small wardrobe and a table on which the equipment needed for the framework of the study was placed (laptop and joystick), were used. In order to reduce the presence of disruptive elements and so as to avoid autistic bend, the room was left bare [38].

• Robot

A mobile robot, called "GIPY-1", which is cylindrical shaped with a diameter of 20 cm and a height of 30 cm, was created for use in the experiment. A representation of a neutral facial expression constitutes the cladding of the robot: the round eyes and nose triangle were dyed olive green and the elliptical mouth was dyed red (Figure 1). Everything was covered with a transparent plastic sheet. The simplicity of the robot was driven by the preference of autistic children for simple and predictable toy design [39]. An operator manipulated the robot via a wireless remote control using a joystick connected to a laptop. The robot could move forward, backward and turn on itself at low speed. These movements were constant.

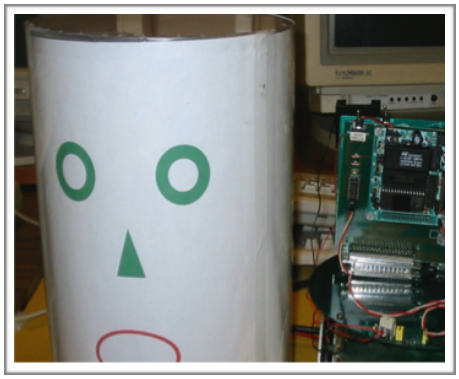


Figure 1. GIPY I

• *Protocol for the dyadic and the triadic interactions*

The duration of the session was 5 minutes. The robot was placed on the ground beforehand, in the center of the room, its stylized face toward the entrance. The game play session began as follows: when the child and the adult entered the room, the tele-operated robot carried out three movements (move forward, move back, 360° swivel). As in real social interaction, the child and the robot altered their responses. If the child approached, the robot moved back and conversely. If the child moved away from the robot, i.e., ignored the robot, the robot followed the child in order to attract its attention. If the child remained motionless, the robot approached or turned itself around in order to focus the attention of the child. All movements were standardized.

• *Analysis for the dyadic and the triadic interactions*

Two independent judges unfamiliar with the aim of the study completed the observations of the game play skills. Both performed the analyses of video sequences with Elan software [40]. Prior to assessing game play improvement, inter-judge reliability was assessed to ensure that both judges who analyzed videotapes were consistent in their analyses. Inter-judge reliability was assessed using intra-class coefficients to make the comparison between them. The inter-judge reliability was good (Cohen's kappa=0.63).

The dependent variable was the time of child-robot interaction for the dyadic interaction and the time of child-robot and adult for the triadic interaction. Accordingly, we calculated the duration of all the characteristics of each criterion. This was defined as the duration between the onset time and the offset time of each child's behavior toward the robot. Four criteria were defined for the dyadic interaction: 1) eye contact, 2) touching, 3) manipulation, 4) posture. Based on the hypothesis that cognitive interaction could be lead to the expression of an emotional state, an additional fifth criterion (5) was defined for the triadic interaction. This criterion was: positive emotion (Table II).

TABLE II. CHARACTERIZATION OF EACH CRITERION

Posture	***Touching***	***Eye contact***	***Manipulation***	***Positive Emotion***
S/he sits downs in front of robot; S/he bends towards the robot; S/he bends over the robot; S/he squats and bends over the robot; S/he steps over the robot	S/he puts the left hands on the robot; S/he puts the right hand on the robot; S/he touches the robot with both hands	S/he looks at the immobile robot; S/he watches the robot turning; S/he watches the robot going away; S/he watches the robot approaching	S/he seizes and blocks the robot with the two hands; S/he lifts of the robot; S/he stops the robot with both hands; S/he catches the robot; S/he returns the robot; S/he tilts the robot around itself and looks of its wheel; S/he puts back the robot upright	S/he smiles to the adult; S/he laughs to the adult; S/he expresses tenderness to the adult; S/he looks happy with the adult; S/he looks pleased to the adult

The duration of each criterion was calculated in seconds and was considered independent of the others. Concerning, for example, the characteristic "s/he looks at the immobile robot" ("eye contact") the onset time corresponded to the time when the child looked at the robot and the offset time to the moment when the child looked away from the robot. We calculated the duration of all the characteristics of each criterion. We summed up the duration corresponding to each criterion. Only the total duration is presented in the results section.

3. Results

• *Dyadic interaction*

The mean time of dyadic interaction was 238.7 sec. In other words, the children spent nearly 80% of their time (156 seconds for the first, 289 seconds for the second, 269 seconds for the third and 241 seconds for the forth child) playing with the robot. The duration of each robot-child interaction is presented in Figure 2. The duration of "eye contact" is similar for all the children. However, the analysis of the duration of "touching", "manipulating" and "posture" possibly reflects inter-individual differences related to different forms of autism. This analysis also showed how autistic children's behavioral interaction with the robot changes over a period of time. As such, this analysis suggests that in free game play a mobile toy robot could help autistic children to reduce repetitive and stereotypical behavior.

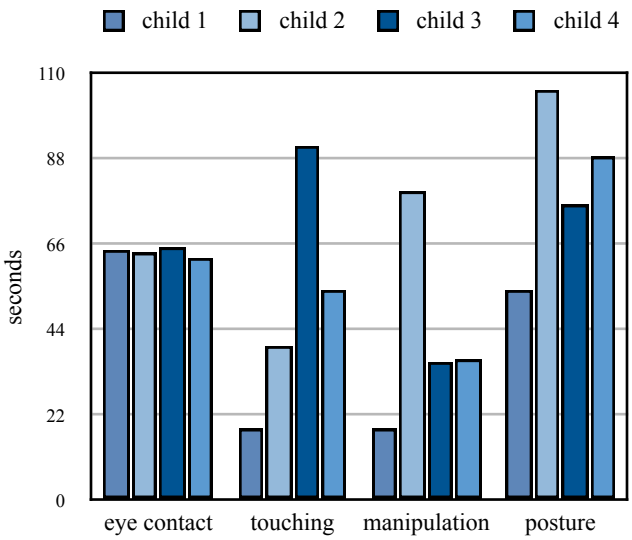


Figure 2. Duration of dyadic interaction for each criterion

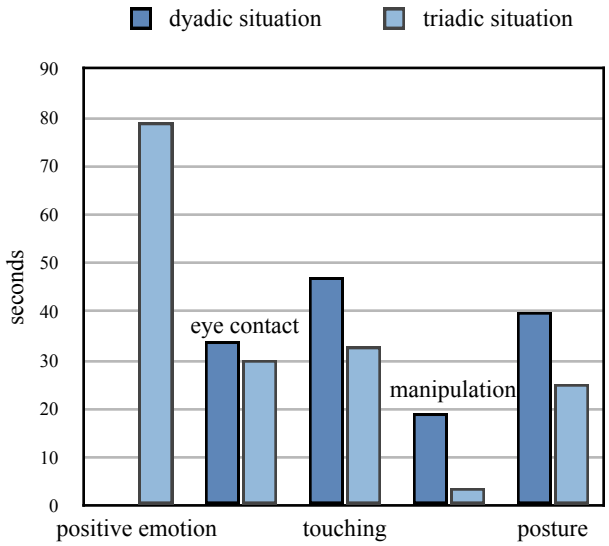


Figure 3. Duration of dyadic and triadic interactions for each criterion

• *Triadic nonverbal and emotional interaction*

The mean time of dyadic interaction was 25 sec; the mean time of triadic interaction was 30 sec. In other words, the child spends half the time playing with the robot and the half the time playing with the robot and the adult.

The duration of dyadic and triadic interaction is presented in Figure 3. The duration of "eye contact" and of "touching" is similar in both situations. However, the duration of "manipulation", of "posture" and of "positive emotion" differs between the two situations. As we can observe, positive emotion is more easily expressed when the child interacts with the adult and the robot than when the child interacts only with the robot. This difference reflective of the changes in autistic child behavior with the robot over a period of time also tells us that a mobile robot could be used as a mediator for social and emotional interaction. This is an encouraging conclusion with regard to the potential of human-to-human interaction.

B. Dyadic verbal and nonverbal interaction

1. Participants

Eleven children (8 boys and 3 girls) participated in this study. Their chronological ages ranged from 7 to 8 years old (mean 7.3 years; sd 6 months); their developmental age ranged from 5 to 6 years old (mean 6 years; sd 4 months). The mean age when first words appeared was 38 months (sd 5 months). The children were diagnosed according to the DSM IV-TR criteria of autism [20]. The Childhood Autism Rating Scale [21] had been administrated at the age of 6 years by an experienced psychiatrist. The present population is composed by middle and moderate autistic children. They were all verbal (Table III). The study was approved by the local ethics committee and was in accordance with the Helsinki convention. All the parents were formally informed and agreed to the participation of their children in this study. Anonymity was guaranteed.

TABLE III. GENERAL CHARACTERISTIC OF POPULATION

A. CHILDHOOD AUTISM RATING SCALE

Subjects	***Developmental age***	***Sex***	***C.A.R.S*** [a]
1	5.7	M	34
2	6.2	M	35
3	5.6	F	32
4	6.5	M	36
5	5.9	M	36
6	6.7	M	34
7	5.8	F	32
8	6.7	M	36
9	5.7	F	34
10	5.7	M	33
11	5.5	M	35

2. Material

• *Robot*

A mobile robot, called "POL", which is animal-shaped, was used: a mobile chicken (Figure 4). An operator manipulated the robot via a wireless control.

• *Protocol*

The study was conducted in two day hospitals: one outside and one inside of Paris. The experiment took place in a familiar room to all the children. We have defined two conditions: one *"*with*"* and another *"*without*"* game play.

"*Without game play":* Children's observation behavior with the immobile robot placed on the ground beforehand, in the center of the room. There is no game play session.

"With game play": Children's observation with the mobile robot. The robot was placed on the ground beforehand, in the center of the room. The game play session was unfolded as in the previous studies (see the protocol described previously).

The two conditions were counterbalanced across the children. The inter-condition interval was about 2 minutes. The duration of each condition was 10 minutes.

Figure 4. Child during game play with "POL" robot

• *Analysis*

In both conditions, two dependent variables (DV) were utilized: a) the duration of child-robot interaction; b) the frequency of nouns and verbs expressed by the children.

For the first DV, four criteria defined as in the previous analysis (dyadic and triadic nonverbal interaction). As in the above studies, we have measured the duration of all the characteristics of each criterion (Table II).

For the second DV, we have calculated the frequency of nouns and verbs, i.e., expressive language.

As in our previous studies, two independent judges unfamiliar with the aim of the study completed the observations of the whole protocol (*"with"* and *"without"* game play) performing the analyses of video sequences with Elan software [40]. The inter-judge reliability was good (Cohen's kappa =0.67).

3. Results

The distribution of duration according to the criteria in the two conditions approximates a non parametric shape. With such distribution, the median has been chosen as a central index for the comparisons. The statistical comparisons have been conducted with the Chi-Square Test ($\chi 2$ Test); relationship between cognitive nonverbal and verbal interactions was analyzed with the nonparametric Spearman rank correlation coefficient (Spearman's ρ correlation coefficient).

In *"without game play"* condition, the children interact less with the robot (1 minute and 57 sec) than in *"with game play"* (8 minutes and 40 seconds) ($\chi 2=6.89$, $p<0.01$). The results show that the median duration of "eye contact" is longer in the *"with game play"* condition (4.27 sec) than in the *"without game play"* (1.48 sec) ($\chi 2=7.12$, $p<0.01$). Similarly, the median duration of "touching", "manipulating" and "posture" is higher in *"with game play"* condition than in *"without game play"* condition, i.e., 2.36 sec vs. 1.24 sec; 1.16 sec vs. 0.66 sec; 1.51 sec vs. 0.97 sec respectively; ($\chi 2=6.07$, $p<0.025$; $\chi 2=4.7$, $p<0.05$; $\chi 2=4.01$, $p<0.05$ respectively) (Figure 5).

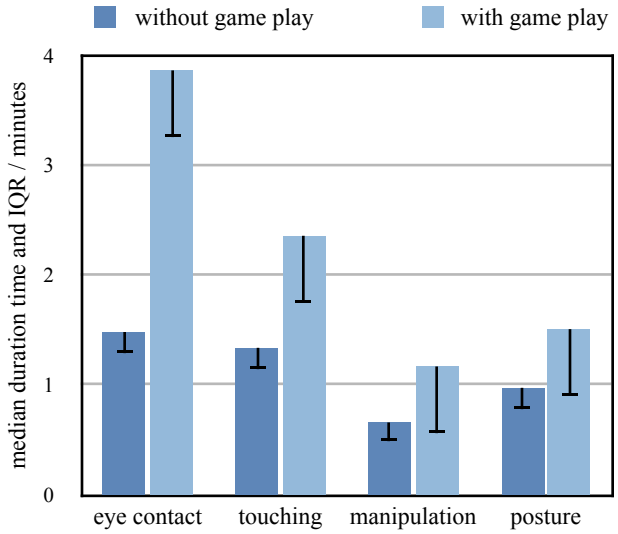


Figure 5. Duration of multimodal cognitive nonverbal interactions

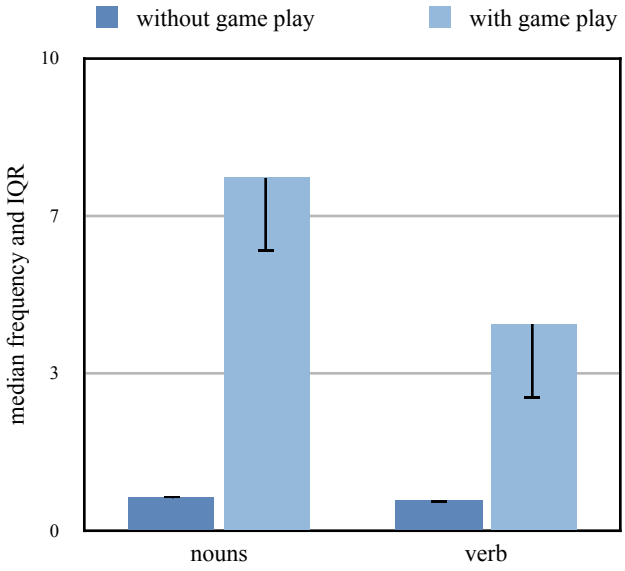


Figure 6. Median frequency of nouns and verbs

As the Figure 6 shows verbal expression was more frequent in *"with game play"* condition (7.45 median frequency for nouns; 4.36 for verbs) than in *"without game play"* condition (0.73 median frequency for nouns; 0.64 for verbs) ($\chi 2=7.16$, $p<0.01$ for the nouns; $\chi 2=6.99$, $p<0.01$ for the verbs) (Figure 3). Only in *"with game play"* condition, the children express three nouns (nice, beautiful, good) and one verb (like), which involve positive emotion ($\chi 2=3.99$, $p<0.05$ for the nouns; $\chi 2=3.88$, $p<0.05$ for the verbs).

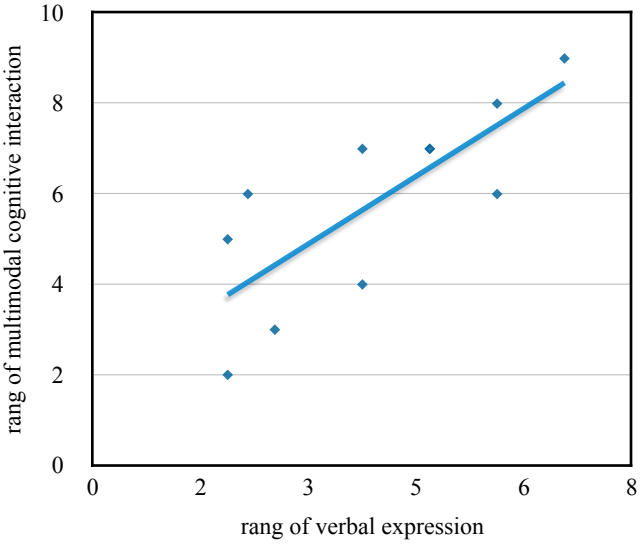


Figure 7. Relation between multimodal cognitive nonverbal and verbal (positive and neutral) information

A positive correlation exists between expressive language (nouns and verbs) and multimodal cognitive information in the *"with game play"* condition (Spearman's ρ correlation coefficient=0.747, p=0.01 one-tailed Test) (Figure 7).

In contrast, there is no positive correlation between expressive language (nouns and verbs) and multimodal cognitive information in *"without game play"* condition (Spearman's ρ correlation coefficient=0.23, p >0.05 one-tailed Test).

III. DISCUSSION

• *Dyadic nonverbal interaction*

Consistent with our hypothesis, the children were quasi-constantly in interaction with the mobile robot using a variety of ways. As autism is a spectrum disorder where a large variation in abilities and interests among autistic children is apparent, the interaction of children and robots was evaluated on the level of each individual child. Coherent with various studies, the present study shows that the use of robots engages autistic children in interaction [25], [29-32], [41-44]. In our case, we have computed the duration of robot-child cognitive nonverbal interaction during free, spontaneous game play. By doing so, the behavior of autistic children vis-a-vis the robot based on four nonverbal criteria has been analyzed and a temporal quantification of dyadic interaction with respect to the duration was performed. The analysis revealed that the duration of "eye contact" behavior was similar for each child. Inter-individual differences were identified for the duration of "touching", "manipulating" and "posture" behavior. These differences might be related to different expression of autism. The data demonstrated that the autistic children not only visually explored the robot [34] but also engaged in different kinds of play with the robot. In other words, the autistic children seem take an interest in playing with the mobile robot. It seems that free game play could be a relevant ecological situation, i.e., near to everyday life, where an autistic child spontaneously interacts with the robot. Moreover, mobile toy robot could help autistic children to reduce repetitive and stereotypical behavior. These findings also reveal that free, spontaneous game play with robots is possible with severe autistic children and could better facilitate the transfer of social and learnt abilities to everyday life.

Nevertheless, what is important to demonstrate is whether and how autistic children could generalize learnt abilities during play with the robot to adults, i.e., proving that the robot could be used as a neural mediator tool for the enrichment of child-human interaction. This later assumption has been analyzed using a triadic approach.

• *Triadic nonverbal and emotional interaction*

In this case study, we analyzed the ingredients of child-robot two-pronged interaction and child-robot-adult three-pronged interaction. Consistent with our hypothesis, the child first establishes a relationship with the robot and then uses the robot as an "instrument" to initiate the interaction with the adult (study 3). At first glance, our results are compatible with recent findings according to which the presence of a robot, are more effective than other environments in allowing autistic children to express social interest towards the robot [27-28], [30], [39], [45-46]. In these studies, researchers have used robots for treating autistic children. However, the relationship between robot and child has been studied solely based on the analysis of a single mode of interaction. Furthermore, the studies have been conducted using fixed robots. Our results go beyond these findings because we have demonstrated, as far we know for the first time, that in spontaneous, free game play, an autistic child uses the robot to interact with the adult and to express positive emotion. As such, on the one hand, we have shown that the dyadic interaction is based on a cognitive state and, on the other, that the child uses the robot as a mediator to express positive emotion playing with the adult.

More precisely, in this study, as in our previous studies [47-48], we have demonstrated that visual, haptic, tactile perception and posture, i.e., multimodal perception, are on the basis of the interest the child displays towards the robot. This is because, in our approach (as in Quinn & Eimas approach [53]), perception and cognition are considered to be a single domain rather than two distinct entities. The criteria we have chosen are assumed to represent the state of the child's cognitive nonverbal processes, as expressed by the interest the child exhibits towards the robot in spontaneous, free game play. As the present study has shown, once this state is established, the child develops a triadic relation, i.e., with the robot and the adult, thereby displaying enjoyment, which is a positive emotion. The expression of positive emotion could be related to the emergence of a cognitive state, which is multimodal in our case. This expression appears when the child interacts with the adult using the robot. In our study, the child "A" is in constant interaction with the robot, expressed by a multimodal cognitive state that, according to us, allows him to express positive emotion with the adult. When "A" interacts with both the robot and the adult, he changes his behavior. We think that the robot as a mediator could bring about neurocognitive improvements to the autistic child.

In both studies, the findings seem indicate that free game play, i.e., near to everyday life, encourages an autistic child to interact with the robot in a spontaneous manner and could reduce repetitive and stereotypical behavior. They also reveal that free, spontaneous game play with robots is possible with autistic children and could better facilitate the transfer of learnt abilities to everyday life.

One limitation of these studies is the small number of autistic children, which makes impossible inferential analysis. Additional studies are required with a substantial number of children. We also need to confirm the importance of free game play in improving children's nonverbal but also verbal performances. In that context, we have explored the multimodal cognitive nonverbal and verbal interactions between a mobile toy robot and autistic children with and without free game play.

- *Dyadic verbal and nonverbal interaction*

In that study, our results indicated that the duration of multimodal cognitive nonverbal interactions (visual contact, manipulation, touching, posture) is longer when free spontaneous game play with the robot is possible (with game play condition) than when game play is impossible (without game play condition). Consistent with the studies presented here above, these new results show, once again, that a mobile toy robot engages autistic children in multimodal nonverbal interactions (visual, tactile, manipulation, posture). Taken together, these studies seem demonstrate that autistic children's behavioral interaction with a mobile robot changes over a period of time. Free game play, which is very close to a everyday life situation, encourages autistic children to interact with the robot in a spontaneous manner [22].

Coherent with the above is the fact that language is expressed only during free game play. Even if the children of our study suffer from middle or moderate autism and are verbal, these results show that the expression of language is possible when the children interact with the robot (in free game play) using a multimodal mode. In the same vein, children produced three nouns and one verb, which connote positive emotion only during game play. Moreover, positive correlation between multimodal cognitive nonverbal information and verbal expression is significant when the children spontaneous interact with the mobile robot: the more the multimodal nonverbal interaction, the more the verbal expression. As such, these results are consistent with our hypothesis suggesting that multimodal cognitive nonverbal interactions could be considered as the basis of verbal expression.

- *Dyadic, triadic verbal, nonverbal and emotional interaction*

The present three studies seem indicate that mobile robots could not only be used as a mediator for nonverbal [48] and emotional interaction ([1], [22], [50], [51]) but also for verbal expression ([52]), which is the distinguishing characteristic of the inter-human communication. This is a comforting issue with regard to the potential of human-robot interaction. As such, the data suggests, that more mobile that immobile robots could be efficient for training, education and neurorehabilitation of autistic children. In other words, an artificial environment such as mobile toy robots could provide the source of emergence of multimodal cognitive nonverbal information, which in turn, could be combined with emotional [49], [50] and verbal information [52] in a coordinated manner. The mobile robots (which can be considered as a neural orthesis), could pave the root for the development of synergistic dialogues between autistic children and human environment. As our data has shown free game play (which is close to everyday life) is more auspicious than the absence of free game play to improve severe, middle and moderate autistic children's brain multimodal activity.

The reporting data converges to say that autistic behavior can be improved via mobile robots, i.e., artificial environments. This is coherent with the assumption that cognitive nonverbal/verbal and emotional development is the result of a complex process with three foci at least, one in the central nervous system, one in the mind and one in the child's dynamic interactions with the environment [22]. The human brain undoubtedly has its own dynamic that allows neurons to interact, which in turn, affects the development and function of the brain areas [21]. In the case of autism, the brain activity is characterized by an hypofunctioning. An artificial environment like a mobile robot, i.e., neural orthesis seems improve the neural activity (and consequently) the behavior of autistic children: autistic children interact with the robot multimodally only in free game play (Figure 8).

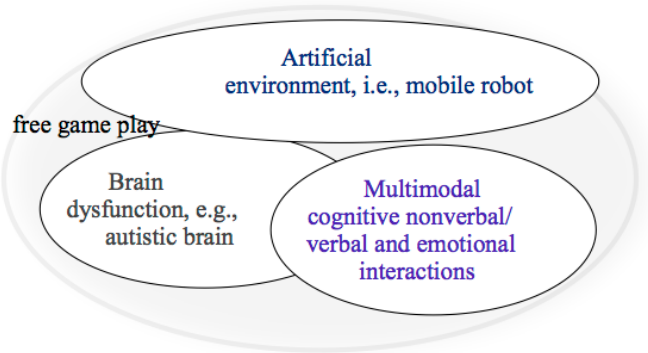


Figure 8. Principle of Neurorehabilitation during free game play

Our hypothesis is that this emerging brain multimodality is crucially shaped by the children's interactions with the environment during free game play. Nonverbal cognition, language and emotion develop at the interface between neural processes. They arise from the dynamic interaction between the developing brain and the artificial environment, i.e., the robot [22].

Our approach, actually in progress, attempts to understand "how" artificial environments could be considered as the root of neuronal organization and reorganization ([21], [22]). Based on the brain's intrinsic properties, neuroplasticity and the fact that the brain is neurodynamic, our studies try to demonstrate that a mobile robot could be used as a neural orthesis with the

intention to support the embodiment of cognitive nonverbal, emotional and verbal information processing in free game play.

To our knowledge, this data represents some of the first to analyze the relationship between multimodal cognitive nonverbal information and verbal expression during spontaneous free game play in severe, middle and moderate autistic children using mobile toy robots.

IV. CONCLUSION

With these studies (part of our project actually in progress), we have demonstrated evidence for the view that spontaneous free game play with toy robots could be a source from which the brain of autistic children can take advantage. This condition could become the beginning of most of one's knowledge base for autistic children: visual, haptic, tactile perception, body posture as well as verbal expression. It is of great interest, particularly when considering that nonverbal information is probably at the origin of what is arguably one of the trademarks of human cognition: the capacity to generate thoughts and concepts for ourselves and for the others, which can be verbally expressed. To better understand the base and the nature of the verbal expression we observed, future studies should extend this work through systematic analyses within a larger sample of autistic children in a follow up design.

ACKNOWLEDGMENT

The author would like to thank all the children who participated to the studies, their parents, ANR's supporting, as well as the French Department of Education.

REFERENCES

[1] I. Giannopulu, "Multimodal Human-Robot Interactions: the Neurorehabilitation of Severe Autistic Children," Proceedings of the Sixth International Conference on Advances in Computer-Human Interactions (ACHI), IARIA, 2013, pp. 68-73.

[2] L. Crane, S.E. Lind, and D.M. Bowler, "Remembering the past and imaging the future in autism spectrum disorder," *Memory*, vol. 21, no. 2, pp. 157-166, 2013.

[3] J.B. Plavnick and S.J. Ferreri, "Establish verbal repertoires in children with autism using function-based video medeling", *J Appl Beh Analysis*, vol. 44, pp. 747-766, 2011.

[4] L. Jorde et al. "Complex segregation analysis of autism," *Am. J. Hum. Genet.,* vol. 49, pp. 932-938, 1991.

[5] N.H. Sykes and J.A. Lamb, "Autism: The quest for the genes," *Exp. Mol. Med.,* vol. 9, pp. 1-15, 2007.

[6] P. Szatmari et al. "Mapping autism risk loci using genetic linkage and chromosomal rearrangements," *Nat. Genet.,* vol. 39, pp. 319-328, 2007.

[7] K.A. Pelphrey and E.J. Caster, "Charting the typical and atypical development of the social brain," *Dev. Psychopathol*., vol. 20, pp. 1081-1102, 2008.

[8] B.A. Corbett et al. "A functional and structural study of emotion and face processing in children with autism," *Psych. Res.,* vol. 30, 173, pp. 196-205, 2009.

[9] L. Brothers, "The social brain: a project for integrating primate behaviour and neurophysiology in a new domain," Concepts. *Neurosc.,* vol. 1, pp. 27–51, 1990.

[10] U. Frith and C.D. Frith, "Development and neurophysiology of mentalizing," *Philos. Trans. R. Soc. London,* vol. 358, pp. 459-473, 2003.

[11] A. Saitovitch et al. "Social cognition and the superior temporal sulcus: Implications in autism," Rev. Neurol. (Paris), 2012 Sep 13. pii: S0035-3787(12)00902-2. doi: 10.1016/j.neurol.2012.07.017.

[12] R. Adolphs and D. Tranel, «Intact recognition of emotional prosody following amygdala damage," *Neuropsychol.,* vol. 37, pp. 1285–1292, 1999.

[13] J.P. Aggleton, The Amygdala: A Functional Analysis. Oxford: University Press, 2000.

[14] B.M. Nacewicz et al. "Amygdala Volume and Nonverbal Social Impairment in Adolescent and Adult Males with Autism," *Arch. Gen. Psychiatry.,* vol. 63, pp. 1417-1428, 2006.

[15] U. Frith, J. Morton, and A.M. Leslie, "The cognitive basis of a biological disorder: autism," TIN14, pp. 433-438, 1991.

[16] A.M. Leslie, "Pretending and believing: issues in the theory of ToMM," *Cognition,* vol. 50, pp. 211-238, 1994.

[17] T. Reed and C. Peterson, " A comparative study of autistic subjects' performance at two levels of visual and cognition perspecctive taking," *J. Autism. Dev. Disord.,* vol. 20, pp. 555-567, 1990.

[18] A. Escalona, T. Field, J. Nadel, and B. Lundy, "Brief report: Imitation effects on children with autism," *J. Autism. Dev. Disord*., vol. 32, pp. 141-144, 2002.

[19] B.A. Taylor et al. "Manipulating establishing operations to promote initiations toward peers in children with autism," *Res. Dev. Disabil.,* vol. 26, pp. 385-392, 2005.

[20] E. Williams, V. Reddy, and A. Costal, "Taking a closer look at functional play in children with autism," *J. Autism. Dev. Disord.,* vol. 31, pp. 67-77, 2001.

[21] I. Giannopulu, "Contribution à la compréhension des représentations multimodales chez l'homme sein et chez des patients avec atteinte neuropsychologique: une approche life span". H.D.R, UPMC-Paris VI, 2011.

[22] I. Giannopulu, "Multimodal interactions in typically and atypically developing children: natural vs. artificial environments", *Cogn Proc.*, vol. 14, pp. 323–331, 2013.

[23] B. Scassellati, H. Admoni, and M. Mataric, "Robots for use in autism research," *An. Rev. Biomed. Eng.,* vol. 14, pp. 275-294, 2012.

[24] M. Davis, B. Robins, K. Dautenhahn, C.L. Nehaniv, and S. Powell editors. "A comparison of interactive and robotic systems in therapy and education for children with autism," European Conference for the Advancement of Assistive Technology in Europe (AAATE'05), pp. 353-357, September 6-9 2005, IOS Press Lille.

[25] F. Michaud and S. Caron, "Roball, the rolling robot," *Auto. Rob*., vol. 12, pp. 211-222, 2002.

[26] K. Dautenhahn, "Socially intelligent robots: dimensions of human-robot interaction," *Philos. Trans. R. Soc.,* vol. B.362, pp. 679- 704, 2007.

[27] B. Robins, K. Dautenhahn, and K. Dubowski, "Robots as Isoalators or Mediators for children with autism ? A Cautionary Tale," Proc's of the AISB 05 Symposium on Robot Companions, pp. 82-88, 2005.

[28] A. Duquette, F. Michaud, and H. Mercier, "Exploring the use of a mobile robot as an imitation agent with children with low- functioning autism," *Auto. Rob. – Special Issue.,* vol. 2, pp. 147-157, 2004.

[29] F. Michaud, T. Salter, A. Duquette, and J.F. Laplante, "Perspectives on mobile robots used as tools for pediatric rehabilitation," *Assist. Technol*., vol. 19, pp. 14-19, 2007.

[30] H, Kozima, C Nakagawa, and Y. Yasuda, "Children-robot interaction: a pilot study in autism therapy," *Prog. Brain. Res.,* vol. 164, pp. 385-400, 2007.

[31] J. Nadel, A. Revel, P. Andry, and P. Gaussier, "Toward communication: first imitations in infants, low-functioning children with autism and robots," *Inter St: Soc. Beh. Commun. Biol. Syst*., vol. 5, pp. 45-54, 2004.

[32] E.S. Kim, L.D. Berkovits, E.P. Bernier, D. Leyzberg, F. Shic, R. Paul, B. Scassellati, "Social Robots as Embedded Reinforcers of Socail Behavior in Children with autism," *J Autim Dev Disord,* DOI 10.1007/ s10803-012-1645-2, 2012.

[33] B. Scassellati, "Quantitative metrics of social response for autism Diagnosis," IEEE International Conference on Intelligent Robots and Systems, pp. 1134-1138 vol.2, 2002.

[34] K. Dautenhahn, and I. Werry, "A quantitative technique for analysing robot-human interactions," IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 1132-1138, 2002.

[35] DSM-IV-TR Manuel diagnostique et statistique des troubles mentaux, Paris, Editions Masson, 2003.

[36] E. Schopler, R.J. Reichler, R.F. De Vellis, and K. Daly, "Toward objective classification of childhood autism: Childhood Autism Rating Scale (CARS)," JADD10, pp. 91-103, 1980.

[37] International Classification of Diseases, Revision, 10, 1990.

[38] P. Planche E. Lemonnier, K. Moalic, C. Labous, and A. Lazartigues, "Les modalités de traitement de l'information chez les enfants autistes," *Ann. Med. Psychol.,* vol. 160, pp. 559-564, 2002.

[39] J. Pedersen and J. Schelde, "Behavioral aspects of infantile autism: an ethological description", *Eur. Child. Adolesc. Psy.,* vol. 6, pp. 96-100, 1997.

[40] M. Tacchetti, Elan Linguistic Annotator. The Language Archive MPI for psycholinguistics Nijmegen, The Netherlands. 2006.

[41] K. Dautenhahn and I. Werry, "Towards interactive robots in autism therapy," *P & CT*, 121, pp. 1–35, 2004.

[42] H. Kozima and H. Yano, Designing a robot for contingency- detection game. Working Notes Workshop Robotic & Virtual Interactive Systems in Autism Therapy, 2001 27-28 September, U.K.Hatfield. University of Hertfordshire, 2001, Technical Report No 364.

[43] J. Martineau, S. Cochin, R. Magne, and C. Barthelemy, "Impaired cortical activation in autistic children: Is the mirror neuron system involved?" *Int. J. Psychophysiol.,* vol. 68, pp. 35–40, 2008.

[44] T. Salter, I.P. Werry, and F. Michaud, "Going into the wild in child-robot interaction studies-Issues in social robotic development," *Int. J. Robot.,* vol. 1, pp. 93-98, 2007.

[45] Y. Yokoyama, "The Possibility of the Psychiatric Treatment with a Robot as an Intervention From the Viewpoint of Animal Therapy," Proc. of Joint 1st International Conference on Soft Computing and Intelligent Systems and 3rd International Symposium on Advanced Intelligent Systems, paper number 23Q1-1, in CD-ROM Proc. 2002.

[46] C. Kozima and Y. Yasuda, "Children-robot interaction: a pilot study in autism therapy," *Prog. Brain. Res.,* vol. 164, pp. 385-400, 2007.

[47] I. Giannopulu and G. Pradel, Mobile toy robots can be used in autism therapy: an example of application. IEEE Proceedings in the IROS 2009 paper number SuT8.pdf in the workshop CD Proc. 2009a.

[48] I. Giannopulu and G. Pradel, "Interactions multimodales en situation de jeu libre entre enfants autistes et un robot mobile". *7ème Journée Nationale de la recherche en Robotique, http:// jnrr09.lms.sp2mi.univ poitiers.fr/IMG/pdf/ Giannopulu_Pradel.* Novembre, Domaine de La Grande Garenne 18330 Neuvy-sur- Barangeon, 2009b.

[49] I. Giannopulu and G. Pradel, "Multimodal interactions in free game play of children with autism and a mobile robot," *NeuroRehabilitation*., vol. 27, pp. 305-311, 2010.

[50] I. Giannopulu, "Cognitive and emotional interactions between autistic child, mobile toy robot and therapist," *Front Comput Neuro,* doi:10.3389/ conf.fncom.2011.52.00002, 2011.

[51] I. Giannopulu and G. Pradel, "From child-robot interaction to child-robot-therapist interaction: a case study in autism," *Appl Bio Biomech., vol* 9, pp. 173-179, 2012.

[52] I. Giannopulu "Embedded Multimodal Nonverbal and Verbal Interactions Between a Mobile Toy Robot and Autistic Children," Late Breaking Paper, HRI 2013, Proceeding of the 8th ACM/IEEE International Conference on Human-robot Interaction, pp. 127-128, IEEE Press Piscataway, NJ. USA.

[53] P.C. Quinn and P.D. Eimas, "The emergence of category representations during infancy: are separate perceptulal and conceptual processes required?" *J. Cogn. Dev.,* vol. 1, pp. 55-61, 2000.

EMuRgency – A Basic Concept for an AI Driven Volunteer Notification System for Integrating Laypersons into Emergency Medical Services

Jesko Elsner, Philipp Meisen, Sebastian Thelen, Daniel Schilberg, and Sabina Jeschke
Institute of Information Management in Mechanical Engineering
IMA/ZLW & IfU - RWTH Aachen University
Dennewartstr. 27, 52068 Aachen, Germany
{Jesko.Elsner, Philipp.Meisen, Sebastian.Thelen, Daniel.Schilberg, Sabina.Jeschke}
@ima-zlw-ifu.rwth-aachen.de

***Abstract*—In case of Sudden Cardiac Arrest, an untreated time interval of only a few minutes usually means the victims' death. Due to a variety of parameters, e.g., the current traffic situation and the mere traveling distance, emergency medical services often arrive too late in order to procure efficient cardiopulmonary. Given this premise, it is necessary to find alternative ways for providing immediate first aid measures. One promising approach is the implementation of a Volunteer Notification System (VNS) – integrating laypersons and medically trained volunteers into the professional medical services by notifying those potential helpers who are, at the time of incident, close to the victim. By tracking the users' location, the system is able to notify those volunteers who can arrive on scene fast enough to provide the urgently needed measures. In September 2011, the European research project "EMuRgency" started the development of such a system. Whereas a running prototype has already successfully been developed, a more sophisticated solution is required to determine the relevant volunteers. While the actual distance is an important parameter, it does not necessarily determine the time of arrival at the scene. Possible obstacles might be in the way and both physical performance and the type of movement have a direct influence on the traveling speed. Furthermore, secondary criteria apply; e.g., the current situation a volunteer is in, knowledge of the area, general engagement and his or her medical expertise. This publication will give you an overview of the "EMuRgency" project and discuss the current stage of development of the VNS; furthermore, this paper will introduce the main concept for determining the relevant volunteers within an ongoing emergency scenario and the advantages of an artificial intelligence approach to enable an efficient volunteer selection will be discussed.**

Keywords*—*Volunteer Notification System, First Responder, Emergency Medical Services, Sudden Cardiac Arrest, Cardiopulmonary Resuscitation, Telemedicine, eHealth, mHealth

I. INTRODUCTION

Due to the way today's professional emergency medical services (EMS) are organized, victims in need of urgent medical care are facing a lethal problem. Depending on the type of emergency, the time interval between the incoming emergency call and the arrival of the professional helpers at the scene is simply too long [1]. In Bavaria (Germany) for example, a region with good infrastructure and an advanced medical system, reoccurring studies are made every four years, analyzing the effective time interval local EMS need until arriving at the place of incident. The Institute For Emergency Medicine in Munich (INM) states in a recent study that professional EMS in the area of Bavaria require approximately nine minutes until arriving on scene [2]. The severity of this time deficit generally correlates with the infrastructure a country can provide, resulting in intensification for less advanced countries and less populated regions. While most emergencies do not involve an immediate life danger for the victim, in case of a Sudden Cardiac Arrest (SCA) the first minutes are of utter importance. Jan Bahr states that as little as three minutes is most likely enough for victims of SCA to suffer permanent brain damage and Karin Grassl states that survival without serious permanent damages are practically zero if a victim was left untreated for more than five minutes [3] [4]. Victims suffering SCA are in need of urgent medical care that professional EMS alone cannot always sufficiently provide. The implementation of a Volunteer Notification System (VNS) is a possible solution [1]; integrating laypersons and medically trained volunteers into the EMS by notifying those potential helpers who are, at the time of incident, close to the victim in order to provide early first aid measures and, therefore, gap the time until the professional helpers arrive on scene.

A. Structure

The first section of this paper describes the medical emergency of a Sudden Cardiac Arrest (SCA), introduces the basic concept of a Volunteer Notification System (VNS) and gives a brief introduction into the main problems regarding the determination of the relevant volunteers in an ongoing medical emergency [5]. Starting by identifying and discussing comparable systems, the second section analyzes the technological state-of-the-art of mobile phones and messaging technologies; herby determining the possibilities and restrictions for a VNS approach. Furthermore, the basic concepts of artificial intelligence (AI) that are required to enable an efficient selection of the relevant volunteers will be introduced. The third section introduces the research project "EMuRgency". One focus of the project is the actual implementation of a VNS, the current state of development is being discussed; describing the main system components

and corresponding architectural details, followed by an introduction to the concept of "Prescient Profiling". The last section provides an outlook on the upcoming research within the project and shortly discusses the relevance on related topics of general location-based services.

B. Sudden Cardiac Arrest

The human heart has an electrical conduction system that controls the rate and rhythm of the heartbeat. Problems with this electrical system can cause irregular heartbeats, so called arrhythmias, which can lead to Sudden Cardiac Arrest (SCA) – a condition in which the heart suddenly and unexpectedly stops beating [6]. The hereby resulting loss of blood flow prevents the brain and any other vital organ from getting oxygen. Without immediate treatment, the victim dies within minutes [3]. It is a common misconception that SCA is the same as a heart attack, while in reality, they are quite different. SCA is an "electrical problem" that prevents the heart in its whole from functioning, whereas a heart attack occurs when part of the heart's blood supply is reduced or blocked causing the heart muscle to become injured or die [6].

C. The basic concept of a Volunteer Notification System

One possible solution for offering faster response treatment is the concept of involving volunteers into EMS by implementing a VNS, which may be defined as an IT system with the following core functionality: by tracking the location of all registered users, the system will be able to notify exactly those potential helpers who are, at the time of the incoming emergency call, close enough to the place of incident in order to provide Cardiopulmonary Resuscitation (CPR) to the victim and, therefore, gap the time until the professional EMS arrive on scene [1]. Whereas the term "close" is suitable for describing the general concept of a VNS, the actual determination of which volunteers are to be notified in an ongoing emergency situation cannot be answered without understanding further implications [5]; this topic will be addressed in part D and E of this section and a possible approach regarding an efficient volunteer selection will be discussed in Section III by introducing the concept of "Prescient Profiling".

The general concept of a VNS does not interfere with the local corresponding emergency standard procedures, but can rather be described as an optional add-on to existing EMS; the responsible dispatcher decides if to involve this optional feature. The potential volunteers are not a replacement for emergency physicians or any professional helper that is normally involved in a medical emergency workflow; their main purpose is to arrive at the victim fast enough to provide CPR and, therefore, bridge the untreated time interval until the professional helpers arrive on scene [1]. While no exact definition or specification of a VNS exists so far, it is part of this publication to discuss its architecture and introduce its main components. One possibly technical implementation of a transnational VNS, utilizing standard smartphones as client devices for receiving notifications, is the focus of the European research project "EMuRgency", which current state of development will be discussed in Section III.

D. Who are the relevant volunteers during an ongoing medical emergency ?

The decision, which volunteers are to be notified, requires a variety of information in order to notify the most promising helpers. False or unnecessary notifications – thus, notifications that will appear immediately irrelevant to the recipient – will have a negative effect on the user acceptance and the system performance, whereas not notifying potential helpers who have high chances to arrive on scene fast enough, will greatly decrease the system's value [5]. Therefore, a selection algorithm is needed in order to produce the best possible set of volunteers at any given moment.

A simple solution for selecting volunteers is the implementation of a notification radius; setting a maximum distance around the place of incident and notifying exactly those volunteers who are within this maximum radius. This approach will provide a set of helpers who are geographically close to the victim, but will they also arrive faster than those potential helpers outside the radius? To answer this question more information is needed, on both the infrastructural situation and on individual user details.

While the actual distance is an important parameter to be considered, it does not necessarily determine the time of arrival at the scene. Due to possible obstacles, the beeline calculation obviously does not offer a suitable background for estimating the arrival time; but even considering roadmap material to calculate the shortest way, does not provide sufficient information without further assumptions. Thus, the type of movement, the physical performance of a volunteer and the current traffic situation, directly influence the approximate traveling time. Furthermore, limiting the relevant decision parameters to merely distance or traveling time appears inadequate and secondary criteria apply; e.g., the potential volunteers' medical expertise, his or her individual knowledge of the area and the current situation this volunteer is involved in [5].

Fig.1 illustrates the general problem; a car driver on a highway (since no highway exit is in reach) and a pedestrian on the opposite side of a river (the next bridge is too far away) are both within the notification radius but will most likely have problems getting to the victim in time. On the other hand, assuming that the train goes in the right direction, a volunteer traveling in the train will, due to the incident being very close to the next train-station, arrive at the victim within a short amount of time even though he or she is far out of the notification radius. If the bicycle rider in the scenario will get to the victim in time (even though again out of the notification radius) depends on the physical performance of the rider and if the road is actually going downhill or uphill.

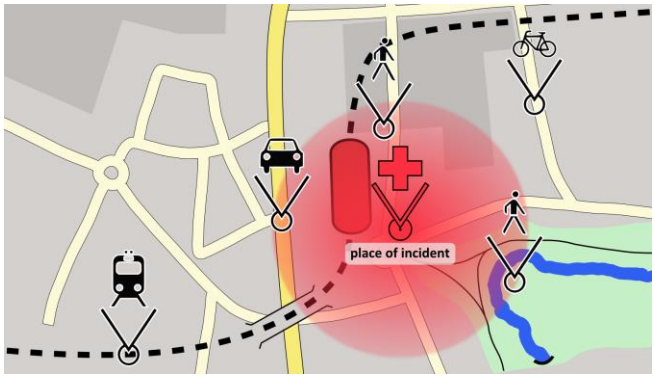


Figure 1: The general problem of an intelligent volunteer selection

At this point, it becomes apparent that a scientifically based answer to the question, "which volunteer should be notified for an individual case?" requires detailed information on the available volunteers. This approach is characterized by an extensive user profiling, therein gathering and processing data from different sources to enable an efficient decision within a limited information environment. The goal is to create a prescient system, which uses historical data and various concepts of artificial intelligence (AI) in order to calculate reliable predictions on any given aspect of the based user profile. Details on this approach and the underlying concepts will be discussed in the upcoming sections of this paper.

E. How can the selected volunteers be notified ?

The general approach within the VNS is to notify potential helpers by using available smartphone technology; a variety of messaging technologies and frameworks are existent at the moment, but no matter the usage they all have one thing in common: any location that is requested on the mobile client and any data that is sent from the client to the server consumes battery power. In recent tests, done within the project scope and with up-to-date smartphones, continuous real time connectivity from different mobile clients to the VNS server, sending any available location update and data changes instantly to the server without taking any precautions regarding battery life, drained the battery to zero within approximately five hours. Therefore, in order to provide an acceptable solution that is suitable for every day scenarios without having constant access to energy sources, a compromise between the server staying up-to-date and the mobile clients' battery load has to be found.

Another parameter to take into account is the probability of weak internet coverage in specific zones. These can result in non-frequent location updates and connectivity issues regarding the connection reliability between the mobile client and the VNS server. Thus, notifications sent might never arrive or the systems' selection algorithm might consider volunteers whose actual location differs significantly from the last successful location update sent to the server. While the actual connectivity and the quality of the technical implementation have a direct influence on the location data that is available on a user, this data is also known to have fluctuations by default, which require additional measures in order to restrain further implications on the selection algorithm. Technical details regarding "how" to notify the relevant volunteers are being discussed in Section II as part of the existing mobile and messaging technologies.

II. State Of The Art

A. Existing notification systems

While some local approaches to implement notification systems (e.g., [7] [8]) exist already, those approaches generally do not have an academic motivation or background. Therefore, publications on the aspect are still rare and the corresponding projects are neither opening their expertise nor the source codes to the public. The only publically available resources are the corresponding application download, some basic usage documentation and a reference document for the Advanced Programming Interface (API) – which merely offers functionality for providing the systems with data [7].

Based on reviews and the appearance in media all over the USA, the PulsePoint Foundation for example offers one of the most advanced software implementations in the field of emergency notifications at the moment [9]. Formerly known as the "Fire Department App" and developed for iOS only, the new version is available under the name "PulsePoint" for Android and iOS [7] [8]. Even though this application is surely great for offering everyday people a possibility to save lives, based on the available documentation, it is a US-only solution without open interfaces. From an academic point of view, it is regrettable that the achieved competences are not shared and the source codes are not publically accessible, which makes it is nearly impossible to use the project as a base for a scientific work. Furthermore, the implementation approach is rather static, allowing two types of mobile devices (Android and iOS) as recipients and no other but US specific regulations, legal circumstances and network characteristics are supported. There are a few smaller projects with less impact and publicity, but the problems stay the same and none of the available solutions actually considers an "intelligent" user selection or a bidirectional communication stream.

Beside the difficulties stated above, the available solutions are implemented as local solutions that cannot easily be adapted to other countries, regions or new legal environments. Fundamental changes are needed in order to use these systems with other than the original parameters and the underlying model itself does not provide a reasonable extension of functionality without making changes to the actual source code itself. In summary, the currently available systems lack essential interfaces, public tools for gathering and extracting information, an efficient communication flow and basic concepts for extensibility. Thus, the demand for an intelligent system, that can efficiently forecast the volunteers' location and performance, requires an entirely

new approach, which is introduced in Section III as part of the "EMuRgency" project.

B. Mobile technologies

Advances in mobile technologies and the continuous growing popularity for portable digital devices with internet access in nowadays society offer a great starting point for VNS. Without supplying any special devices, a VNS is able to communicate with a huge variety of volunteers by simple using the existing hardware and infrastructure that people own and use every day. Modern smartphones for example offer a diversity of features that may be used to aid potential helpers in their mission to arrive on scene as early as possible. Some notable built-in features are real time internet connections, notification options with vibration and sound, photo and video modes, a variety of sensors to enable situation based functionality like a compass, and the fact that actually any modern mobile device is running an operation system (OS) that supports programmatic solutions for individual software [10].

Based on the basic definition of a VNS, the core functionality of any VNS is the effective localization of the volunteers. The actual localization of mobile devices within a network is a complex matter, while the reliability of the results generally depends on the corresponding network provider and its infrastructure [10]. Different companies and research groups are working on this topic, offering a variety of Advanced Programming Interfaces (API's) with base functionality to access localization data for different types of devices. One of the most advanced examples is Android's Location API, which is part of the Android software platform, developed by Google in conjunction with the Open Handset Alliance (OHA) [10].

The OHA is a consortium of various companies, working on developing and advancing open standards for mobile devices. The consortium, led by the Google Incorporation, includes some of the biggest mobile operators like Telekom and Vodafone, as well as some important manufacturers of mobile devices like Samsung and HTC. As an open-source project, the Android source code is publicly available and can be accessed freely [10]; this reflects in a high user acceptance and fast development progress due to contributions from the open source community. The comScore Incorporation frequently publishes reports on the mobile subscriber market worldwide and recently published the numbers for the first quarter of 2013, showing that Android is holding an average of more than 70% market share within the biggest countries of Europe [11]. In a press release from August 2012, the International Data Corporation (IDC) states that the Android market share is continuously increasing [12]. Both studies are based on device sales in the corresponding regions and therefore reflect the general tendency within the segment of smartphones and other mobile devices with internet access.

Even though restricting the notification recipients to exclusively smartphones and similar devices running Android is questionable, it seems to be a reasonable decision for rapid prototype development in order to provide an early running system as soon as possible. It must clearly be stated that a limitation of this kind can only be temporary and that a final model of a state-of-the-art VNS has to provide a generic communication approach in order to support a broad variety of different devices. A more detailed discussion on the topic of a possible generic approach will follow in the upcoming sections of this paper.

C. Messaging Technologies

The Hypertext Markup Language (HTML) defines the core language of the World Wide Web (WWW). With the HTML 5 specification becoming the new standard for web interactivity, a lot of features are accessible for programmers to enable client and server technologies to communicate with each other. While a detailed discussion on server push technologies and HTTP requests would clearly exceed the context of this paper, it is important to note that the HTML 5 specification includes full support for so-called WebSockets. WebSockets specify an API as well as a protocol, while the protocol defines the HTTP handshake behavior to switch from an existing HTTP connection to a lower level connection; a so-called WebSocket connection. While a common approach over the last years was to simulate a server push channel over HTTP, a WebSocket connection enables bidirectional communication natively [13].

The Communication between VNS server and volunteer's devices (the clients) is a central part of the VNS' system architecture. To support the VNS core functionality it has to transport different types of messages:

- Frequent location updates from the mobile clients to the VNS server.
- Information streaming from the server to the clients, containing information on upcoming events, general content or medical tutorials.
- Case notifications and bidirectional real-time updates between the server and the clients, which have been selected as potential volunteer in an ongoing case.

The architecture and technological choices for this communication have major implications for the system's openness regarding device platforms, the timeliness of message delivery and both client and server performance. Client performance and thus the user experience, is most notably influenced by their device's battery life time while the VNS client application is active on the device. As mentioned in Section I, not draining a device's battery is an important non-functional requirement to the system. Server performance can be characterized by the number of clients that can be handled concurrently by a defined set of physical hardware.

Whereas a manifold of approaches exist to realize this communication, a first architectural choice is to use a message queuing based communication pattern, which conforms to the current state of the art and reduces the solution space considerably. This communication pattern defines that peers communicate by means of asynchronous

message exchange. Messages are sent by publishing them to one or more queues. The published messages are then received by subscribers of these queues. This communication pattern – often referred to as publish/subscribe or short pub/sub – is well established in enterprise integration and many protocols, frameworks and implementations exist, which realize this pattern. This pattern usually involves a central message broker or a cluster there of which holds the queues and routes the messages between all connecting peers [14].

Using this communication pattern with mobile devices imposes requirements onto realizations of this pattern, which are not present when communicating merely between fixed peers; like the frequently changing quality of service and bare availability of network connections between a mobile device as peer (MDP) and the broker. Also the hibernating of the MDP – in order to reduce power consumption –results in new requirements, since an MDP is (ideally) continuously connected to the broker but will not be able to receive or send messages while hibernating.

Based on the discussed requirements and corresponding with the introduced pattern, the following three frameworks have been chosen for a detailed assessment:

- Simple Text Oriented Messaging Protocol (STOMP) in combination with WebSockets as initial take, in order to use a web technology as basic transport and STOMP on top of it for the pub/sub semantics [15].
- Google Cloud Messaging for Android (GCM) is a native Android messaging framework. It delivers great reliability but implements a merely one-way communication (server to client push) [16].
- MQ Telemetry Transport (MQTT) is, according to its developers "…a publish/subscribe, extremely simple and lightweight messaging protocol, designed for constrained devices and low-bandwidth, high-latency or unreliable networks. The design principles are to minimize network bandwidth and device resource requirements whilst also attempting to ensure reliability and some degree of assurance of delivery." [17].

The assessment's results were gathered by performing a review of literature, developer resources and the web as well as by implementing prototypes and performing preliminary tests within the project's system implementation; they are summarized in Table I. The frameworks were judged by the following four parameters:

- Connection Management refers to a framework's ability to handle the special network conditions of mobile devices. I.e., does the framework have versatile capabilities in this regard or is it necessary to cope with the challenges in the client code?
- Platform Interoperability refers to a framework's ability to be used on different mobile phone platforms such as Android and iOS.

Table I: Assessment of messaging frameworks for the VNS

	Connection Management	Platform Interoperability	Resource Efficiency	Messaging Semantics
STOMP	-	+	-	o
GCM	o	-	+	o
MQTT	+	+	+	+

- Resource Efficiency refers to a framework's utilization of network bandwidth and central processing unit as well as its ability to let a device enter lower sleep states and thus hibernate in order to save battery life time. This parameter is a very rough estimate based on a framework's architecture and not based on actual measurements. It thus has to be taken with great care.
- Messaging Semantics refers to the versatility of a framework's messaging protocol regarding message handling like Quality of Service (QoS) aspects, publication/subscription schemes and message life time.

Whereas the VNS implementation within the project was originally based on WebSockets in the early prototype versions, a variety of connectivity issues were encountered during the prototype evaluation. With HTML5 and therefore WebSockets still being a new and a not yet fully established technology, any detailed discussion on possible explanations would clearly exceed the intention of this paper. As a result to the unsatisfying performance of the WebSocket implementation, recent builds of the EMuRgency VNS are entirely based on the MQTT connectivity protocol, which will be discussed in the upcoming section.

D. The MQ Telemetry Transport

MQTT is designed as a protocol that transfers the enterprise integration pattern of publish/subscribe message queuing to telemetry nodes connecting to a message broker over constrained networks [17]. The current version 3.1 of the protocol specification, as provided by IBM and Eurotech, has been submitted to OASIS for general standardization [18] [19]. In detail description of the protocol can be acquired from various online resources, such as [17] [20]. As our own reviews show, scientific literature covering the protocol is sparse at the moment. The following gives a brief summary of its main distinguishing features.

The publish/subscribe scheme provided by MQTT exclusively uses topics, which can be arranged in hierarchies. In order to subscribe to multiple topics in a given hierarchy, wildcards can be used for different parts of the hierarchy when subscribing. For strict unicast messages, a topic per peer must be used. As there is no explicit notion of queues in the MQTT protocol, all peers subscribing to the

same topic get their own copy of a message, thus enabling easy broadcast realization by shared topics.

Three important features of MQTT are Durable Connections, Retained Messages and Wills. If a client defines its connection as durable, the corresponding broker will use a storage location in order to persistent the relevant subscriptions and store non delivered messages; e.g., while the client is disconnected. On reconnect, these messages are then published to the client. The most recent message per topic, which was marked as retained, is stored by the broker and delivered to every newly connecting client. Upon connection, a client can define a so called "will message" on the broker. In case the broker detects an abnormal disconnection of the client (i.e., a time out), it publishes the client's will message, which can therefore inform potential subscribers of the up-to-date system status and invoke local adjustments.

E. Positioning, Prediction and Profiling

Section I.D states, that a major problem is the answer to the question "which volunteers should be notified in case of an incident?" and that the rather simple solution "Just notify everyone within a specific radius" will often lead to counterproductive notifications and is not always technologically achievable. The technological constrains introduced in Section I.E result in an increased lack of information, which therefore must be forecasted based on the historical data that is available within the system. The next three sections describe the state-of-the-art concerning the possibility to use infrastructural knowledge for positioning, generate a volunteer profile and predict the actual position and behavior of an individual volunteer.

F. Positioning using map-matching algorithms

Map-matching is a process, which was initially defined to map the inaccurate position of a person or vehicle retrieved via GPS to a valid position of a provided road network [21]. The map-matching problem is illustrated in Fig. 2.

Different map-matching algorithms were introduced over the last years, which are generally grouped into geometric, topological and probabilistic algorithms [22]. Depending on the input-data requirement, different algorithms to determine the actual position are available. Geometric map-matching algorithms expect a transmitted position and a map, whereby topological algorithms also expect the availability of previous positions (i.e., a trajectory). Probabilistic algorithms use mostly the same input as the topological algorithms, but additionally predict several actual positions with a probability [23]. There are a few enhanced algorithms, which use further data (i.e., temporal information) and additional techniques like dead reckoning

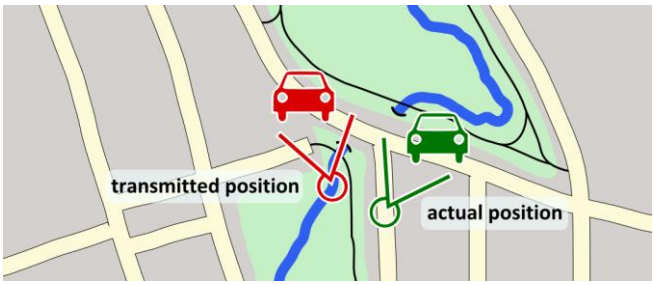


Figure 2: The map-matching problem

or trajectory prediction, in order to increase the quality of the map-matching [24] [25] [26]. Those algorithms are often adapted to a specific use-case and therefore difficult to apply to other cases.

Furthermore, the map-matching algorithms can be divided into so called on-line and off-line algorithms [22] [27]. On-line algorithms are mainly used to determine the actual position in real-time, whereby the latter are used to determine the trajectory (or just points of it) in a post-processing step having the complete set of previously obtained geo-referenced points available.

1) Positioning based on dead reckoning techniques

Dead reckoning techniques are used to approximate the position of an observed unit (e.g., a pedestrian, vehicle or ship) based on some stationary information (e.g., the last valid position) and collected, non-positioning information (e.g., traveling direction, destination, or average speed). Dead reckoning is widely used and applied in different research areas; e.g., autonomous driving [28], robotics, pedestrian or indoor positioning [25] [29] [30]. The techniques can be divided into the variants, which are illustrated in Fig. 3.

The techniques presented in this publication often vary in the sensors used (e.g., compass, accelerometer, motion measurement unit or odometer) and the way they utilize the available data. In general, dead reckoning techniques are good alternatives or complements when positioning data (e.g., GPS, WIFI- or cellular-positioning) is unavailable or insufficient (e.g., indoor, underwater, bad reception) but additional information like acceleration, speed, or direction are available. Furthermore, dead reckoning techniques are one approach to project a possible future position [31].

2) Trajectory prediction

A projection into the future is realized by a so called "trajectory prediction". This prediction is based on an initial location and other none-positioning sensor values. Another approach to forecast the position of a unit is based on data-mining techniques. Those techniques normally work in two phases.

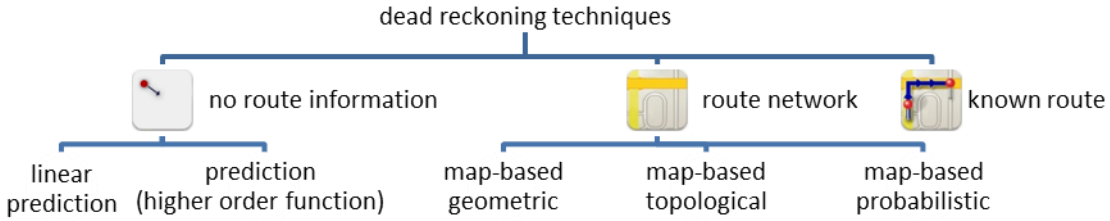


Figure 3: Variants of dead reckoning techniques

The first "learn-phase" is used to detect the trajectory patterns within the collected data. In the second "predict-phase" the current trajectory is matched against the discovered patterns. The matching patterns are finally used to project the current trajectory into the future [32]. Some techniques also combine the two phases and use the currently observed data to continuously enhance the patterns on-line [33]. The two phases of data-mining for trajectory prediction are illustrated in Fig. 4.

Generally, there are two data-mining techniques used for the "learn-phase", i.e., the pattern detection. These techniques are clustering [34] [35] and sequence mining [36] [37]. The difference between them is how they interpret (i.e., model) a trajectory within the system. Therefore, the results and usability vary concerning the kind of patterns that are looked for (e.g., region-, trajectory-, behavior-patterns). The "predict-phase" is depending on the technique applied and the usage of the discovered patterns, based on Kalman Filters, Hidden Markov Models or directly on the patterns [32] [33].

3) Profiling

The context of the historical trajectories used to determine the patterns (as described in the previous section) often depends on the use-case. For example: It might be necessary to find patterns within the collected data of a specific day and from a specific person, or it might be sensible to look at all the trajectories of a specific group of people (e.g., clustered "a priori" by age or job). Profiling does address this type of problem, i.e., the grouping or categorizing of data, so that a generalized profile is determined. A profile can thereby be generated for one entity (e.g., a person, a city) or a group of entities (e.g., all persons of a specific age, all cities with a specific population).

Profiling does use and combine different techniques from

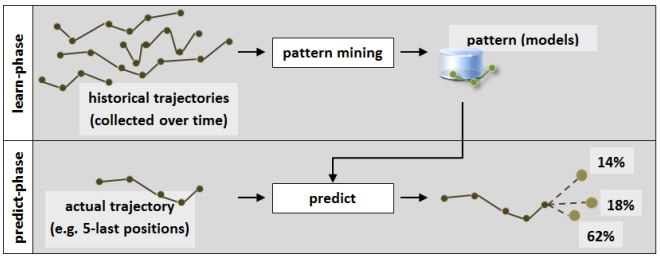


Figure 4: Two phases of data-mining techniques used for trajectory prediction

different research areas, e.g., often data mining and machine learning techniques. A state-of-the-art overview concerning profiling is very extensive and extends the intention of this paper. The goal of this publication is to highlight the necessity of an intelligent user selection and give a short introduction on the topic of volunteer-profiling in order to predict his/her current situation concerning the discussed restrictions. Especially a reliable prediction concerning the current position and situation is important, so that infrastructural constraints and situational limits can be foreseen. The introduced algorithms and techniques of the previous section cover the state-of-the-art concerning possible infrastructural constraints [38]. Some of them are also likely to be adapted in order to detect situational limitations. Thus, the type of used vehicle (e.g., in a train, afoot), possible company (e.g., with children) or individual circumstances (e.g., under time-pressure) can be implemented case by case or the system has to adapt itself over time by implementing learning adapted algorithms.

III. THE EMURGENCY PROJECT

The European research project "EMuRgency" started in September 2011. Research facilities from Germany, the Netherlands and Belgium are working together on modeling and implementing an integrated Volunteer Notification System (VNS) to gap the time between an incoming emergency call and the arrival of professional helpers at the scene. The name of the project is a composition of the two words "emergency" and "urgent" and refers to urgent help that is needed in case of SCA. The three upper case letters "EMR" identify the regional base of the project; the "Euregio Maas-Rhein" (Eng. "Meuse-Rhine Euroregion").

A. Definition of the term "volunteers" within a VNS

Before describing the system, its components and the technical details, it needs to be clarified, which group of people can actually participate as volunteers within a VNS. A volunteer can be anyone with basic skills in first aid and CPR (Cardiopulmonary Resuscitation) who is willing to help in case of an emergency. It is important to differentiate this definition from the term "first responder", which was defined by US National Highway Transportation Safety Administration as "the first medically trained responder who arrives on scene of an emergency" [5]. While the definition of a first responder includes groups like police officers, firefighters and EMS, it does not include laypersons since those generally do not have medical training. Still,

laypersons might be able to provide the needed measures in order to help victims of SCA and thus should be included as potential helpers within a VNS [39]. Within the EMuRgency project, the term "volunteer" is referring to any potential helper, medically trained or not, willing to aid other people in an ongoing emergency [1].

B. Integration of VNS and professional EMS

Whenever an incident is reported to an emergency dispatch center that might involve SCA, the dispatchers will do what they normally do: send professional help – but optionally also invoke the VNS. It is important to stress that the VNS, at this time of development and based on the way EMS are organized today, is a merely optional feature. This means that the responsible dispatcher may or may not involve the VNS, depending on their analysis of the case and personal motivation. In order for the optional integration to be achieved, the VNS has to provide a user-interface where the dispatcher can initiate a case by forwarding its exact location and some optional information to the system. During SCA, time is of utter importance, so this user-interface has to be as simple and efficient as possible.

Within the project, interviews were made to determine the acceptance and motivation of the dispatchers to integrate a VNS within the general workflow; even though all the interviewed dispatchers agreed on a potential benefit, it became rather clear that the general acceptance of new systems seems to directly correlate with the extra work that is involved in order to use it. Taking into account the discussed optionality and the still early stage of development within the project, a manual integration will be the starting point towards involving the VNS within the professional EMS workflow. The implementation of an integrated system that gets activated fully automated during a reported emergency is surely desirable, but requires detailed collaboration with the corresponding software providers. At the current stage of development, different regional dispatch centers expressed their goodwill to work together with the project consortium and enable a real-life test scenario within the next evaluation period upcoming in 2014. The details on this cooperation are not yet fully discussed and are also a matter of various legal restrictions regarding the different countries [40].

C. The Decision System

As soon as the system receives information on a new emergency, no matter if automated or manually initialized from the dispatcher, the VNS will determine a set of volunteers, which are to be notified of the incident. In order for this to be possible with minimum time effort, the system needs to be "aware" of all potential volunteer locations at the time of the incoming emergency call. This awareness can be achieved by making all connected clients publish their locations to the server in pre-determined time intervals or whenever a significant change occurred. It is important to understand that this location updates arriving on the server will not be interpreted as the effective actual location, but rather as a possible user location with a significant high probability.

As discussed in Section I, the location is not the only decision parameter for a volunteer to be considered. Secondary criteria apply and the actual decision process is influenced by different criteria and a variety of available data on each potential volunteer; this data is either statically available or dynamically collected within the system and then taken into consideration, using the criteria, in order to create a set of results, the decision. The following three categories represent the different types of data available within a volunteer notification system [5].

The first category of data is requested during the actual registration of a new volunteer; the medical competence level, the contact data and personal details like the birthdate are some examples. This data represents a static layer of information; the content of this information is unlikely to change and potential changes are generally not requested by the system but instead triggered by the volunteer – e.g., supplying new certificates of medical training or changing the address.

The second category of data is characterized by dynamic data and actively collected by the system. The actual location of a volunteer, represented by a combination of latitude, longitude and altitude, or an available acceleration index on a mobile client, are some examples for frequently changing data that is frequently published to the system.

The third category is represented by data that is generated by the system itself. Section II discussed various concepts and techniques of data mining in order to create new datasets. This category characterizes the concept of "Prescient Profiling", creating abstracted behavior patterns and route approximations based on the collected data of an individual volunteer.

As shown in Fig. 5, the three identified types of data are building the input base for the actual decision system, whereas the final output will be a set of volunteers to be considered. With various types of data available, a decision has to evaluate the relative importance of each of them; this evaluation must not be constant but instead occur frequently and adjust over time, assessing the quality of past decisions in correlation with the current learned knowledge base [41]. Different learning algorithms apply and need to be evaluated against each other. Whereas the basic data (static and/or dynamic) will by itself result in a plausible set of volunteers already, an enhanced set of volunteers will be selected when processing the third category of generated data. In order to produce valid assumptions on the efficiency of the selection algorithm, both sets need to be analyzed and compared.

D. Webservices

In order to enable external sources, for example mobile clients or applications from project partners, to access specific functions on the server, accessible interfaces are required. Specific APIs provide any kind of clients the necessary functions in order to communicate with the VNS without having to access a web browser. The same approach provides the different project partners with an interface to integrate alternative ways of volunteer registration into the main system. Commonly used solutions for offering a limited scope of functionality to external clients are so called

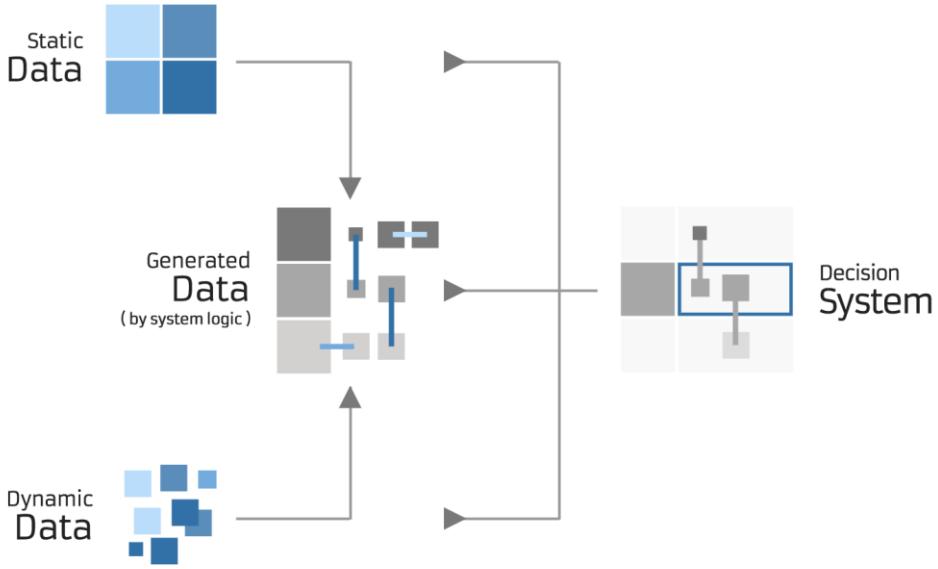


Figure 5: Different types of data utilized by the decision system

WebServices or more specific, REST based Web APIs [42]. By implementing Webservices, a predefined set of functions will become externally available; different devices will be able to communicate with each other over a given network infrastructure; like for example, the World Wide Web (WWW).

E. The messaging architecture

At the current state of development, an advanced prototype of the VNS is available and running in different test scenarios. The upcoming section will introduce the used messaging architecture and describe the essential components implemented for the different types of communication.

As long as a client is not involved in an ongoing case, the main communication occurring between the mobile clients and the server is defined by a client-to-server location publish; that automatically gets invoked in a predefined time interval or whenever the location of the mobile client significantly changed. As current implementation of the location framework, the Google Maps V2 API has been utilized, which requires Google Play Services in order to work [43]. While the first implementation approach within the early prototype versions was based on HTTP requests for posting and pushing this data [1], the current messaging architecture merely utilizes the MQTT protocol to exchange data between the client and the server [18]. HTTP requests in general produce high latencies and require a strict order of POST and GET request. Furthermore, they do not allow a server to client message push, which is required for the actual notification event. Furthermore, bidirectional real time connectivity is required to enable a variety of supportive features; for example chat-channels for notified helpers and the option to enable a graphical feedback for the dispatcher of the current situation on each volunteer.

The WWW was originally not intended to support bidirectional communication and therefore does not include the corresponding specifications or protocols. With HTML 5 introducing the WebSocket JavaScript interface, a native solution for bidirectional communication is available. Once a WebSocket connection has been established between client and server, instant data communication is supposed to be possible between both sides without explicitly having to deal with technical differences. The experiences with this technology, as stated in the sections above, were disappointing, regarding the reliability and quality of the connection. After various test runs, the decision was made to entirely switch to the MQTT messaging protocol, which has been introduced in Section II.

The MQTT implementation within the server-side applications is a more or less standard procedure. A connection to the broker is established and kept alive with reoccurring pings. In case of an unexpected disconnect, a new connection will be established. Since the server is running both the MQTT broker as well as the servlets utilizing the connection to the broker, incoming messages will be handled by the corresponding functionalities without having to care about lifetime-cycles or explicit queuing. Receiving and sending MQTT messages in this architecture are both straight forward implementations regarding the service-side.

The mobile client side of the MQTT implementation is an entirely different matter. Without going too much into detail about mobile client programming, there are a variety of architectural features that have to be taken into account. As mentioned in Section I, the battery consumption is of high importance, since a high CPU load can drain a fully

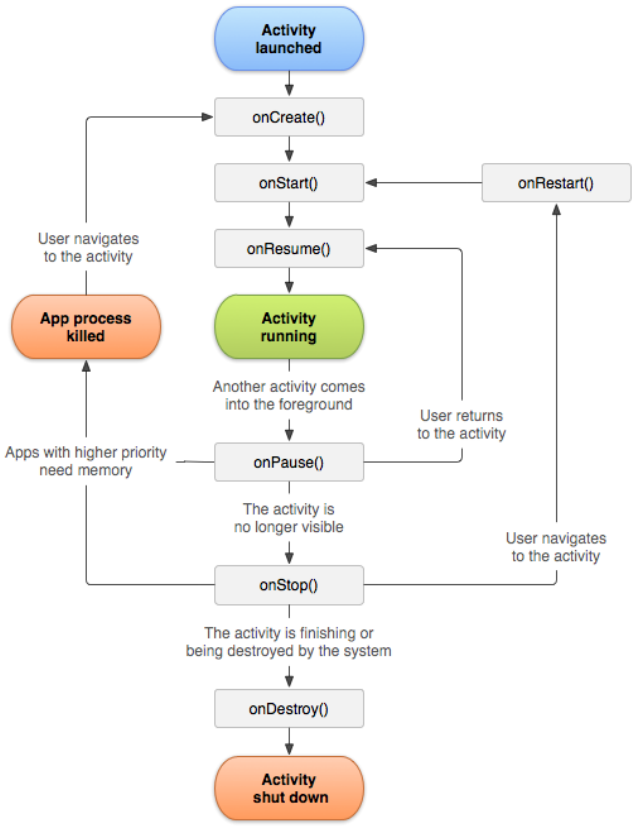


Figure 6: Lifecycle of an Android activity

charged battery within a few hours. To prevent this, smartphones usually have a so called "sleep mode", which is generally divided into a partial (only the screen turn off) and a full (the CPU goes into sleep as well). The problem is that a mobile device, which is sleeping, has very limited functionality, actually only device specific services are still running, everything else is paused; including running applications. This has various implications on the way an MQTT connection is kept alive, as well as on how incoming messages can be received or handled. The android operation system uses so called activities as "views" within an application. Their lifecycle is illustrated in Fig. 6. An application can have multiple activities running simultaneously and influencing each other's state. When the Android device goes into sleep mode, all activities will be paused (even the last displayed one) and no more functions will be called, not even if explicitly invoked by a running service [44]. Furthermore, Android might recycle any running activity during sleep mode. Since a continuous MQTT connection is required in order to implement the needed functionalities (esp. the notification), an Android background service has to handle the connection and wake the CPU (partial wakelock) every now and then to tell the broker that it is still valid or reconnect in case of a timeout. The service has to run in the same thread in order to prevent leaking. This approach delivers a steady MQTT connection on the mobile client, but does not solve the problem of paused activities. In order to handle incoming messages, the MQTT service therefore has to wake the CPU and afterwards broadcast the message to its subscribers (i.e, broadcastReceiver). Due to the lifecycle of created activities, the subscription to the broadcasted messages must not occur for succeeding instances and must be canceled in case of specific event.

Since the delegating service actually runs independently from the application and any activity, a MQTT wrapping framework has been developed within the project, which delivers an elegant way of keeping the MQTT connection alive by also providing base functionality for sending and receiving messages from a configurable MQTT broker. During tests the actual messaging implementation delivered reliable connectivity with low power consumption and enables an integrated approach of a consistent messaging architecture connecting heterogeneous devices.

To sum up the implemented messaging architecture at the current state of development, a single messaging protocol (MQTT) is implemented instead of combining various frameworks and technologies (HTTP, GCM, WebSockets), as done in the early prototype versions. While the actual implementation required the refactoring of nearly all system components, the first results are very promising and will be evaluated in detail in the upcoming weeks.

F. An integrated VNS platform

As a scientific project with partners from both technical and sociological research fields, the project is focusing on many more aspects than a simple technical approach for a notification system. Users of the system will be informed on ongoing events or urgent news and since a common interest level of registered people can be implied, communication channels for exchanging know-how and general information are being implemented. A real-time information flow regarding the aspects of first aid and CPR is extending the core notification functionality. While also developing concepts on raising public awareness on SCA, educational content is displayed at frequently used public places; and if in digital form, the streamed content will be enriched or synchronized with data from the VNS. Furthermore, in order to receive substantial scientific results and to determine the potential benefit of a VNS, corresponding reporting and analyzing features are being designed. Open interfaces supply options for non-project members to change or extend functionalities. An integrated VNS platform combines the different research topics with the diversity of requirements that are to fulfill [1].

G. Regional differences for involving laypersons into professional EMS

It is important to understand the actual role of an occurring registration and the resulting implications on the system and the user. A newly registered user for example, obviously wants to help, but comparing different countries, potential differences between the way that laypersons are legally allowed to be integrated into EMS, must be considered [40]. Some regions for example might not allow the integration of laypersons in EMS at all; and is a layperson with first aid skills but without corresponding

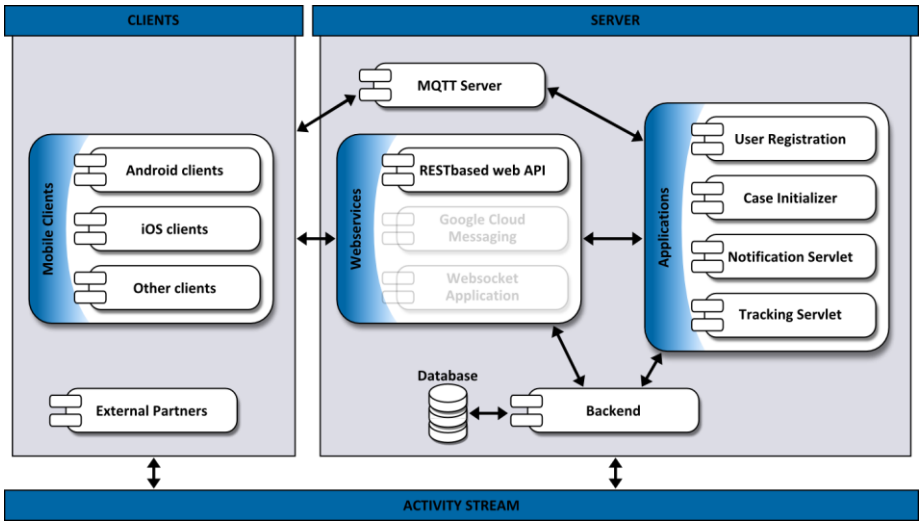


Figure 7: Core components of the EMuRgency VNS

certificates still a layperson? While those questions will not be discussed further within this paper, expert legal advices for different countries are contracted, to validate this matter. The direct implication concerning the VNS is that a new user by default will be "unconfirmed" and will not be considered a potential volunteer until "confirmed"; the confirmation process on the other hand is implemented in a separate administrative component, whereas the final details of this component are again a matter of regional differences that are, at the current stage of development, not yet fully established. With the first real-life evaluation of the system planned for 2014, many of the regional specialties will be addressed in the upcoming months.

H. An overview of the primary components within the VNS

Within the past sections of this paper, the core components of a VNS have been introduced in their corresponding context. Fig. 7 shows an overview of these components while the following paragraph will give some additional information on how they have been implemented.

User Registration: Implemented as web application, this component offers base functionality for new users to register to the system; existing profiles can be edited and specific settings can be configured by the user. Due to the implemented Webservices, alternative ways for sending a user registration are available; users can for example directly register using the mobile client.

Case Initializer: This application constitutes the actual data provider for new emergencies until the first integration into a dispatch center is available. Intended for dispatchers only, this web application provides the possibility to manually initiate a new case by providing general information on an ongoing emergency and the corresponding unique location, as pair of latitude and longitude. In order to supply a user friendly interface, the location itself is automatically calculated as an approximation for a given address. This component demonstrates a basic approach towards supplying the actual notification system with case data and is implemented as various JavaScript applications embedded within a shared Vaadin servlet [45].

Server-side applications: This component bundles the server-side functionalities of tracking, localizing and notifying the volunteers; a steady connection to the MQTT broker is established within a messaging servlet and utilized as a shared resource. The Case-Initializer is part of this component and utilizes the shared Vaadin framework as well as the MQTT functionalities. This approach provides fast interaction times when invoking new emergency events or communicating with the Backend and enables the efficient implementation of various secutiry features.

Backend: The Backend represents the interface for persisting data and provides functions to enable communication between the database and other system components. The database implements the Spring-MongoDB ORM framework [46]. Moreover, to ensure a consistent data usage within the system, all referenced data models and structures are defined within this component.

Webservices: Consisting mainly of the REST based Web API, this component actually represents an intermediate communication layer; providing predefined functions for external modules and other clients to exchange data with the server. Compared to earlier versions of the VNS approach, this component no longer utilizes Websockets or the GCM Messaging framework, but instead merely delegates external requests to the shared resources.

IV. Conclusion and Outlook

During the project, sociological and technical aspects are being combined. Country-specific differences in a variety of discussed parameters have been balanced against each other and are being implemented into an integrated VNS platform. Many fundamental difficulties were identified within the

past sections of this paper whereas an advanced prototype of the software is available already - Fig. 8 shows screenshots of the Android client application. This prototype implements the essential components introduced in Section III and enables base notification functionality for nearby helpers. The current system has been evaluated in various test runs and is planned to be integrated in real-life scenarios in the near future. The seamless integration of the different project partners with their interdisciplinary research topics has formed into a complex VNS platform. Current discussions with different regional dispatch centers underline the necessity of such a VNS and will hopefully lead to long-term collaboration possibilities regarding the topic of volunteer integration.

There are a variety of additional features that are being discussed within the project consortium at the moment; they will be shortly introduced in the following section.

A. Aditional features

One promising feature is the integration of existing applications or services that provide information on nearby automated external defibrillators (AEDs). Although the time critical aspect of CPR is the main concern for the first volunteer who arrives on scene, it can prove very useful for further helpers, to have reliable information on nearby AED devices. The project is right now hosting a separate web application gathering data on AEDs within Germany, the Netherlands and Belgium.

The integration of chat channel functionality, enabling direct communication between the dispatcher and all volunteers who accepted a specific case, is another reasonable feature. By utilizing the introduced advantages of the current platform approach and subscribing each client on a case-specific MQTT topic, dispatchers and volunteers will be able to exchange information in real-time. By communicating with each other, the potential helpers will be able to share local information or arrange an AED pickup in the near vicinity of the victim. The dispatcher can thereby supply additional information that has not been included in the original notification.

There are multiple scenarios in which the use of game design elements (gamification) within a VNS can be used to influence the user behavior; the adoption of a score system for attended courses or participated cases is one example. The general idea of gamification elements is to increase the user acceptance and motivation, whereas a sensible consideration is needed in order not to distract the users' attention from the main topic.

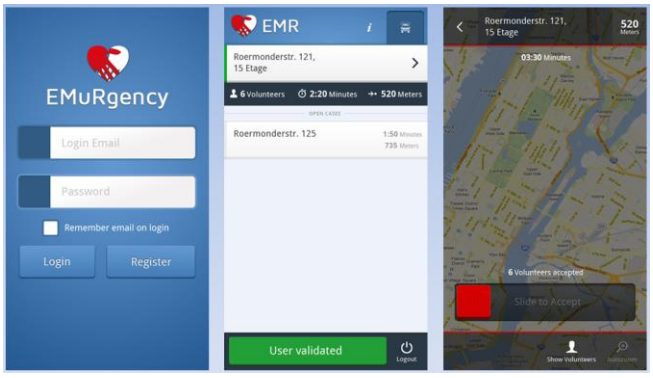


Figure 8: Screenshots of the EMR mobile client (Android version)

Modern smartphones and many other portable devices with internet access offer build-in functionality for photos and videos. With internet bandwidth increasing continuously, a real-time media streaming from the place of incident is another possible feature with high benefit. The integration of telemedical concepts (e.g., the regional project TemRas) becomes possible and thereby enables an approach of professional helpers using their expertise to analyze the streamed data, supporting the volunteers at the scene with valuable information [47].

B. Specification of a VNS and future development

Within this paper, the basic concept of a VNS has been described and both the core functionalities and additional functionalities have been discussed. At the current stage of development, a formal specification of a VNS is not yet possible, since many parameters are not yet fully determined and legal issues are still being discussed. Diverse aspects (legal, social and technical) still need to be finalized and the system has to be evaluated in real-life, rather than within theoretical test scenarios. The upcoming integration into the professional EMS workflow for a limited period of time will prove very useful for producing valuable data on many of these aspects.

Maintaining a highly agile programming approach will assure a continuous development and a fast adaption of new functionalities, whereas the implementation of the core system reached a notable status regarding stability and reliability already. The mobile clients will be enriched with individual information channels, integrated with the server side profiling and ergonomic design will replace the so far functional driven user interfaces.

The research focus for the upcoming months will be a seamless integration of the core system with new components implementing the introduced concept of "Prescient Profiling" in order to enable an intelligent volunteer selection and produce scientific output regarding the question of "who are the relevant volunteers in case of time critical medical emergencies".

ACKNOWLEDGMENT

This paper is based on work done in the INTERREG IVa project EMuRgency (www.emurgency.eu). The project is partially financed through the European Regional Development Fund (ERDF) and co-financed by several regions of the Meuse-Rhine Euroregion and partners of the EMuRgency consortium.

REFERENCES

[1] J. Elsner, M. T. Schneiders, M. Haberstroh, D. Schilberg, and S. Jeschke, "An Introduction to a Transnational Volunteer Notification System Providing Cardiopulmonary Resuscitation for Victims Suffering a Sudden Cardiac Arrest", Proc. of the Fifth International Conference on eHealth (eTelemed2013). IARIA, April 2013, pp. 59-64.

[2] Studie des Instituts für Notfallmedizin und Medizinmanagement (INM) - Klinikum der Universität München, "Untersuchung zum Notarztdienst und arztbegleitenden Patiententransport in Bayern", April 2010.

[3] J. Bahr, "Lay resuscitation in the German rescue system. Some basic information", vol. 10, no. 3, Notfall Rett. Med 2007, pp. 197-200, doi:10.1007/s10049-007-0910-y.

[4] K. Grassl, "Herzdruckmassage im Rahmen der Laienreanimation: Einsatz eines audiovisuellen Echtzeit-Rückmeldesystems". Dissertation, LMU München: Faculty of Medicine, 2009.

[5] J. Elsner, M. T. Schneiders, D. Schilberg, and S. Jeschke, "Determination of Relevant First Aiders within a Volunteer Notification System", Proc. of the International eHealth, Telemedicine and Health ICT Forum for Educational, Networking and Business (Med@Tel 2013), Luxembourg, 2013, pp. 245-249.

[6] D. L. Kasper, E. Braunwald, S. Hauser, D. Longo, J. L. Jameson, and A. S. Fauci, "Harrison's principles of internal medicine", 16th ed, 2004.

[7] PulsePoint Foundation, www.pulsepoint.org, downloadable as "PulsePoint" App on Google Play and Apple iTunes, (accessed 08/12/2013).

[8] The San Ramon Valley Fire Protection District, www.firedepartment.org, "Fire Department iPhone App", (accessed 08/12/2013).

[9] Tracy Rosecrans , "Top 10 Heart Disease Apps of 2012", Healthline article, September 2012.

[10] K. Roebuck, "Location-Based Services (LBS): High-impact Strategies - What You Need to Know: Definitions, Adoptions, Impact, Benefits, Vendors", Emereo Pty Limited, 2011.

[11] Electronic Publication: comScore Reports June 2013 U.S. Mobile Subscriber Market, comScore Incorporation, 2013.

[12] Electronic Publication: International Data Corporation (IDC) - Press Release, "Android and iOS Surge to New Smartphone OS Record in Second Quarter", August 8, 2012.

[13] P. Lubbers, B. Albers, and F. Salim, "Pro HTML5 Programming: Powerful APIs for Richer Internet Application Development", Apress, 2010.

[14] G. Hohpe and B. Woolf, "Enterprise integration patterns: Designing, building, and deploying messaging solutions.", Addison-Wesley, Bosten 2004.

[15] http://stomp.github.io/ (accessed 08/10/2013).

[16] http://developer.android.com/google/gcm/index.html (accessed 08/14/2013).

[17] http://mqtt.org/ (accessed 08/02/2013).

[18] IBM and Eurotech, MQTT V3.1 Protocol Specification. Available: http://www.ibm.com/developerworks/webservices/library/ws-mqtt/index.html (accessed 08/20/2013).

[19] https://www.oasis-open.org/committees/mqtt (accessed 08/18/2013).

[20] V. Lampkin, W. T. Leong, L. Olivera, S. Rawat, N. Subrahmanyam, R. Xiang, G. Kallas, N. Krishna, S. Fassmann, M. Keen and D. Locke, "Building Smarter Planet Solutions with MQTT and IBM WebSphere MQ Telemetry", IDM RedBooks, September 2012, Available: http://www.redbooks.ibm.com/redbooks/pdfs/sg248054.pdf (accessed 08/20/2013).

[21] D. Bernstein and A. Kornhauser, "An introduction to map matching for personal navigation assistants.", Technical report, New Jersey TIDE Center Technical Report, 1996.

[22] M. A. Quddus, W. Y. Ochieng and R. B. Noland, "Current map-matching algorithms for transport applications: State-of-the art and future research directions", Transportation Research Part C: Emerging Technologies, vol. 15, issue 5, October 2007, pp. 312-328, ISSN 0968-090X, http://dx.doi.org/10.1016/j.trc.2007.05.002.

[23] W. Ochieng, M. Quddus and R. Noland, "Map-Matching in complex urban road networks", Brazilian Journal of Cartography (Revista Brasileira de Cartografia), vol. 55, pp. 1–18, 2003.

[24] Y. Lou, C. Zhang, X. Xie, Y. Zheng, W. Wang, and Y. Huang, "Map-Matching for Low-Sampling-Rate GPS Trajectories," GIS '09 Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, pp. 352–361, 2009.

[25] S. H. Shin, C. G. Park, and S. Choi, "New Map-Matching Algorithm Using Virtual Track for Pedestrian Dead Reckoning," ETRI Journal, vol. 32, no. 6, pp. 891-900, 2010.

[26] J. Eisner, S. Funke, A. Herbst, A. Spillner, and S. Storandt, „Algorithms for Matching and Predicting Trajectories", 13th Meeting on Algorithm Engineering and Experiments (ALENEX), pp. 84-95, 2011.

[27] F. C. Pereira, H. Costa and N. M. Pereira, "An off-line map-matching algorithm for incomplete map databases", European Transport Research Review, vol. 1, issue 3, Springer, pp. 107-124, ISSN 1867-0717, 2009.

[28] Z. Yan, S. Peng, J. Zhou, J. Xu, and H. Jia, "Research on an improved dead reckoning for AUV navigation", 2010 Chinese Control and Decision Conference (CCDC), pp. 1793-1797, 2010.

[29] C. D. C. Randell and H. L. Muller, "Personal position measurement using dead reckoning", in Proceedings of the Seventh International Symposium on Wearable Computers, C. Narayanaswami, Ed. IEEE Computer Society, October 2003, pp. 166–173.

[30] S. Beauregard and H. Haas, "Pedestrian Dead Reckoning: A Basis for Personal Positioning", Proceedings of the 3rd Workshop on Positioning, Navigation and communication (WPNC'06), 2006.

[31] A. Leonhardi, C. Nicu, and K. Rothermel, "A map-based dead-reckoning protocol for updating location information", Parallel and Distributed Processing Symposium., Proceedings International, IPDPS, December 2001.

[32] D. Vasquez, T. Fraichard, O. Aycard, and C. Laugier, "Intentional motion on-line learning and prediction", Machine Vision and Applications, vol.19, no. 5-6, pp.411-425, 2008.

[33] A. Monreale, F. Pinelli, R. Trasarti, and F. Giannotti, "WhereNext: a location predictor on trajectory pattern mining", KDD 2009, pp. 637–646, 2009.

[34] S. Gaffney and P. Smyth, "Trajectory clustering with mixtures of regression models", in Knowledge Discovery and Data Mining, pp. 63–72, ACM Press, 1999.

[35] J. Alon, S. Sclaroff, G. Kollios, and V. Pavlovic, "Discovering clusters in motion time-series data", 2003.

[36] P. Meisen, T. Seidl, and K. Henning, "A Data-Mining Technique for the Planning and Organization of Truck Platoons", Proceedings of International Conference on Heavy Vehicles (HHVT2008), Paris, 2008.

[37] F. Giannoti, M. Nanni, F. Pinelli, and D. Pedreschi, "Trajectory pattern mining", KDD 2007, pp. 330-339, 2007.

[38] I. H. Witten, E. Frank, and M. A. Hall, "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann, Amsterdam, 3rd ed, 2011.

[39] J. Bahr, H. Klingler, W. Panzer, H. Rode, and D. Kettler, "Skills of lay people in checking the carotid pulse", Resuscitation 35, August 1997.

[40] K. Fehn, Legal Advice for the project “SMS Retter”, pp. 13-20, November 2009, unpublished.

[41] M. Hutter, “Universal Artificial Intelligence - Sequential Decisions based on Algorithmic Probability”, Springer, 2005.

[42] L. Richardson and S. Ruby, “RESTful Web Services”, O'Reilly Media, 2007.

[43] Google Maps API V2 as part of the Google Play services SDK for Android (https://developers.google.com/maps/documentation) (accessed 08/14/2013).

[44] R. Meier, “Professional Android 4 Application Development”, Wrox Press, May 2012.

[45] Vaadin}> “Thinking of U and I”, developer homepage: (https://vaadin.com/home) (accessed 08/10/2013).

[46] M. Pollack, T. Risberg, O. Gierke, C. Leau, J. Brisbin, and T. Darimont, “Spring Data MongoDB - Reference Documentation “, (http://docs.spring.io/spring-data/data-mongo/docs/1.3.3.RELEASE/reference/html/) (accessed 08/20/2013)

[47] S. Thelen, S. Bergrath, M.-T. Schneiders, J. Brokmann, T. Brodziak, D. Schilberg, R. Rossaint, and S. Jeschke, “Telemedizin in der präklinischen Versorgung - das Forschungsprojekt TemRas”, e-Health2012, Medical Future 2011, pp. 241-244.

User-Centered Design and Evaluation of an Ambient Event Detector Based on a Balanced Scorecard Approach

Rainer Planinc, Martin Kampel
Computer Vision Lab
Vienna University of Technology
Vienna, Austria
{*rainer.planinc; martin.kampel*}*@tuwien.ac.at*

Stefan Ortlieb, Claus-Christian Carbon
Department of General Psychology and Methodology
University of Bamberg
Bamberg, Germany
{*stefan.ortlieb; ccc*}*@uni-bamberg.de*

***Abstract*—The user-centered design of an e-health system is a complex endeavor: the conflicting interests of multiple user perspectives have to be balanced throughout the design process. Moreover, a common understanding of the interplay between technical, psychological and business aspects has to be developed among multi-disciplinary project partners. In this paper, we describe how the European ambient assisted living project *fearless* deals with these challenges of a user-centered design process based on a balanced scorecard approach.**

Keywords-Ambient assisted living, ambient event detector, user-centered design, balanced scorecard.

I. Introduction

The European ambient assisted living (AAL) project *fearless*[1] ("Fear Elimination As Resolution of Elderly's Substantial Sorrows") is dedicated to fall detection and inactivity monitoring in the homes of solitarily living elderly. For this purpose an autonomously operating ambient event detector is being developed in a user-centered design process [1]: technical specification of the first prototypes is based on two multi-cultural user requirements surveys targeting elderly people (N=259), their relatives and trusted persons (N=215) as well as 22 representatives of care taker organizations. The prototypes of the ambient event detector have been tested extensively under laboratory conditions (e.g., [2]–[6]) yielding promising results. At present these results are being verified in two longitudinal field pilot studies including 45 elderly test users and care taker organizations from Austria, Germany, Italy, and Spain.

The *fearless* consortium comprises ten multi-disciplinary partners from Austria, Germany, Italy, and Spain. While technical partners are working on fall detection algorithms and setting up a telematics platform for alarm handling, two care taker organizations are responsible for the recruitment of elderly test users. Tools for the analysis of user needs and a continuous user-centered evaluation process are designed and deployed by cognitive psychologists from the department of General Psychology and Methodology at the University of Bamberg. Medical experts are counseling the project consortium in terms of ethical issues. Last but not least, business experts are responsible for bringing the *fearless* technology to the market. Market research is conducted and a business model is designed taking different national health care systems into consideration.

User-centered design is essential for the development of assistive technologies. Yet, designing an e-health system becomes a highly complex endeavor as soon as multiple stakeholders are involved: first and foremost, the conflicting interests of complementary user perspectives have to be balanced throughout the design process. Moreover, a common understanding of the interplay between technical, psychological and business aspects has to be developed among multi-disciplinary project partners. In the following sections of this paper, we describe how the AAL-project *fearless* deals with these challenges based on a balanced scorecard (BSC) approach: Section II is dedicated to the current state of technology used in the *fearless* project. In Section III the concept of the balanced scorecard is introduced. An overview of the different stages of user integration within the *fearless* project is given in Section IV. Subsequently, the results of a primary and a secondary user requirements survey are presented in Section V and Section VI. Section VII deals with the Technological Impact Assessment Model, which integrates the results of these two user requirements surveys. In Section VIII, the specific challenges and trade-offs are discussed, which have aroused in the course of a user-centered design process including multiple user perspectives. Finally, in Section IX conclusions are drawn and an outlook is given.

II. Current State of Technology

The structure of the *fearless* system is depicted in Figure 1, showing all relevant interfaces and involved end users. The proposed e-health system consists of sensor units (Xtion Pro™ + small PC for data processing) installed at the elderly's house or flat. The system is adaptable, hence standardized interfaces to third parties are provided (e.g., burglar alarm system, gas detector). Unusual events (e.g.,

[1]This work is supported by the European Union under grant AAL 2010-3-020.

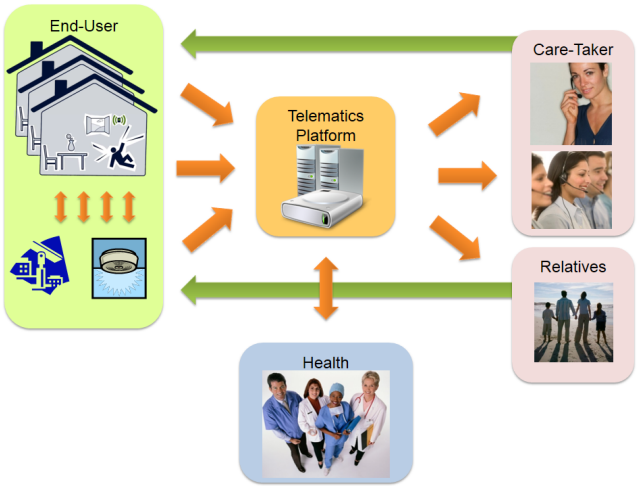


Figure 1. Organizational workflow

falls) are detected automatically and alarms are sent to the telematics platform. This platform enables relatives and care takers to handle alarms. Furthermore, the telematics platform offers interfaces to different standardized electronic health record systems (e.g., ELGA in Austria) to include health professionals.

The *fearless* system works autonomously and raises alarms wihtout any user intervention, since it is important to reduce the cognitive load on the user, especially when dealing with dementia [7]. The use of computer vision is feasible, since it can overcome the limitations of other sensor types [8] and no devices need to be worn. Zweng et al. [6] show that the accuracy of their fall detection approach is higher compared to 2D cameras when using a calibrated camera setup and a 3D reconstruction of a person. Hence, we propose to use a Kinect™ / Xtion pro™, since 3D information is available for distances up to ten meters without the need for a calibrated camera setup. However, the SDK is optimized for a range from 0.8 to 3.5 meters and thus not all features provided by the SDK can be used for higher distances (e.g., NITE). Moreover, the scene can be analyzed in more detail (e.g., estimation of the ground plane) in comparison to standard cameras. Due to the sensor range of ten meters, one sensor is able to cover one room. The total number of sensors per flat highly depends on the layout of the rooms and user preferences. Users with a high risk of falling may choose to equip their flat with many sensors, whereas user with a lower risk of falling may choose to place sensors in rooms with a higher probability of falling (i.e., living room) only. The *fearless* system also works with multiple persons and detects falls not only directly, but also includes additional heuristics to detect "unusual behaviour" indicating a fall (e.g., person has disappeared behind a sofa).

Nevertheless, the use of computer vision raises privacy issues. Due to this fact, the Kinect™ respectively the Xtion Pro™ sensor is used, since depth data can be used to detect falls accurately (e.g., [3], [4]). Using depth data respects the privacy of elderly, since neither the person nor the surrounding can be identified from depth images. A depth image only visualizes the position and the distance to the sensor. Figure 2 (left) shows an example of a depth image illustrating a person, tables and a mat lying on the floor. This visualization illustrates the distances of subjects and objects to the sensor. The brighter the color in the depth image, the further away the person or object is. On the other hand, the darker the object is, the closer to the sensor it is. Furthermore, black indicates that there is no data available (e.g., due to sunlight or reflections). In contrast, the corresponding RGB image is shown in Figure 2 (right), representing the same scene.

The workflow of our approach is depicted in Figure 3: starting with a depth image, the skeleton and ground plane data is extracted by the use of OpenNI [9]. The skeleton joints of the shoulder, spine and the center of the hip are extracted and analyzed using fuzzy logic. Based on the results of the fuzzy logic, a decision is made if the person is in an upright position or lying on the floor. Since only the skeleton joints are used, the privacy of the elderly is respected due to the use of an anonymous and abstract visualization only using lines and dots. An example of this visualization is shown in Figure 4: the dots are representing the upper part of the body, whereas the line represents the major body orientation and the ground floor. In case of an alarm, the alarm including this abstract visualization is sent to the telematics platform, depicted in Figure 3. Moreover, the alarm is stored and forwarded to the appropriate care taker organization or relative by using this platform.

III. Balanced Scorecard (BSC)

The user-centered design of an e-health system is a complex endeavor: multiple user perspectives and their conflicting interests have to be balanced throughout the design process. Moreover, a common understanding of the interplay between technical, psychological and business aspects has to be created among the multi-disciplinary project partners. A performance management system can help reduce complexity by drawing our attention to those aspects critical

Figure 2. Depth image (left) and corresponding RGB image (right)

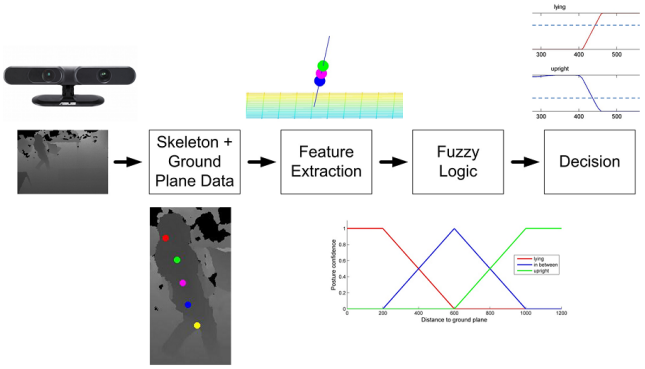


Figure 3. Technical workflow of the *fearless* system

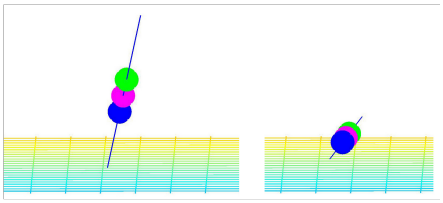


Figure 4. Illustrations used for verification a person standing upright (left) vs. a person lying on the floor (right)

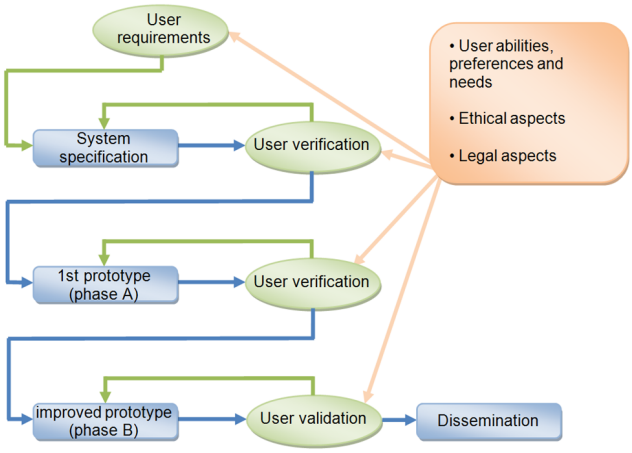


Figure 5. User involvement in the *fearless* project

for success. In 1992, Kaplan and Norton have introduced a new type of performance measurement system to the field of business administration [10]: the Balanced Scorecard (BSC). Based on an explicit "theory of business" [11, p. 110] a BSC combines measures and indicators that reflect different stakeholder perspectives and their conflicting interests. Thus, a BSC enables executives and project managers to "see whether they have improved in an area at the expense of another" [10, p. 1]. Basically, a BSC has to give answers to the following questions:

- Who are the relevant stakeholders?
- What are the needs and expectations of these stakeholders?
- How are stakeholder perspectives interrelated?
- What are appropriate measures and indicators for tracking these needs and expectations?
- Which aspects have to be balanced in the user-centered design process?

In the following sections, we will answer these questions for the *fearless* project. The resulting *fearless* scorecard will be used for project evaluation in the user-centered design process of an e-health system.

IV. Concept and Phases of User Involvement

Figure 5 illustrates the integration of the end users within the *fearless* project: at the beginning of the project, two multi-cultural user requirements surveys were conducted in order to assess the needs (e.g., fears) and expectations of primary and secondary end users. According to the user requirements, an initially defined system specifications are evaluated by test users and redefined on a regular basis in the course of two field pilot studies. Field pilots in combination with semi-structured interviews are conducted to ensure that the *fearless* system is tailored to the end user's needs. Since end users provide regular feedback, the technical specification of the *fearless* system is adapted throughout the project.

Who are the stakeholders of the projected e-health system? In the *fearless* project the design process is centered on two groups of users: Older adults (aged 60+) who wish to use the e-health system in their private home. In the following, this group of users is referred to either as primary users or as clients. Together with their closest relatives and other trusted persons these potential users are involved during user analysis and field pilots. The second user perspective is represented by members of care taker organizations that offer services related to the projected e-health system. Just like the primary users these people will be interacting with the *fearless* system on a daily basis: as technicians they will be engaged in hardware installation and maintenance, as members of the call-center staff they will interact with the telematics platform handling the alarms. In the following this group of users is referred to as secondary users. Secondary users are also involved throughout the design process: from user requirements analysis to field pilot evaluation.

During the field pilots the *fearless* system is installed in elderly's flats in Austria, Germany, Italy, and Spain. The field pilots consist of two different phases: during phase one, the first prototype of the fall detection system is installed in 16 flats (four flats in each country) to obtain first results of the system. Due to these results the prototype is enhanced before phase two with more than 40 installations will be conducted. The aim of the field pilots is not only to test

the fall detection system itself, but also to assess technological and psychological aspects (e.g., housing-related control beliefs) as well as integrating care taker organizations and relatives throughout the project. Furthermore, also ethical commissions are involved during the field pilots to verify the feasibility of the *fearless* system from a legal and an ethical point of view.

The project benefits from the different interdisciplinary perspectives, from which the results of the field pilots are analyzed. From a technical point of view, the fall detection algorithm [4] is tested under real settings and is adapted to the end users' needs while overcoming the lack of realism when performing falls in the laboratory. Furthermore, the overall system including its interfaces as well as the feasibility of the system setup are evaluated. From an organizational point of view, end user organizations are able to integrate the system into their workflow and provide feedback to adapt the system to their needs. Elderly are involved to provide essential feedback to the technical partners. Additionally, we expect to reduce their fears by providing safety while using our system. Since the field pilots are conducted with medical and psychological support, changes and benefits for elderly can be determined and these assumptions can be verified.

In summary, the following user groups are involved in the *fearless* project:

- **Elderly and relatives** are involved during the user requirements analysis and during the field pilots. Elderly install the system in their flats during the field pilots, whereas their relatives can receive the alarms if unusual events are detected.
- **Care takers** are involved during the user requirements analysis and are fully integrated during the field pilots. Hence, the care takers' call centers are integrated and alarms are forwarded to the appropriate call center. This allows for an evaluation of the overall workflow in case of an alarm.

V. Primary User Requirements Survey

What are the needs and expectations of prospective users? Of course user involvement starts with a thorough analysis of primary and secondary user needs. Thus, two multi-cultural user requirement surveys were conducted prior to technical specification of the first *fearless* prototypes [12]. The first survey addressed the needs and expectations of older adults (aged 60+), their closest relatives and other trusted persons. Based on the results of this survey we sought for answers to the following questions: What do elderly people actually fear? Which functions of an e-health system would be desirable in order to resolve these fears? Besides, two more specific questions were targeted: Where do falls occur? How much are users willing to spend on an ambient event detector and the services related to it? 259 potential primary users from Austria, Germany, Italy, and Spain (Catalonia) took part in this survey. The participants were aged between 59 and 101 years ($M = 73.6$; $SD = 8.3$; $Mdn = 73.0$). Additionally, we surveyed 215 relatives and trusted persons of potential primary users. All participants either filled in a questionnaire or participated in a standardized face-to-face interview. Primary users answered questions about resources and deficits of their private home, previous falls and fears related to a broad variety of critical incidents (e.g., falls, fire, housebreaking, etc.). Besides, they were asked to specify preferred functions and an appropriate pricing for a custom-tailored ambient event detector. Relatives and trusted persons answered these questions from a third-person perspective [12]. Across all cultures suffering a stroke in the absence of others was perceived as the most troubling event by primary users and relatives alike followed by events related to falling (Figure 6). Elderly people from Austria, Italy and Spain were particularly worried about housebreaking whereas solitarily living elderly from Austria and Germany reported fear of social isolation. The most preferred functions for a new ambient event detector were fall and fire detection followed by a burglar alarm function, gas detection and inactivity monitoring (Figure 7). In terms of pricing primary users and their relatives stated that hardware must not cost more than 200 Euros whereas monthly expenses for services related to the *fearless* system should be less than 50 Euros. From the survey data a set of seven requirements was derived: maintenance of social networks, stabilization of internal housing-related control beliefs, enhancement of falls efficacy and mobility, reliable fall and fire detection, adaptability (system should allow for additional functions, e.g., burglar alarm), and last but not least affordability [12].

VI. Secondary User Requirements Survey

A second requirement survey was conducted in order to capture the needs and expectations of professional care takers [12]. On the basis of this survey data we sought for answers to the following questions: Which functions should an innovative e-health system provide in order to meet the needs and expectations of care taker organizations? Which aspects of a new e-health system are critical for success? 22 participants from Austria ($n=3$), Germany ($n=7$), Italy ($n=10$), and Spain ($n=2$) either filled in a questionnaire or participated in a standardized face-to-face interview. They answered questions about their previous experience with current tele-care devices (e.g., panic button) and their expectations towards innovative e-health technology. Furthermore, they were asked to specify preferred functions and an appropriate pricing for an ambient event detector. In spite of its size the expert sample reflects a broad variety of different organizations and professions from the field of healthcare: members of research organizations, SMEs and large care taker organizations were included. The participants were employed either by profit ($n=8$) or non-profit ($n=14$) organizations [12]. Across different healthcare systems and individual areas of expertise four functions were favored by

Critical Incidents	Austria		Spain		Germany		Italy	
	Elderly	Relatives	Elderly	Relatives	Elderly	Relatives	Elderly	Relatives
Suffering a stroke	X	XX	X	XX	X	X	X	X
Burglars break into the house	X		XX	XX	(x)		(x)	
Slipping in the bathroom		(x)	X	XX	(x)	(x)	(x)	
Falling down the stairs			X	XX	(x)	(x)	(x)	
Fire caused by a hot plate			X	XX				X
Fire caused by electrical defect			X	XX				X
Fire caused by a hot iron			(x)	XX				X
Fire caused by lightning strike			X	X				X
Burst of a water main			X	X				
Gas leakage			X	X				

Each critical incident was rated on a 5-stepped rating scale from 1= *does not worry me* to 5= *worries me very much*

(x) = average rating between 2.5 and 3.0

X = average rating between 3.0 and 4.0

XX = average rating above 4.0

Figure 6. Prevalent fears of elderly people and their relatives across the four cultures

professionals: fall and fire detection, inactivity monitoring, and gas detection (Figure 7). Requirements for an innovative ambient event detector could be identified as follows: data protection, usability, accreditation, interoperability, and reasonable costs for hardware and services. Due to a severe shortage of trained care personnel in their countries, experts from Germany put special emphasis on low staff intensity in terms of installation, maintenance and handling of the e-health system [12].

VII. Technological Impact Assessment Model

How are the two user perspectives interrelated? What are appropriate measures for a user-centered outcome evaluation? In order to answer these two main research questions the *Technological Impact Assessment Model (TIAMo)* has been developed by psychologists from the University of Bamberg [12]. One of the basic assumptions of the TIAMo is that user perspectives can be arranged hierarchically according to the value added chain. First of all, an innovative e-health system has to meet the needs and expectations of service providers. Otherwise corresponding services will not be available for primary users in the first place - of course this does not apply to the cases in which the *fearless* system is used independently from a care taker organization (e.g., by primary users and their relatives). Thus, our model suggests that the secondary user perspective forms the basis of the TIAMo (Figure 8). In the following sections, we will take a closer look at the requirements for each user perspective. The interdependencies between single requirements are described by arrows. The arrows point to the necessary conditions for each requirement. For example, whether the e-health system is affordable or not depends among other factors on the costs for services, which again depend on staff intensity.

A. Measures and Criteria for the Secondary User Perspective

In accordance with the results of the corresponding requirement survey the secondary user perspective is represented by a set of eight variables: (1) privacy and data protection, (2) usability, (3) reliability of fall and fire detection, (4) staff intensity, (5) accreditation, (6) costs for hardware and (7) services as well as (8) interoperability, illustrated in Figure 9.

Privacy and data protection. Of course compliance with national and European data protection laws is a sine qua non for an e-health system. What are appropriate criteria for evaluating privacy and data protection? Within the European Union Germanys data privacy act is the most restrictive. Thus, it will be applied to the *fearless* system as a standard for the handling of personal data.

Usability. Besides privacy and data protection usability is the most essential requirement. An e-health system has to be reliable and easy to handle. Important pieces of information (e.g., the current status of the client) have to be displayed

	Austria		Spain (Catalonia)		Germany		Italy		Healthcare professionals
	Elderly	Relatives	Elderly	Relatives	Elderly	Relatives	Elderly	Relatives	
Fall detection	X	X	X	X	(x)	(x)	X	XX	XX
Fire detection	X		XX	X	X	(x)		X	X
Burglar alarm			XX	X	(x)		X	X	
Gas detection			X	X			X	X	X
Inactivity monitoring			X	X	(x)	(x)			X
Flooding			X	X					
Monitoring of daily routines			X	X					
Light triggered by motion	X				(x)				

(x) = functions appreciated by at least 30% (only applied to German samples)

X = functions appreciated by at least 50%

XX = functions appreciated by at least 70%

Figure 7. Preferred functions across user perspectives and different cultures

comprehensibly. Usability aspects are closely related to staff intensity: only if the *fearless* system functions reliably it may relieve rescue staff. Only if it is easy to operate it will not require intensive training. What are appropriate measures for usability evaluation? A log-file template has been created to document any technical problem that occurs during pilot implementation. It is filled in by the professional care taker in charge and forwarded to the project partners responsible for technical support and outcome evaluation. In addition, a self-devised "Expert Checklist" was developed by psychologists from the University of Bamberg is handed out to the professional care takers after a single test installation has been completed. They are asked to rate the overall usability of the *fearless* system on a five point rating scale (Item E09: "Compared with the tele-care devices we have been using so far usability and handling of the *fearless* system are very good" ; 1 = "I strongly disagree" to 5 = "I strongly agree"). Apart from the "Expert log-file" and the "Expert Checklist" the usability of the telematics platform is also evaluated in a separate usability study, which comprises a participatory observation and a subsequent interview. The participatory observation takes place in a laboratory at the University of Bamberg. Participants are asked to navigate the telematics platform and to respond to a series of test alarms that are generated by the conductor of the study. Afterwards the participants are interviewed about their user experience and point out areas of improvement.

Reliability. A crucial aspect of usability of the *fearless* system is its reliability in terms of fall detection. To measure the reliability of the system, the number of false alarms / false positives (FP) is evaluated. This number can also be compared to similar alarm systems (e.g., panic button) since a button push without the actual need for help can be considered as false positive. How do we assess reliability of fall detection? The applied fall detection approach is introduced and evaluated under laboratory conditions by Planinc and Kampel [4], [13]. Falls are simulated according to the scenarios defined by Noury et al. [14] and by adding additional scenarios defined by Planinc and Kampel [13]. These scenarios are simulated by two subjects, simulating each scenario twice. This results in an overall set of 72 videos, containing 40 falls and 32 no-falls. The approach used in the *fearless* system yielded an accuracy of 98.6 % on 72 videos, resulting in one FP in the whole dataset. This FP occurs due to a tracking error after a fall, since the person is not tracked correctly while getting up again. Hence, a second

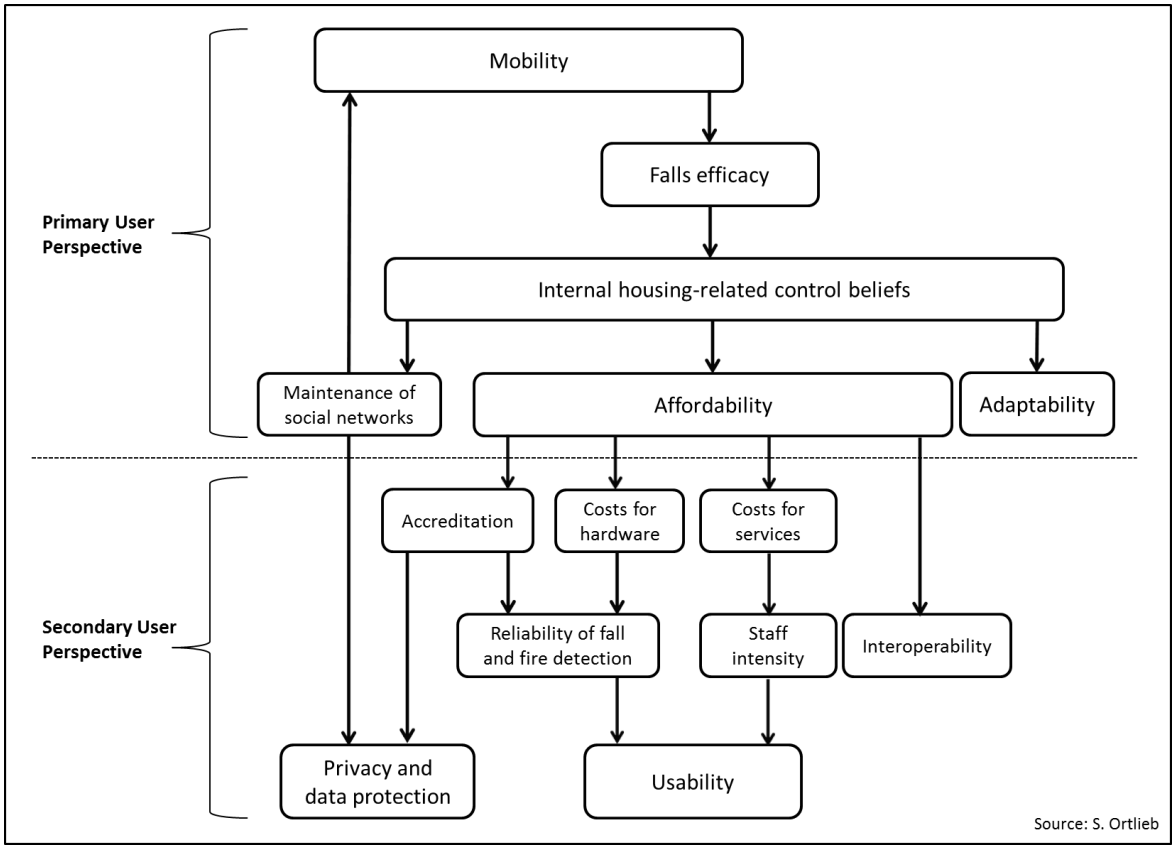


Figure 8. Technological Impact Assessment Model (TIAMo) describing the interdependencies of user requirements

fall is detected within the same sequence but as this fall does not occur in the time interval specified in the ground truth annotation, it is marked as a FP. Furthermore, the use of the Kinect™ offers practical advantages: it is robust to changing lighting conditions, also works also during the night and the installation in real homes is simplified by using only one sensor without the need for a complex calibration.

Staff intensity. Especially for care taker organizations from Germany it is difficult to recruit qualified care personnel. Also for the sake of labor costs operating the *fearless* system should be less labor intensive than a conventional tele-care systems (e.g., panic button). How do we assess staff intensity? The "Expert Checklist" contains three items targeting staff intensity ("Compared with the tele-care devices we have been using so far the *fearless* system is very efficient in terms of staff intensity."), maintenance ("Compared with the tele-care devices we have been using so far the *fearless* system needs very little maintenance.") and training effort ("Compared with the tele-care devices we have been using so far the *fearless* system can be operated without intensive training.") related to the *fearless* system. Each of these items features a five point rating scale ranging from 1 = "I strongly disagree" to 5 = "I strongly agree".

Accreditation means that the *fearless* system is officially certified and listed as an assistive device. As accreditation is a major precondition for reimbursement it contributes to the affordability of the projected e-health system. What are appropriate criteria for accreditation? National as well as European criteria for certification are applied to the *fearless* system.

Affordability / Costs for hardware and services. The projected e-health system and the services related to it should be affordable even for elderly people with limited financial resources. What are appropriate criteria for affordability? According to our survey data the hardware should be less than 200 Euros and monthly expenses for services related to it should not exceed 50 Euros.

Interoperability means that hardware and software of the e-health system should be compatible with the given IT-environment of care taker organizations. An e-health system that is interoperable with an existing infrastructure contributes to affordability as it does not require subsequent investments. How do we assess interoperability? Interoperability of the *fearless* system is operationalized by one item of the "Expert Checklist": "In general the *fearless* system is compatible with existing hardware and software and provides all necessary interfaces". Employees of care taker organizations who have worked with the *fearless* system

during pilot implementation phase are asked to rate their approval of this statement on a five point rating scale ranging from 1 = "I strongly disagree" to 5 = "I strongly agree".

B. Measures and Criteria for the Primary User Perspective

The TIAMo characterizes the primary user perspective by a set of five requirements (Figure 8): (9) housing-related control beliefs, (10) fear of falling/falls efficacy (older persons confidence in performing everyday tasks without falling) (11) mobility and (12) maintenance of social networks, illustrated in Figure 10. In addition, overall housing satisfaction is assessed. The requirement "affordability" is covered by "costs for hardware" and "costs for services" in the secondary user perspective.

Housing-related control beliefs. People who state a strong internal housing-related control belief are convinced that they are in control of their private home environment [15]. By contrast people with external control beliefs feel that their life is either controlled by powerful others (e.g., "I rely to a great extent upon the advice of others when it comes to helpful improvements to my apartment / house" [15]) or by mere chance (e.g., "Having a nice place is all luck. You cannot influence it; you just have to accept it" [15]). Older people displaying a high level of externality are prone to feelings of helplessness and depression. Therefore, reducing external control beliefs and/or sustaining an internal housing-related control belief must be considered as prerequisites for fear resolution and psychological well-being. How do we assess housing-related control beliefs? In order to capture housing-related control beliefs 16 items from the Housing-related Control Beliefs Questionnaire (HCQ) by Oswald et al. [15]) were included into our feedback questionnaire. The HCQ is a 24-item questionnaire "based on the widely used psychological dimensions of Internal Control (8 items, sum-score), External Control: Powerful Others (8 items, sum-score), and External Control: Chance (8 items, sum-score)" [16, p. 192]. Participants are instructed to rate their approval of certain statements (e.g., "Having a nice place is all luck. You cannot influence it; you just have to accept it" [15]) on a five-point rating scale ranging from 1 = "not at all" to 5 = "very much". Due to the poor psychometrical quality of the internal control scale only the two external control subscales (8 items + 8 items) of the HCQ were included into our feedback questionnaire.

Fear of falling refers to the worries of an older person in terms of falling in his/her home. Fear of falling has serious consequences for older people [17]. Elderly who are very worried about falling tend to limit their physical activities in order to reduce their risk of falling. Since fear resolution is the main objective of the *fearless* system this variable will be examined closely during pilot implementation. For the measurement of fear of falling we prefer the construct falls efficacy. Falls efficacy is defined as an "older persons confidence in performing a series of everyday tasks without falling" [18, p. 299]. How do we assess fear of falling/falls efficacy? According to Tinetti et al. [18] the Falls Efficacy Scale (FES) can be considered as a valid estimate for fear of falling. Thus, our feedback questionnaire includes the complete Falls Efficacy Scale (FES) by Tinetti et al. [18]. The FES consists of 16 Items describing activities that may be challenging for elderly (e.g., "Reaching for something above your head or on the ground"). For each activity participants are asked to rate their concerns in terms of falling in the course of this particular activities on a five point rating scale ranging from 1 = "not at all concerned" to 5 = "very concerned".

Mobility refers to the test users level of physical activity in terms of locomotion. By enhancing falls efficacy we expect the fearless system to disinhibit mobility among our test users. Ideally, this effect is not limited to the range of the ambient event detector. Thus, for mobility assessment we have combined indicators for indoor as well as outdoor activities (e.g. grocery shopping). Enhancing mobility is an important objective of the *fearless* system since it is a prerequisite for social participation.

Social participation refers to the frequency of test users social activities and the number of different people involved in these activities. Social bonds are a powerful source of self-confidence. Installing an automated event detector in the private home of an elderly person may affect social activities. The impact of the *fearless* system on social activities has to be examined throughout the pilot phase.

How do we assess mobility and social participation? The Nordic mobility-related participation outcome evaluation of assistive device intervention (NOMO) by Brandt et al. [19] allows for a combined assessment of mobility and social participation: 21 items of the NOMO capturing the "frequency of mobility-related participation and ease/difficulty in mobility during participation" [19, p. 18] have been included into the feedback questionnaire. Each item describes a certain activity of daily living, which requires physical activity (e.g., washing clothes or garments). The participants are instructed to estimate how frequently they conduct this activity (e.g., "How often do you do grocery shopping?"). For our feedback questionnaire these questions have been slightly modified: Participants are asked to estimate how frequently they have conducted certain activities in the course of the last month (e.g., "In the last month how often did you do grocery shopping?"). Moreover, the multiple choice format of the NOMO has been replaced by a cloze ("About *x* times."). The feedback questionnaire reflects the test users personal view in terms of mobility. For outcome assessment this self-report data is very valuable, yet it may be biased (e.g., by selective memory processes). Thus, every test user will be equipped with a portable step counter device in order to collect more objective data on his/her level of physical activity.

Housing satisfaction describes the overall "satisfaction

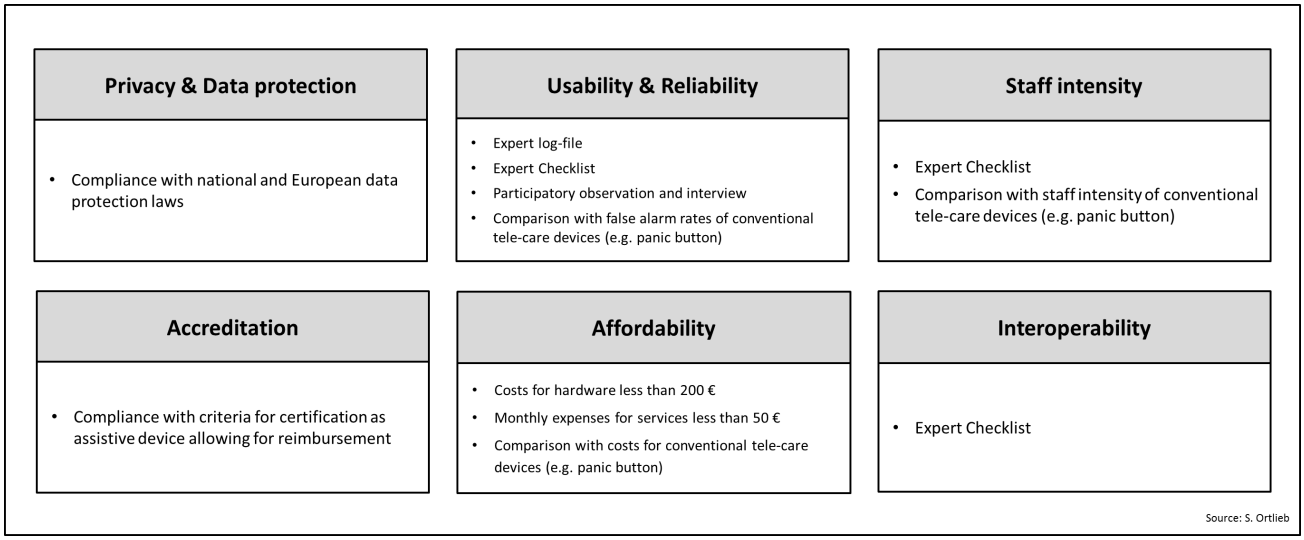


Figure 9. Overview of measures and indicators based on data from the secondary user perspective

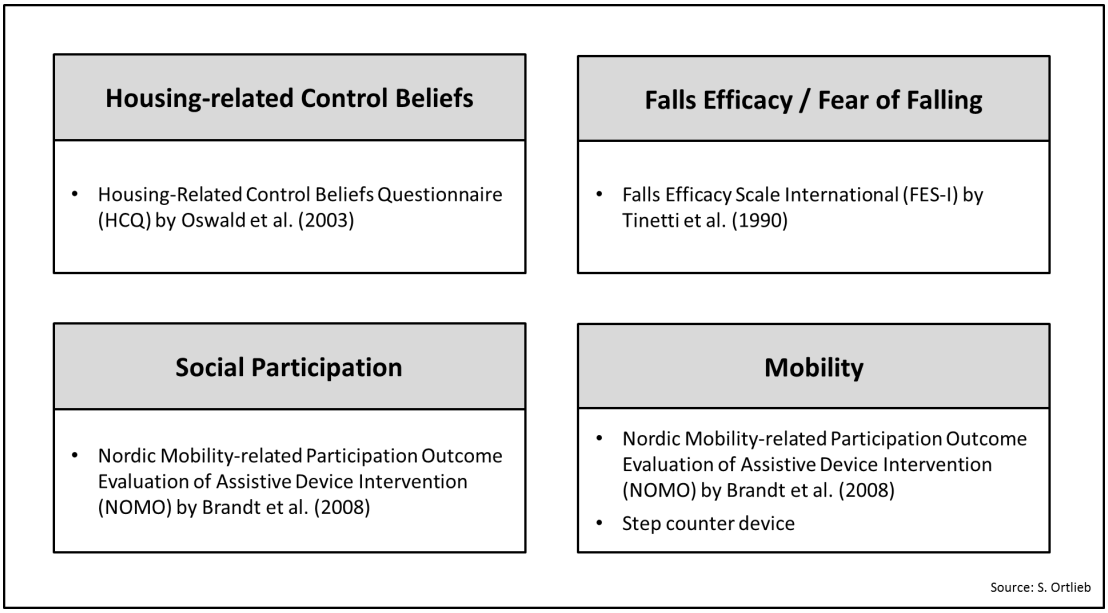


Figure 10. Overview of measures and indicators based on data from the primary user perspective

with the condition of the house“ [16, p. 192]. Introducing a novel assistive technology such as the *fearless* system to the homes of our test users is likely to affect their housing satisfaction. In our feedback questionnaire housing satisfaction is captured by a single item measure (“Are you happy with the condition in your home?”). This item was taken from the more extensive Housing Options for Older People (HOOP) by Heywood et al. [20]. It is answered on a five point rating scale ranging from 1 = “definitely not” to 5 = “yes, definitely”.

VIII. Challenges in the User-Centered Design Process

Which aspects have to be balanced in the user-centered design process? At first glance, the requirements described in the TIAMo seem perfectly compatible. Yet a closer look reveals at least four goal conflicts that have to be addressed and balanced in the process of system development and evaluation: (1) affordability versus technical performance, (2) need for control versus automation, (3) security versus privacy, (4) social needs versus personnel intensity. These four trade-offs are described and design recommendations are discussed in the following sections.

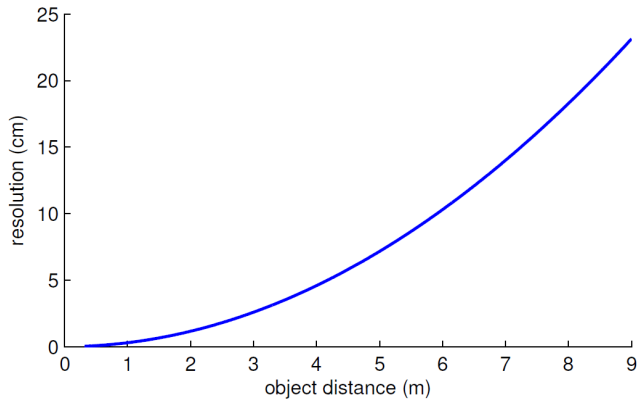


Figure 11. Resolution of the Kinect™ sensor depending on the distance [21]

A. Affordability versus Technical Performance

Research in the field of depth-sensor technology is very dynamic. Depth-sensors are relentlessly being improved in terms of accuracy and range. Although cutting-edge technology excels todays off-the-shelf sensors it is still comparatively expensive. Thus, a trade-off between technical performance and affordability has to be found. Since the Kinect™ sensor was introduced by Microsoft in 2010, a cheap depth sensor (in comparison to other depth sensors) is available. An evaluation of the depth sensor by Pramerdorfer [21] showed the feasibility of the Kinect™ sensor for fall detection. The precision of the sensor is analyzed depending on different distances, lighting conditions and surfaces. The accuracy of the sensor is reduced at higher distances but results in resolution being high enough to detect falls properly. Figure 11 shows the dependency of the accuracy (here resolution per pixel) and the distance: at a distance of one meter, the resolution is less than one centimeter. Whereas at a distance of nine meters, the resolution is above 20 cm per pixel, resulting in a reduced accuracy. However, in comparison to more accurate but more expensive devices (e.g., Pmd CamCube, Fotonic P70, Argos 3D) the accuracy of the Kinect™ is sufficient to detect falls reliably.

The Kinect™ uses structured light to obtain depth information from a scene. This results in the drawback, that the Kinect™ can only be used indoors since direct sunlight interferes with the structured light and thus no depth information can be obtained. This trade-off was made in the *fearless* project due to the goal of providing an affordable system. However, the new version of the Kinect™ presented in 2013 already uses time of flight for obtaining depth information and thus will be more stable to different lightening conditions (especially including direct sunlight).

B. Need for Control versus Automation

In 2008, guideless underground trains were introduced to the public transportation system of the city of Nuremberg, Germany [22]. Many senior citizens opposed to these plans. They felt uncomfortable being at the mercy of an autonomously operating transportation system and signalized that in the future they would choose other means of transportation where conductors of “flesh and blood” are available in case of an emergency. This example illustrates a second challenge we are facing in the user-centered development process of an automatically operating event detector: primary users need for control versus the benefits of automated fall detection. Internal control beliefs are defined as a strong confidence in ones abilities to control ones own life. On the contrary, people displaying external control beliefs feel that their life is controlled either by environmental factors (e.g., powerful others) or by mere chance. A strong notion of internal control is essential for psychological well-being in general and fear resolution in particular [23]. To a certain extent control beliefs vary across different life domains. Housing-related control beliefs refer to a persons private home: elderly people stating strong internal housing-related control beliefs are convinced that they can exert control over their private home environment. By contrast, people with external control beliefs feel that their life is either controlled by powerful others or by mere chance. Older people displaying a high level of externality are prone to feelings of helplessness and depression [24].

How does this relate to the *fearless* system? From a technical point of view reliability and accuracy of fall detection and inactivity monitoring can be improved by automation. A conventional panic button for example is useless if the client passes out or forgets to wear it. Yet from a psychological point of view, automation always implies giving away control to a technical device. In case of a panic button an alarm is actively released by the client himself/herself. Thus, introducing an autonomously operating ambient event detector to the private home of an elderly person must not undermine his/her internal housing-related control beliefs. How can we compensate this loss of control due to automation? First of all, we recommend a user interface, which provides immediate feedback and allows for active control of the ambient detector. Immediate feedback means that there is a user interface, which gives answers to the following questions: What is the current status of the system? Is it working properly? Does it require maintenance? Has an alarm been released? Active control means that the system can be switched on and off, that alarms can also be released manually and that false positives can be cancelled by the user.

From a technical point of view, these requirements can be integrated in the system easily: LEDs can symbolize the current status of the system (e.g., light blue = system active, red = alarm triggered, no light = system off) and an easy accessible on/off switch maintains the control beliefs of elderly since they are able to switch the system off at any time. Furthermore, the system can be combined

with the already established contact devices already used in combination with the panic button: in case of a triggered alarm, voice communication to a call center, care taker organization or relative can be established and thus the *fearless* system is an additional device enhancing the safety of elderly in their homes.

C. Security versus Privacy

In case of an alarm the *fearless* system generates a visualization of the scene at the clients home. This visualization is sent to the care taker organization for alarm verification. Based on these visualizations, employees of the care taker organization have to discriminate between, e.g., a fall and a false alarm. The quality of these visualizations gives rise to a third goal conflict: fast and correct alarm verification requires very detailed images, whereas for the sake of privacy protection a low level of resolution is desirable (see [25], [26]).

How can we solve this goal conflict? For instance, alarm verification via depth-images could be replaced by direct voice contact in case of an alarm. Many care taker organizations already use direct voice contact in combination with a panic button device: in case of an alarm the client receives a verification call from the call-center of the care taker organization. If the client answers this call, his/her need for assistance can be specified. If not, further measures are taken (e.g., an ambulance is sent to the clients home). If this well-tried routine was combined with the *fearless* sensors the dilemma between security and privacy could be resolved. Allowing for direct voice contact between clients and care taker staff brings us to the next crucial trade-off between the clients need for communication and personnel intensity on the service providers side. However, this solution only applies to the final system. During the technical development within the *fearless* project and the pilot phases A and B, another trade-off needs to be found. The verification image is needed not only for care taker staff to decide whether a fall has occurred or not, but also for technical staff developing the system and thus not only information if a false alarm has occurred but also why this false alarm has occurred need to be gathered. Hence, different visualizations are developed as shown in Figure 12. Most information is included in RGB images, allowing to verify if a fall occurred and identifying the problem of false alarms easily. However, due to the lack of privacy protection, no RGB images are transmitted at any time. Depth images are used during the pilot phase in order to verify if a fall occurred and the system is working properly. These images were found to be a trade-off during the pilot phases were the system is still under development but the privacy aspects need to be considered. However, if the technical system is working properly, more abstract visualizations (e.g., top-view and 3 dots representing the upper body with respect to the ground floor) can be applied in order to allow the verification of false alarms by care takers. Evaluation showed that the visualization containing only the ground floor, three dots and a line is even more helpful than a top view image showing the shape of a person lying on the floor since it can be interpreted more easily. Nevertheless, a direct voice communication is seen as the best way to verify if a fall has occurred while respecting the privacy of elderly.

D. Social Needs versus Personnel Intensity

In the field of AAL a strong emphasis is usually put on self-determination and independence of elderly people. Yet social relatedness and the feeling of belonging to a valued group (e.g., family, neighborhood, religious community, etc.) are equally important for our psychological well-being as humans. Moreover, our needs for autonomy and affiliation are dynamically interrelated and cannot be treated as separate entities [27]. For instance, the ability to interact and communicate with members of a valued group adds to an elderly persons notion of competence. As an elderly persons level of functioning is gradually decreasing, social bonds play an increasingly important role in stabilizing control beliefs. Thus, the relevance of affiliation for the resolution of elderlies fears has to be taken into account.

In our primary user requirement survey we have found that social isolation is a problem particularly for solitarily living elderly from Austria and Germany [12]. In addition, our secondary user requirement survey shows that many false positive alarms that are raised by client's wearing a panic button actually reflect their need for affiliation: they press the button because they are longing for someone to speak to. As a consequence some care taker organizations do not consider these events as false positives but as a different kind of alarm, which requires as much attention as for example a severe fall. For these organizations providing social support is part of their mission. On the other hand, dealing with these "false positives" is staff intensive. In our secondary user requirement analysis the issue of staff intensity was often raised by representatives of care taker organizations from Germany. Since mandatory civil service has been abolished in Germany trained care personnel is scarce and labor costs are increasing.

How does this relate to the *fearless* system? In case of a fall the *fearless* system releases an alarm automatically. Verification images of the scene can be used for alarm verification by members of the care taker organization. Thus, no social interaction (e.g., via direct voice contact) takes place between clients and care taker personnel. For the staff of care taker organizations this may be beneficial in terms of staff intensity. For primary users this procedure is likely to yield acceptance issues: as in the case of the guideless underground trains elderly users might prefer a care taker of "flesh and blood" to an efficient yet anonymous e-health system. This results in the fact, that the *fearless* system may server as additional system for fall detection, when

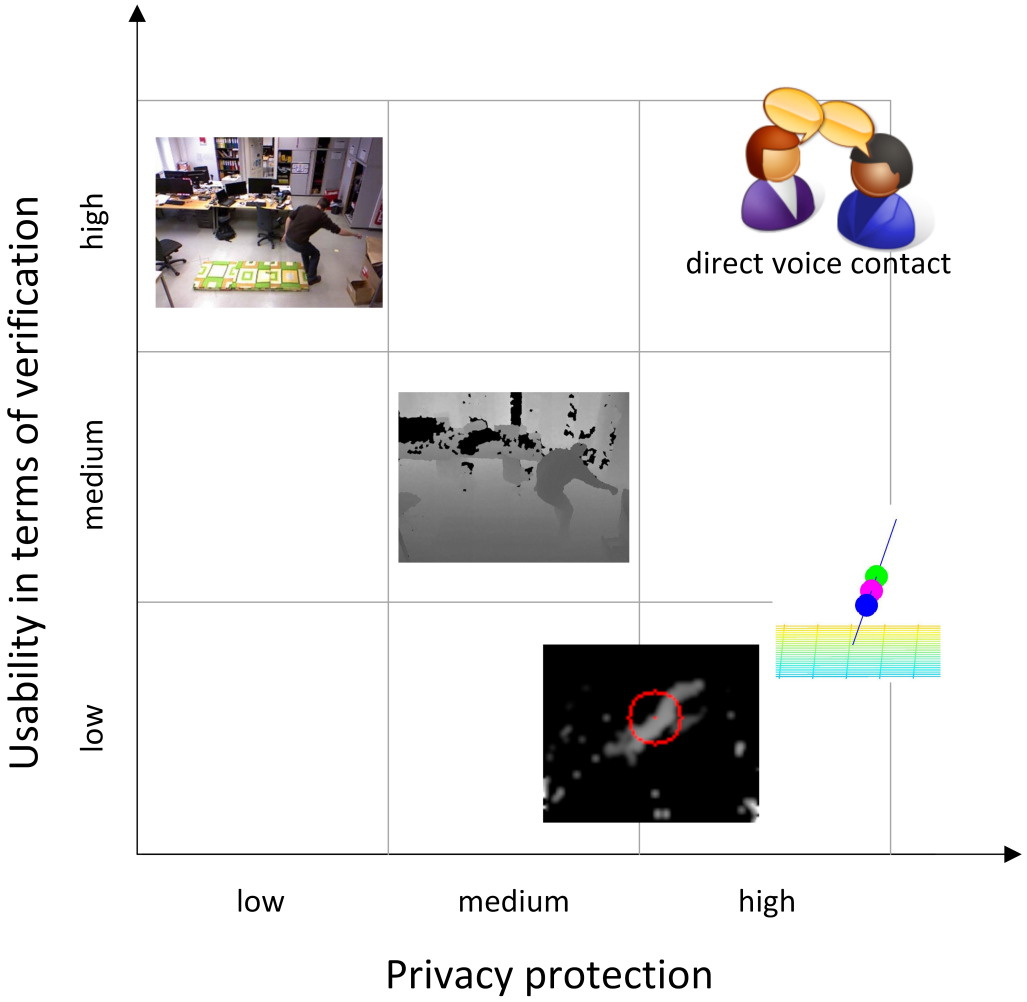


Figure 12. Usability versus privacy protection

a client is not able to push the button but not replacing an already established panic button system. However, this increases staff intensity on the care taker's side. Thus, yet another trade-off has to be found.

IX. Conclusion

In this paper, we have outlined a performance measurement system that has been tailored to the specific challenges of the AAL-JP project *fearless*. Based on the Technological Impact Assessment Model (TIAMo) its measures reflect two different user perspectives: older adults as primary users and care taker organizations as secondary users. Moreover, it describes the interplay of technical, psychological and business aspects related to the projected e-health system. By creating a common understanding of different user perspectives among our multi-disciplinary project partners this balanced scorecard will guide our actions in the next stages of the user-centered design process. What are the next steps to be taken? A pilot study is planned during which the *fearless* system will be installed in the private homes of 45 test users from Austria, Germany, Italy, and Spain. This pilot study lasts four months including pre-test and follow-up test, during which test users and care taker personnel will be contacted on a regular basis to assess the measures described in the *fearless* scorecard.

Acknowledgment

The authors want to thank the whole *fearless* project team, since techniques and ideas of many researchers involved are used.

References

[1] R. Planinc and M. Kampel, "E-Health System Development based on End User Centered Design," in *Proceedings of The Fifth International Conference on eHealth, Telemedicine, and Social Medicine (eTELEMED 2013)*. Nice, France: IARIA XPS Press, 2013, pp. 83–86.

[2] D. Anderson, J. M. Keller, M. Skubic, X. Chen, and Z. He, "Recognizing falls from silhouettes." in *28th Annual International Conference of the IEEE on Engineering in Medicine and Biology Society, 2006*, vol. 1, New York, Jan. 2006, pp. 6388–6391.

[3] G. Mastorakis and D. Makris, "Fall detection system using Kinects infrared sensor," *Journal of Real-Time Image Processing*, Mar. 2012.

[4] R. Planinc and M. Kampel, "Introducing the use of depth data for fall detection," *Personal and Ubiquitous Computing*, vol. 17, no. 6, pp. 1063–1072, May 2012.

[5] C. Rougier, E. Auvinet, J. Rousseau, M. Mignotte, and J. Meunier, "Fall Detection from Depth Map Video Sequences," in *Toward Useful Services for Elderly and People with Disabilities*, ser. Lecture Notes in Computer Science, B. Abdulrazak, S. Giroux, B. Bouchard, H. Pigot, and M. Mokhtari, Eds. Montreal: Springer Berlin / Heidelberg, 2011, vol. 6719, pp. 121–128.

[6] A. Zweng, S. Zambanini, and M. Kampel, "Introducing a Statistical Behavior Model into Camera-Based Fall Detection," in *6th International Symposium on Visual Computing (ISVC'10)*, ser. Lecture Notes in Computer Science, G. Bebis, R. Boyle, B. Parvin, D. Koracin, R. Chung, R. Hammoud, M. Hussain, T. Kar-Han, R. Crawfis, D. Thalmann, D. Kao, and L. Avila, Eds., vol. 6453. Las Vegas: Springer Berlin / Heidelberg, 2010, pp. 163–172.

[7] R. Lubinski, *Dementia and Communication*. B.C. Decker, Inc.: Philadelphia, 1991.

[8] A. Mihailidis, B. Carmichael, and J. Boger, "The Use of Computer Vision in an Intelligent Environment to Support Aging-in-Place, Safety, and Independence in the Home," *Gerontechnology*, vol. 2, no. 2, pp. 173–189, Sep. 2002.

[9] (2013, Nov.) Openni. [Online]. Available: www.openni.org

[10] R. S. Kaplan and D. P. Norton, "The balanced scorecard: measures that drive performance," *Harvard Business Review*, vol. 1, pp. 71–79, 1992.

[11] A. Müller, *Strategisches Management mit der balanced scorecard*. Kohlhammer: Stuttgart, 2000.

[12] S. Ortlieb, G. Streffing, and C. C. Carbon, "There is Nothing More Practical than a Good Theory: Designing an Ambient Event Detector based on a Technological Impact Assessment Model (TIAMo)," in *AAATE*. IOS Press, 2013, to appear.

[13] R. Planinc and M. Kampel, "Robust Fall Detection by Combining 3D Data and Fuzzy Logic," in *ACCV Workshop on Color Depth Fusion in Computer Vision*, J.-I. Park and J. Kim, Eds. Daejeon, Korea: Springer, 2012, pp. 121–132.

[14] N. Noury, P. Rumeau, A. K. Bourke, G. OLaighin, and J. E. Lundy, "A proposal for the classification and evaluation of fall detectors," *Biomedical Engineering and Research IRBM*, vol. 29, no. 6, pp. 340–349, 2008.

[15] F. Oswald, H.-W. Wahl, M. Martin, and H. Mollenkopf, "Toward Measuring Proactivity in Person-Environment Transactions in Late Adulthood," *Journal of Housing For the Elderly*, vol. 17, no. 1-2, pp. 135–152, 2003.

[16] F. Oswald, O. Schilling, H.-W. Wahl, A. Fänge, J. Sixsmith, and S. Iwarsson, "Homeward bound: Introducing a four-domain model of perceived housing in very old age," *Journal of Environmental Psychology*, vol. 26, no. 3, pp. 187–201, 2006.

[17] R. G. Cumming, G. Salkeld, M. Thomas, and G. Szonyi, "Prospective Study of the Impact of Fear of Falling on Activities of Daily Living, SF-36 Scores, and Nursing Home Admission," *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, vol. 55, no. 5, pp. M299–M305, 2000.

[18] M. E. Tinetti, D. Richman, and L. Powell, "Falls Efficacy as a Measure of Fear of Falling," *Journal of Gerontology*, vol. 45, no. 6, pp. pp. 239–243, 1990.

[19] A. Brandt, C. Löfqvist, I. Jónsdottir, T. Sund, A.-L. Salminen, M. Werngren-Elgström, and S. Iwarsson, "Towards an instrument targeting mobility-related participation: Nordic cross-national reliability." *Journal of Rehabilitation Medicine*, vol. 40, no. 9, pp. 766–772, Oct. 2008.

[20] F. Heywood, C. Oldman, and R. Means, "Housing and Home in Later Life," *Housing Studies*, vol. 18, no. 6, Nov. 2002.

[21] C. Pramerdorfer, "Evaluation of Kinect Sensors for Fall Detection," in *Computer Graphics and Imaging / 798: Signal Processing, Pattern Recognition and Applications*. Innsbruck, Austria: ACTAPRESS, 2013.

[22] Stadtanzeiger Nürnberg, "Senioren wollen die fahrerlose U-Bahn meiden," Nürnberg, April 23, 2008.

[23] A. Bandura, "Self-efficacy: The exercise of control," 1997.

[24] P. K. Presson and V. A. Benassi, "Locus of control orientation and depressive symptomatology: A meta-analysis." *Journal of Social Behavior & Personality*, vol. 11, no. 1, pp. 201–212, 1996.

[25] A. J. Bharucha, A. J. London, D. Barnard, H. Wactlar, M. A. Dew, and C. F. Reynolds, "Ethical Considerations in the Conduct of Electronic Surveillance Research," *The Journal of Law, Medicine and Ethics*, vol. 34, no. 3, pp. 611–619, Sep. 2006.

[26] H. Sohn, W. De Neve, and Y. M. Ro, "Privacy Protection in Video Surveillance Systems: Analysis of Subband-Adaptive Scrambling in JPEG XR," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, no. 2, pp. 170–177, Feb. 2011.

[27] N. Bischof, "Das Rätsel Ödipus. Die biologischen Wurzeln des Urkonflikts von Intimität und Autonomie," 2001.

A New Device for Sleep Apnea Treatment Telemonitoring: a Bench Study

Valentina Isetta
University of Barcelona-CIBERES,
Faculty of Medicine
Unit of Biophysics and Bioengineering
Barcelona, Spain
valentina.isetta@ub.edu

Josep M. Montserrat
Sleep Lab, Pneumology Department,
Hospital Clinic-IDIBAPS-CIBERES,
Barcelona, Spain
jmmontserrat@clinic.ub.es

Geraldine Thiebaut
Air Liquide,
Paris, France
geraldine.thiebaut@airliquide.com

Claude Weber
Air Liquide,
Paris, France
claude.weber@airliquide.com

Daniel Navajas
University of Barcelona-IBEC-CIBERES,
Faculty of Medicine,
Unit of Biophysics and Bioengineering
Barcelona, Spain
dnavajas@ub.edu

Ramon Farré
University of Barcelona-IDIBAPS-CIBERES,
Faculty of Medicine,
Unit of Biophysics and Bioengineering
Barcelona, Spain
rfarre@ub.edu

***Abstract*—Patient's compliance is crucial for the effectiveness of continuous positive airway pressure (CPAP) treatment of obstructive sleep apnea (OSA). Unfortunately, up to 50% of patients withdraw CPAP because of treatment side effects. Monitoring a patient's CPAP compliance at home would be useful to early detect underuse and to properly address problems. Air Liquide developed NOWAPI, a novel telemedicine system, which provides a remote monitoring of the CPAP treatment and is designed to be compatible with all CPAP devices under clinical use. The aim of this study was to validate this novel telemonitoring system in a bench test. NOWAPI detects important CPAP treatment efficiency parameters, such as the usage time and residual events, and sends them to a secure server, from which can be downloaded for analysis by the healthcare staff. In this study, NOWAPI was tested when using CPAP devices applied to a model simulating OSA patients. In a first test phase, to assess the influence of NOWAPI sensor unit geometry to the CPAP treatment, the responses of 2 different CPAP machines to a series of 10 disturbed breathing patterns with NOWAPI connected or not to the setting was compared. Then, the telemedicine system performance was tested in 30 simulated patients' sleep periods of CPAP treatment, lasting 4 hours each. They consisted of disturbed breathing patterns built from selected events of real OSA patients' polysomnographic recordings. The recorded data of each test were telemetrically sent to a server by the NOWAPI GPRS module, then downloaded and analyzed. The simulated patients were treated with 3 different currently available CPAP devices. NOWAPI sensor unit connection to the setting did not influence the CPAP treatment in the two CPAP devices analyzed. The difference between the treatment duration estimated by the device and actual values was never higher than 3 minutes over the 4-hour test. The absolute difference between the apnea-hypopnea index estimated by NOWAPI and the actual values, 0.9±1.6 events/hour (mean±SD), was not significantly different from the absolute difference between the AHI estimated by the CPAP machines and the actual values, 0.9±1.0 events/hour (p=0.171). NOWAPI showed an excellent performance in estimating the duration of the CPAP treatment and in detecting residual respiratory events in simulated OSAS patients. NOWAPI system could be a valuable tool for telemonitoring the treatment of obstructive sleep apnea.**

Keywords-component; telemedicine; eHealth; home monitoring; sleep apnea; CPAP.

I. INTRODUCTION

The original version of this paper has been presented at eTELEMED 2013 conference, in Nice, France [1]. This extended version includes more detailed data across all paper sections.

Obstructive sleep apnea (OSA) is a very prevalent disease mainly associated with daytime sleepiness and deterioration of quality of life and is suffered by 2% to 4% of middle-aged adults [2]. OSA entails repetitive partial or total occlusion of the upper airway, which results in significant levels of sleep disturbance and snoring. However, the seriousness of untreated OSA is stressed by its significant consequences, including depression, ischemic heart disease, stroke, hypertension and significantly increased risk of motor vehicle crashes [3][4]. In addition, OSA is strongly related to obesity even though it is also increasingly identified in non-

obese subjects with a particular craniofacial structure. The incidence of OSAS is likely to grow in parallel with the spread of obesity now occurring in many countries.

The treatment of choice for OSA is continuous positive airway pressure (CPAP) applied through a nasal mask during sleep. This constant pressure is transmitted to the pharyngeal area, thereby avoiding upper airway obstruction [5]. Despite the documented clinical efficacy of CPAP, up to 50% of patients suspend or underuse CPAP treatment, mainly due to its discomforting side effects, such as pressure intolerance, claustrophobic reaction to the mask, mask displacement, and machine noise [6][7]. Many of these problems could be easily solved by a closer follow-up, especially during the first weeks, but busy sleep centers have difficulties in giving such support.

If patients do not use CPAP for the recommended minimum of 4 hours per night, clinical outcomes are compromised [8], demonstrating that adherence optimization is a critical aspect of patient management.

Several studies confirmed that treatment compliance could be significantly improved by comprehensive support programmes and timely interventions by health professionals [9]. In recent times, it has been recognized that telemedicine could have a valuable role in improving CPAP therapy adherence [10]. In fact, telemedicine has been used in various studies to promote and reinforce CPAP treatment. In most of them a cognitive behavioural intervention was applied to OSA patients at home, by telephone, the Internet and videoconference. Namely, a randomized clinical trial showed that the use of a telephone-linked communication system, which provided feedback and counselling to OSAS patients at home, improved CPAP adherence, patients' functional status and reduced depressive symptoms [7]. Another study employed an Internet-based informational support service for problems due to CPAP use [11]. Despite the organizational limitations and poor differences between intervention and control group follow-up, they obtained good patients' acceptance of this monitoring strategy. It is also noteworthy that telehealth interventions, such as long-distance visit via videoconference (or "televisits"), have been found to improve CPAP adherence in a small group of nonadherent patients versus a placebo-controlled group [12]. The cost of the interventions, including the telehealth monitor, home installation and telephone charges, was lower than the same number of face-to-face visits. However, larger studies are needed to generalize any conclusion.

Although these previous studies achieved mixed results in terms of significant improvement of CPAP compliance, the potential of telemedicine as part of an integrated care for OSA patients was confirmed.

Two recent randomized studies [13][14] combined elements of psycho-educational interventions together with technological innovation. Usual care was compared to a wireless telemonitoring of CPAP compliance and efficacy data, which physicians were able to daily monitor through a secure web browser and thus contact the patient if needed. Both studies resulted in higher CPAP adherence and improved OSA outcomes and demonstrated that continuous monitoring of patient's compliance could be useful to early detect underuse and to properly address possible problems.

Some existing CPAP and APAP (Automatic Positive Airway Pressure) devices monitor patient's compliance by using different algorithms, but only few of them offer continuous remote monitoring. Air Liquide developed NOWAPI, a telemonitoring system designed to be compatible with all commercially available CPAP/APAP devices.

The hypothesis of this study was that this new telemedicine system for CPAP therapy remote monitoring could provide valuable and useful data about treatment compliance and efficacy for the follow-up OSAS patients.

The specific aims of this study were: a) to assess whether the connection of NOWAPI sensor unit to different CPAP/APAP machines influenced their normal functioning and responses to the disturbed breathing patterns generated by a simulated OSAS patient in a bench; b) to evaluate the NOWAPI's performance in accurately detecting the CPAP/APAP treatment duration and the residual disturbed-breathing events in a bench test.

II. METHODS

A. System Description

NOWAPI system has been designed to remotely monitor the CPAP or APAP treatment of patients with sleep apnea at home. The system overview is depicted in Fig. 1. NOWAPI comprises a small sensor unit (15x4x7 cm), shown in Fig. 2, powered by a rechargeable battery, which contains a pressure and flow sensing module, a specifically developed detection software for the analysis of the measured signals and detection the breathing events, a GPRS communication module, which enables data transmission to a server, and a clinical interface, which enables the physician to visualize and properly evaluate the data downloaded from the server. The NOWAPI sensors unit is connected between the

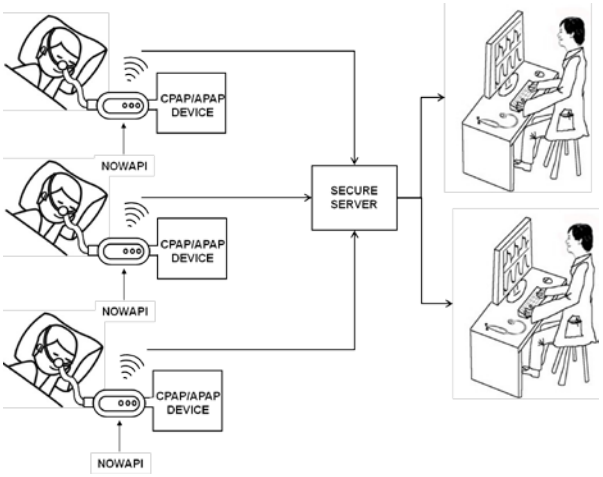


Figure 1. NOWAPI system data flow.

Figure 2. NOWAPI sensor unit.

CPAP/APAP device outlet and the patient's tubing. During the patient's CPAP/APAP treatment, the system detects the pressure and flow signals, which characterize the patient's breathing and estimates the treatment use rate and some important parameters to assess the effectiveness of the therapy, such as the number of apneas, hypopneas, flow limitations, snoring periods, and average breathing flow and nasal pressure. The system stores all data in 2 different files, a detailed file with a sampling rate of 25 Hz and a synthetic file where data are recorded as mean values over 15–minute consecutive periods. The latter file is sent by the GPRS module integrated into the device to a secure server then available to be downloaded and analyzed.

Furthermore, a led in the sensors unit turns red if the treatment duration is less than the minimum standard of 4 hours/night, giving an immediate useful feedback to the patient about his/her treatment compliance.

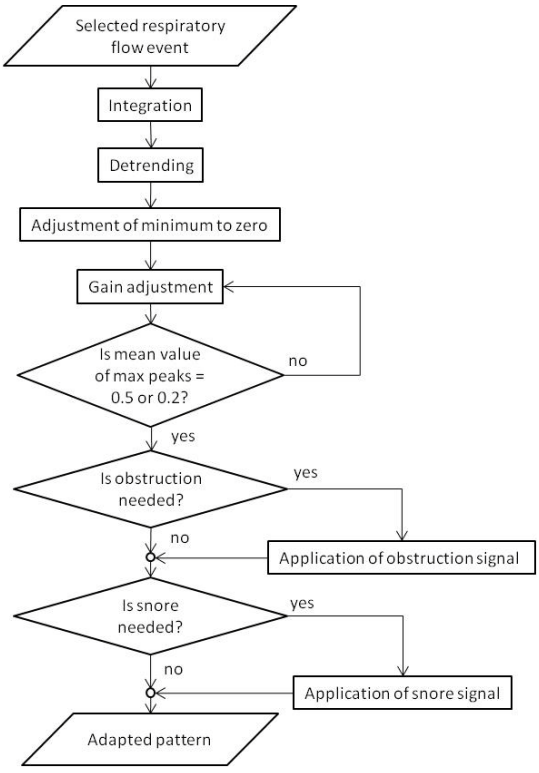


Figure 3. Block diagram of the algorithm implemented to obtain the breathing patterns simulating patients with OSA.

B. *Patterns of disturbed breathing*

NOWAPI was tested with 2 different sets of simulated breathing patterns. In the first phase, a series of 10 waveforms consisting of the successive repetition of apneic or hypopneic events or persistent flow limitation with or without snoring [15][16] was used. For each flow pattern, the waveform generator (see Section C) produced a specific pattern of airway obstruction with the proper magnitude and duration for simultaneously mimicking the obstructive events and the flow shapes observed in OSA patients. The flow and obstructive events were combined for generating 10 different breathing patterns. The description of these disturbed breathing patterns is summarized in Table I. Test 1 simulated a breathing pattern with repetitive apneas with obstruction that correspond to the total airway occlusion. Test 2 simulated a breathing pattern known as central apnea, in which patients stop breathing due to brain's respiratory control centers imbalance, but no airway obstruction is present. Tests 3 and 4 reproduced two different severity levels of partial airway obstruction, called hypopnea. Test 5 simulated a mild hypopnea with the addition of snoring vibrations synchronized with the inspirations.

TABLE I. DISTURBED BREATHING PATTERNS USED FOR THE COMPATIBILITY TEST PHASE.

Test #	Test Description
1	Apnea with Obstruction
2	Apnea without Obstruction
3	Severe Hypopnea
4	Mild Hypopnea
5	Mild Hypopnea with Snoring
6	Prolonged Flow Limitation with Obstruction
7	Prolonged Flow Limitation with Obstruction and Snoring
8	Mouth Expiration
9	Apnea with Obstruction and Leaks
10	Simulated OSA patient

Tests 6 and 7 reproduced the breathing events known as flow limitation that occurs when flow ceases to increase with increasing expiratory effort. In the case of Test 7 snoring signal was superimposed to the flow limitation pattern. Test 8 mimicked the occurrence of patient's expiration through the mouth. In Test 9 a signal mimicking the mask leaks was added to the repetitive obstructive apnea pattern. Test 10

simulated a breathing condition representative of a complete OSA patient breathing pattern that depended on the CPAP applied. Specifically, the generator reproduced apneas if CPAP was less than 5 cmH_2O, severe hypopneas when applied CPAP was between 5 and 7 cmH_2O, moderate hypopneas if CPAP was between 7 and 10 cmH_2O, prolonged flow limitation when CPAP was between 10 and 12 cmH_2O, and normal breathing for CPAP greater than 12 cmH_2O.

In the second phase, the system performance was tested in 30 different test scenarios especially developed for this study, simulating 30 sleep periods of OSA patients under CPAP treatment, lasting 4 hours each. In Table II the 30 breathing patterns generated for this study are described in

TABLE II. NUMBER OF SIMULATED EVENTS ASSIGNED FOR EACH TEST AND EXPERIMENTAL CONDITIONS.

Test number	Average AHI[a]	Number of obstructive apneas	Number of central apneas	Number of hypopneas with snoring	Number of hypopneas without snoring	Number of flow limited events with snoring	Number of flow limited events without snoring	Number of prolonged flow limitations without snoring	Number of prolonged flow limitations with snoring	CPAP machine	With/without comfort mode	With/Without humidifier
1	0	0	0	0	0	2	2			S9 Autoset	without	without
2	1	1	0	2	1	9	73			Goodknight 420E	without	with
3	1,25	1	0	4	0				1	Goodknight 420E	without	without
4	1,75	0	0	4	3			1		Goodknight 420E	without	without
5	2	0	0	2	6	35	15			S9 Autoset	without	with
6	2	0	0	3	5	20	90			S9 Autoset	without	without
7	2,25	0	1	7	1					S9 Autoset	without	without
8	2,5	1	2	7	0				1	S9 Autoset	without	without
9	2,5	3	0	6	1			1		Remstar Auto PR1	without	without
10	2,5	6	0	2	2			1		Remstar Auto PR1	without	with
11	2,75	2	0	6	3				1	Remstar Auto PR1	without	without
12	2,75	2	0	9	0	0	17			S9 Autoset	without	without
13	3,25	1	2	0	10			1		S9 Autoset	without	without
14	3,5	0	0	8	6	2	1			Remstar Auto PR1	without	without
15	4	6	1	7	2					Goodknight 420E	without	without
16	4,25	0	1	14	2					Remstar Auto PR1	without	without
17	4,25	2	3	7	5					Remstar Auto PR1	without	without
18	4,5	5	2	9	2					Goodknight 420E	without	without
19	4,75	18	1	0	0					S9 Autoset	without	without
20	4,75	4	3	9	3					S9 Autoset	without	without
21	4,75	2	0	0	17	38	17			Remstar Auto PR1	without	without
22	5	17	1	2	0	6	89			Goodknight 420E	without	without
23	6	1	1	5	17	14	1			Remstar Auto PR1	without	without
24	7	0	0	18	10					S9 Autoset	without	without
25	8,75	4	5	13	13	0	4			Goodknight 420E	without	without
26	10,8	12	6	25	0					Remstar Auto PR1	without	without
27	15	15	21	10	14	7	11			S9 Autoset	without	without
28	20,3	37	2	32	10					Goodknight 420E	without	without
29	25,3	9	8	48	36					Remstar Auto PR1	without	without
30	30	31	31	29	29	38	17			S9 Autoset	without	without

a. AHI = Apnea-hypopnea Index, which is the number of disturbed breathing events per hour.

detail.

These simulated breathings consisted of realistic airflow patterns built from a library of actual events (e.g., normal breathing, apneas, hypopneas, flow limitations) selected from real OSA patients' polysomnographic recordings. The selected events were exported by using the polygraph software with a sampling frequency of 64 Hz. Then, each event was properly elaborated by an algorithm implemented for this study. The block diagram describing the algorithm, developed by using Matlab computing tool, is shown in Fig. 3. First, the flow event was integrated to obtain the volume signal. Then, the signal was detrended and adjusted in order to have the minimum signal point at zero. Then, to reproduce the typical tidal volumes for normal breathing (0.5 l approx.) and hypopnea (0.2 l approx), the signal gain was iteratively adjusted until the mean value of the signal peaks was 0.5 in the case of normal breathing and 0.2 in the case of hypopnea. Subsequently, the obstruction signal controlling the test bench valve was added to obstructive events. Moreover, a snore signal was added where requested. Then, the processed

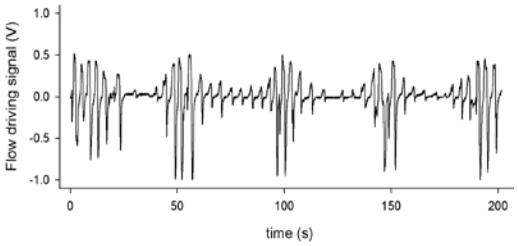


Figure 4. Fragment of a pattern of disturbed breathing which simulated an OSA patient's sleep periods of treatment.

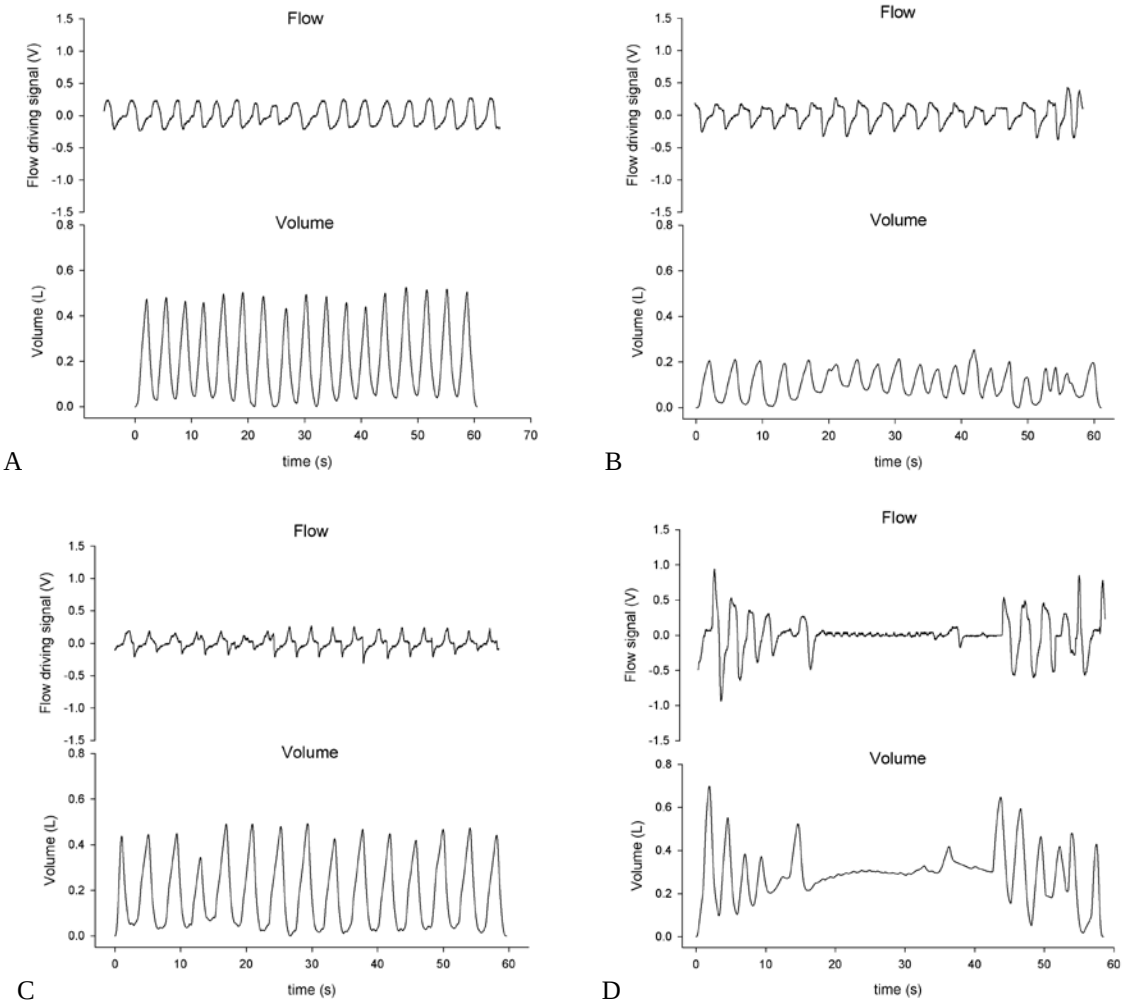


Figure 5. Breathing patterns obtained by using the algorithm developed in this study: (A) Normal breathing, (B) Hypopnea, (C) Flow limitation, (D) Apnea.

events were assembled to obtain the 30 4-hour simulated breathing patterns (Fig. 4). In Fig. 5 some representative breathing patterns obtained by using the algorithm developed in this study are depicted.

C. Measurement Setup and Protocol

NOWAPI sensors unit was plugged between the CPAP/APAP device (or its humidifier) outlet and a model simulating an OSA patient [15][16], as shown in Fig. 6. This computer-driven model comprises a flow generator and an obstruction valve which allows the simulation of obstructive events. Other two valves (the leak and the exhalation valves) allow the simulation of leaks and mouth breathing and a loudspeaker simulates snoring. The test bench is equipped with two sensors, which record pressure and flow signals. A calibrated leak (EP on Fig. 6) simulates the mask leak.

This validation study comprised two phases in which the same test setting (Fig. 6) was employed.

1) First Test Phase

The aim of the first test phase was to verify that the NOWAPI sensor unit connected between a CPAP/APAP machine and the conventional tubing connected to the patient did not modify the normal performance of the CPAP/APAP machine. Two commercially available CPAP/APAP devices (S9 AutoSet, Resmed and Remstar Auto, Respironics) were subjected to a set of 10 breathing patterns described elsewhere [15][16] with 2 alternative settings: with or without their Comfort Mode (CPR) activated and with or without NOWAPI sensors unit connected to the test setting. The responses obtained in the 4 different experimental conditions were compared and evaluated.

2) Second Test Phase

In the second test phase, NOWAPI was subjected to the 30 patterns especially implemented for this study, which simulated 30 sleep periods of OSA patients under CPAP treatment. The aim of this phase was the assessment of NOWAPI's performance in correctly detecting the treatment duration and the residual disturbed-breathing events.

In order to assess the effect of water condensation into the tubing on the measurements, usually caused by patient's breathing, three of the tests were performed with the APAP device humidifier turned on. To guarantee realistic water condensation, humidifier was set to maximum level and the APAP device tubing was immersed in ice.

The simulated patients were treated with 3 different currently available devices for APAP treatment: S9 Autoset (Resmed), Remstar Auto PR1 (Respironics) and Goodknight 420E (Sefam). Each APAP device was connected to the monitoring device with its own tubing. A Whisper Swivel valve (Respironics) was used as exhalation port for all devices.

Each 4-hour test was preceded and followed by a 30-minute period during which the NOWAPI device was functioning but not subjected to either APAP device pressure or patient simulator's breathing. This was to ensure test two epochs in which no treatment time and no events should have been detected.

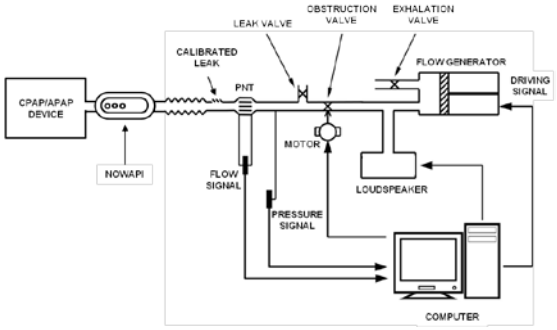


Figure 6. Scheme of the test setting.

The synthetic files for each of the 30 tests, containing the data recorded as mean values over 15–minute consecutive periods, were sent via GPRS to the Air Liquide secure server and then downloaded for analysis. In this study, treatment duration and respiratory events, measured as apnea-hypopnea index (AHI), detected by NOWAPI were considered for analysis and compared to the ones detected by the CPAP/APAP devices and to the actual simulated patterns generated by the bench.

III. RESULTS

To assess whether the connection of NOWAPI sensor unit to the CPAP/APAP machines influenced their responses to the disturbed breathing patterns generated by the OSAS patient simulated by the bench, the absolute differences between the test setting with and test setting without NOWAPI in the circuit of the following parameters were calculated: the time taken by the CPAP/APAP machine to reach the pressure of 10 cmH_2O (T_{10}) and the maximum pressure applied by the machine (P_{max}). These values were calculated for CPAP/APAP machines S9 Autoset (Resmed) and Remstar Auto (Respironics) with and without CPR. Table III summarizes the results, which corresponded to the 4 experimental conditions.

TABLE III. RESULTS OF THE FIRST TEST PHASE

CPAP/APAP machine	Absolute difference with NOWAPI/without NOWAPI (mean±SD)	
	T_{10} (min)	P_{max} (cmH_2O)
S9 Autoset with CPR[a]	0.40±0.43	0.50±0.77
S9 Autoset without CPR	0.19±0.29	0.11±0.14
Remstar Auto PR1 with CPR	1.23±0.98	0.78±1.54
Remstar Auto PR1 without CPR	1.50±1.75	1.26±1.41

a. CPR = Comfort Mode.

In order to assess the intrinsic variability of the CPAP/APAP machines' response to the same breathing pattern, we performed 3 rounds of the same 10 disturbed breathing patterns in both machines without the NOWAPI sensor unit in the circuit. The absolute differences between test rounds of T_{10} and P_{max} are comparable with the ones found previously (Table IV and V), hence they can be imputable to the intrinsic variability of the CPAP/APAP devices' response.

In the second test phase, all data sent to the server via GPRS were successfully received and analyzed. Detailed results of the CPAP treatment duration analysis are summarized in Table VI. The percentage difference between the treatment duration estimated by NOWAPI and actual values was never higher than 1.25% (3 min) and never lower than -0.42% (-1 min).

Detailed results of the residual event detection analysis are summarized in Table VII. The difference in absolute values between the AHI estimated by NOWAPI and the actual values, 0.9±1.6 events/hour (mean±SD), was not significantly different from the difference in absolute value between the AHI estimated by the CPAP/APAP machines and the actual values, 0.9±1.0 events/hour (p=0.171, the normality condition was achieved). This good agreement was confirmed by Bland-Altman analysis of AHI values estimated by NOWAPI in each test versus the actual ones (Fig. 7A). Also, AHI values estimated by NOWAPI showed a very good correlation with the actual values (R^2=0.97), slightly better than the ones estimated by PAP machines (R^2=0.88) (Fig. 7B).

TABLE IV. ANALYSIS OF CPAP/APAP MACHINES' INTRINSIC VARIABILITY ON T_{10} (MIN)

CPAP/APAP machine	Absolute difference with NOWAPI/without NOWAPI (mean±SD)		
	R1 – R2[a]	***R1 – R3***[b]	***R2 – R3***[c]
S9 Autoset	0.77±1.34	0.62±0.79	0.87±1.12
Remstar Auto	4.47±7.11	5.1±6.54	2.6±4.30

a. Round 1 – Round 2; b. Round 1 – Round 3; c. Round 2 – Round 3.

TABLE V. ANALYSIS OF CPAP/APAP MACHINES' INTRINSIC VARIABILITY ON P_{MAX} (CMH_2O)

CPAP/APAP machine	Absolute difference with NOWAPI/without NOWAPI (mean±SD)		
	R1 – R2[a]	***R1 – R3***[b]	***R2 – R3***[c]
S9 Autoset	0.36±0.47	0.32±0.40	0.87±1.12
Remstar Auto	0.87±0.41	0.69±0.54	0.80±0.56

a. Round 1 – Round 2; b. Round 1 – Round 3; c. Round 2 – Round 3.

TABLE VI. DETAILED RESULTS OF THE CPAP TREATMENT DURATION ANALYSIS.

Test number	CPAP	Actual treatment duration (hours)	Estimated treatment duration (hours)	Error on the treatment duration estimation (%)
1	S9 Autoset	4.000	4.017	0.417
2	Goodknight 420E	4.000	3.983	-0.420
3	Goodknight 420E	4.000	4.017	0.417
4	Goodknight 420E	4.000	4.000	0.000
5	S9 Autoset	4.000	4.000	0.000
6	S9 Autoset	4.000	4.017	0.417
7	S9 Autoset	4.000	4.000	0.000
8	S9 Autoset	4.000	4.017	0.417
9	Remstar Auto PR1	4.000	4.017	0.417
10	Remstar Auto PR1	4.033	4.033	0.000
11	Remstar Auto PR1	4.000	4.017	0.417
12	S9 Autoset	4.000	4.000	0.000
13	S9 Autoset	4.000	4.017	0.417
14	Remstar Auto PR1	4.000	4.000	0.000
15	Goodknight 420E	4.000	4.017	0.417
16	Remstar Auto PR1	4.000	4.050	1.250
17	Remstar Auto PR1	4.000	4.017	0.417
18	Goodknight 420E	4.000	4.017	0.417
19	S9 Autoset	4.000	4.017	0.417
20	S9 Autoset	4.000	4.033	0.833
21	Remstar Auto PR1	4.000	4.000	0.000
22	Goodknight 420E	4.000	4.017	0.417
23	Remstar Auto PR1	4.000	4.017	0.417
24	S9 Autoset	4.000	4.017	0.417
25	Goodknight 420E	4.000	4.017	0.417
26	Remstar Auto PR1	4.000	4.000	0.000
27	S9 Autoset	4.000	4.017	0.417
28	Goodknight 420E	3.983	3.983	0.000
29	Remstar Auto PR1	4.000	4.000	0.000
30	S9 Autoset	4.000	4.017	0.417
			Mean	*0.292*
			Max	*1.250*
			Min	*-0.420*
			Standard Dev	*0.312*

IV. DISCUSSION

NOWAPI is a novel telemedicine system, which provides remote monitoring of CPAP/APAP treatment of OSA patients at home. It detects critical parameters to evaluate the patient's adherence (treatment duration), and the effectiveness of the treatment (residual respiratory events) and sends them via GPRS to a secure server. In this way the data can be easily downloaded and revised by the physician or the health professional providing CPAP, who can perform a closer patient's monitoring and timely intervene to improve his/her treatment compliance.

Few systems in the market provide this kind of remote treatment monitoring, which is usually integrated in the CPAP/APAP devices and implemented with a different algorithm for each manufacturer. Since NOWAPI is a stand-alone system, it can be compatible with all the commercially available CPAP/APAP devices currently in clinical use. This fact would make it easy to remotely monitoring any patient, regardless of the specific CPAP device he/she uses.

In this study, NOWAPI system was evaluated in a bench. In a first test phase, two different CPAP/APAP machines were subjected to a previously validated set of disturbed breathing patterns [15][16] with or without NOWAPI device connected between the CPAP/APAP and the bench. The results of this phase demonstrated that the geometry of NOWAPI does not influence the CPAP treatment delivered by the devices considered in this study.

In the second test phase, NOWAPI was subjected to 30 different breathing patterns especially built for this study by assembling real respiratory flow signals recorded during

TABLE VII. DETAILED RESULTS OF THE RESIDUAL EVENT DETECTION ANALYSIS.

Test number	CPAP	Actual AHI	CPAP estimated AHI	Absolute difference AHI CPAP-Actual	NOWAPI estimated AHI	Absolute difference AHI NOWAPI-Actual
1	S9 Autoset	0.00	0.50	0.50	0.53	0.53
2	Goodknight 420E	1.00	1.33	0.33	1.36	0.36
3	Goodknight 420E	1.25	1.33	0.08	1.39	0.14
4	Goodknight 420E	1.75	1.87	0.12	1.56	0.18
5	S9 Autoset	2.00	2.50	0.50	3.11	1.11
6	S9 Autoset	2.00	3.00	1.00	2.37	0.37
7	S9 Autoset	2.25	1.66	0.59	2.49	0.24
8	S9 Autoset	2.50	2.20	0.30	2.37	0.13
9	Remstar Auto PR1	2.50	3.50	1.00	1.95	0.55
10	Remstar Auto PR1	2.48	2.20	0.28	2.84	0.37
11	Remstar Auto PR1	2.75	3.30	0.55	1.67	1.08
12	S9 Autoset	2.75	2.70	0.05	2.50	0.25
13	S9 Autoset	3.25	3.20	0.05	3.24	0.01
14	Remstar Auto PR1	3.50	3.80	0.30	1.52	1.97
15	Goodknight 420E	4.00	4.53	0.53	3.77	0.23
16	Remstar Auto PR1	4.25	4.40	0.15	4.00	0.25
17	Remstar Auto PR1	4.25	3.20	1.05	3.86	0.39
18	Goodknight 420E	4.25	5.63	1.38	3.77	0.48
19	S9 Autoset	4.25	4.70	0.45	4.75	0.50
20	S9 Autoset	4.50	5.20	0.70	3.72	0.78
21	Remstar Auto PR1	4.75	5.30	0.55	4.77	0.02
22	Goodknight 420E	5.00	7.50	2.50	5.25	0.25
23	Remstar Auto PR1	6.00	5.30	0.70	3.62	2.38
24	S9 Autoset	7.00	2.50	4.50	6.75	0.25
25	Goodknight 420E	8.00	16.07	8.07	6.50	1.50
26	Remstar Auto PR1	10.00	10.90	0.90	10.29	0.29
27	S9 Autoset	15.00	13.30	1.70	15.38	0.38
28	Goodknight 420E	20.08	32.82	12.74	17.57	2.51
29	Remstar Auto PR1	25.00	28.30	3.30	22.21	2.79
30	S9 Autoset	30.00	28.10	1.90	21.82	8.18
			Average	*0.93*		*0.87*
			Standard Dev	*1.04*		*1.59*

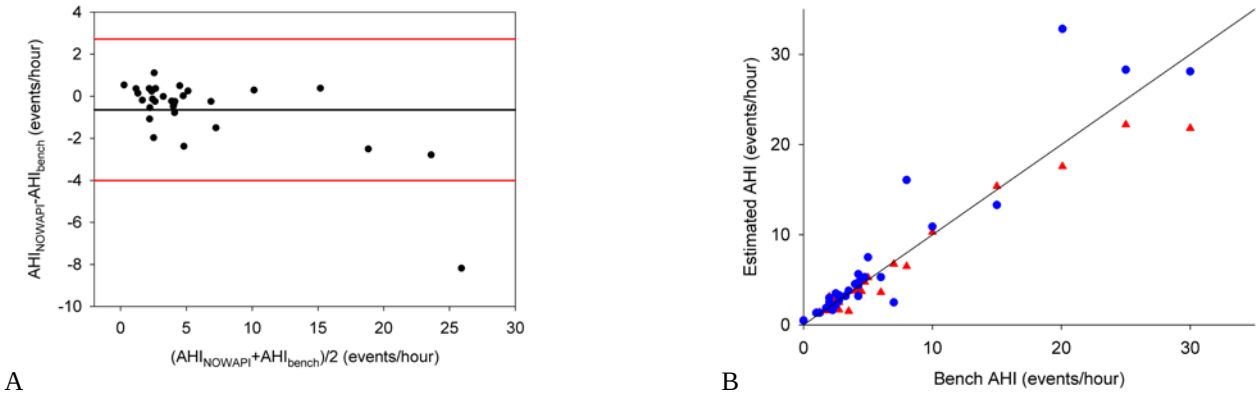


Figure 7. (A) Bland-Altman analysis of AHI values estimated by NOWAPI in each test versus the actual ones; (B) AHI values estimated by NOWAPI (red triangles) and the PAP machines (blue circles) versus the bench ones for each test.

polysomnography in OSA patients. The telemedicine system successfully sent the recorded data to the central server and showed an excellent performance in estimating the CPAP treatment duration and in detecting residual respiratory events.

This validation study was performed on a reduced number of NOWAPI devices. For this reason the conclusions of the study are limited to the tested devices. In addition, since this study was performed on a bench test and not in real patients, it could be argued that the results could lead to limited conclusions. Indeed, subjecting NOWAPI to reference breathing patterns at the bench was a first step for evaluating the performance of the hardware/software implemented in the system. The results of the study should subsequently be confirmed on patients in the clinical routine.

A bench test is a useful tool to validate new systems such as NOWAPI, because it allows the comparison of different devices response when they are subjected to exactly the same patterns of disturbed breathing, which is not possible in patients, due to the biological variability in their disturbed breathing patterns [15]. Actually, bench tests and clinical studies are both useful and should be considered complementary when evaluating a specific system [16]. Subjecting NOWAPI to reference breathing patterns at the bench was a first step for evaluating the performance of the hardware/software implemented in the system.

The encouraging results of this kind of study highlight the potential of Information and Communication Technology applications in the management of patients affected by respiratory diseases and, more generally, by chronic conditions.

V. CONCLUSION AND FUTURE WORK

NOWAPI showed good compatibility with the CPAP machines and an excellent performance in estimating the duration of the CPAP treatment and in detecting residual respiratory events in simulated OSAS patients. The results of this study demonstrated that NOWAPI system could be a valuable tool for telemonitoring the treatment of obstructive sleep apnea.

The results of the study will be verified in a clinical trial with real OSA patients, in which beside the assessment of NOWAPI system performance in a real setting, also other critical aspects will be analyzed, such us usability, implementation viability and cost-effectiveness.

ACKNOWLEDGMENT

The authors wish to thank Miguel Angel Rodriguez for his valuable technical support to the validation tests. The authors would like to thank Dr. Jordi Rigau for his contribution in the development of the bench test.

REFERENCES

[1] V. Isetta, J. M. Montserrat, G. Thiebaut, C. Weber, D. Navajas, and R. Farré, “Validation of a Telemonitoring System for Sleep Apnea Treatment,” eTELEMED 2013: The Fifth International Conference on eHealth, Telemedicine, and Social Medicine. IARIA 2013, pp. 156-161.

[2] T. Young, P. E. Peppard, and D. J. Gottlieb, “Epidemiology of obstructive sleep apnea: a population health perspective,” Am. J. Respir. Crit. Care. Med., vol. 165, May 2002, pp. 1217-1239, doi:10.1164/rccm.2109080.

[3] N. M. Al Lawati, S. R. Patel, and N. T. Ayas, “Epidemiology, risk factors, and consequences of obstructive sleep apnea and short sleep duration,” Prog. Cardiovasc. Dis., vol. 51, Jan. 2009, pp. 285-293, doi: 10.1016/j.pcad.2008.08.001.

[4] J. Duran, S. Esnaola, R. Rubi, and A. Iztueta, “Obstructive sleep apnea-hypopnea and related clinical features in a population-based sample of subjects aged 30 to 70 yr,” Am. J. Respir. Crit. Care Med., vol. 162, March 2001, pp. 685-689, doi: 10.1164/ajrccm.163.3.2005065.

[5] C. A. Kushida, M. R. Littner, M. Hirshkowitz, T. I. Morgenthaler, C. A. Alessi, D. Bailey, B. Boehlecke, T. M. Brown, J. Coleman Jr., L. Friedman, S. Kapen, V. K. Kapur, M. Kramer, T. Lee-Chiong, J. Owens, J. P. Pancer, T. J. Swick, and M. S. Wise, “Practice parameters for the use of continuous and bilevel positive airway pressure devices to

treat adult patients with sleep-related breathing disorders," Sleep, vol. 29, Mar. 2006, pp. 375-380.

[6] I. Rolfe, L. G. Olson, and N. A. Saunders, "Long-term acceptance of continuous positive airway pressure in obstructive sleep apnea," Am. Rev. Respir. Dis., vol. 144, Nov. 1991, pp. 1130-1133, doi:10.1164/ajrccm/144.5.1130.

[7] D. Sparrow, M. Aloia, D. A. Demolles, and D. J. Gottlieb, "A telemedicine intervention to improve adherence to continuous positive airway pressure: a randomised controlled trial," Thorax, vol. 65, Dec. 2010, pp. 1061-1066, doi:10.1136/thx.2009.133215.

[8] L. S. Doherty, J. L. Kiely, V. Swan, and W. T. McNicholas, "Long-term effects of nasal continuous positive airway pressure therapy on cardiovascular outcomes in sleep apnea syndrome," Chest, vol. 127, Jun. 2005, pp. 2076-2084, doi:10.1378/chest.127.6.2076.

[9] R. Zozula and R. Rosen, "Compliance with continuous positive airway pressure therapy: assessing and improving treatment outcomes," Curr. Opin. Pulm. Med., vol. 7, Nov. 2001, pp. 391-398, doi:10.1097/00063198-200111000-00005.

[10] M. Kwiatkowska and N. Ayas, "Can telemedicine improve CPAP adherence?" Thorax, vol. 65, Dec. 2010, pp. 1035-1036, doi:10.1136/thx.2010.140897.

[11] Y. Taylor, A. Eliasson, T. Andrada, D. Kristo and R. Howard, "The role of telemedicine in CPAP compliance for patients with obstructive sleep apnea syndrome," Sleep Breath., vol. 10, Sept. 2006, pp. 132-138, doi:10.1007/s11325-006-0059-9.

[12] C. E. Smith, E. R. Dauz, F. Clements, F. N. Puno, D. Cook, G. Doolittle, and W. Leeds, "Telehealth services to improve nonadherence: A placebo-controlled study," Telemed. J. E Health, Jun. 2006, pp. 289-296, doi:10.1089/tmj.2006.12.289.

[13] C. J. Stepnowsky, J. J. Palau, M. R. Marler, and A. L. Gifford, "Pilot randomized trial of the effect of wireless telemonitoring on compliance and treatment efficacy in obstructive sleep apnea," J. Med. Internet Res., 2007, e14, doi:10.2196/jmir.9.2.e14.

[14] N. Fox, A.J. Hirsch-Allen, E. Goodfellow, J. Wenner, J. Fleetham, C.F. Ryan, M. Kwiatkowska, and N.T. Ayas, "The impact of a telemedicine monitoring system on positive airway pressure adherence in patients with obstructive sleep apnea: a randomized controlled trial," Sleep, vol. 35, Apr. 2012, pp. 477-481, doi: 10.5665/sleep.1728.

[15] R. Farre, J. M. Montserrat, J. Rigau, X. Trepat, P. Pinto, and D. Navajas, "Response of automatic continuous positive airway pressure devices to different sleep breathing patterns: a bench study," Am. J. Respir. Crit. Care Med., vol. 166, Aug. 2002, pp. 469-473, doi:10.1164/rccm.2111050.

[16] J. Rigau, J. M. Montserrat, H. Wohrle, D. Plattner, M. Schwaibold, D. Navaja, and R. Farré, "Bench model to simulate upper airway obstruction for analyzing automatic continuous positive airway pressure devices," Chest, vol. 130, Aug. 2006, pp. 350-361, doi:10.1378/chest.130.2.350.

www.iariajournals.org

International Journal On Advances in Intelligent Systems

ICAS, ACHI, ICCGI, UBICOMM, ADVCOMP, CENTRIC, GEOProcessing, SEMAPRO, BIOSYSCOM, BIOINFO, BIOTECHNO, FUTURE COMPUTING, SERVICE COMPUTATION, COGNITIVE, ADAPTIVE, CONTENT, PATTERNS, CLOUD COMPUTING, COMPUTATION TOOLS, ENERGY, COLLA, IMMM, INTELLI, SMART, DATA ANALYTICS

issn: 1942-2679

International Journal On Advances in Internet Technology

ICDS, ICIW, CTRQ, UBICOMM, ICSNC, AFIN, INTERNET, AP2PS, EMERGING, MOBILITY, WEB

issn: 1942-2652

International Journal On Advances in Life Sciences

eTELEMED, eKNOW, eL&mL, BIODIV, BIOENVIRONMENT, BIOGREEN, BIOSYSCOM, BIOINFO, BIOTECHNO, SOTICS, GLOBAL HEALTH

issn: 1942-2660

International Journal On Advances in Networks and Services

ICN, ICNS, ICIW, ICWMC, SENSORCOMM, MESH, CENTRIC, MMEDIA, SERVICE COMPUTATION, VEHICULAR, INNOV

issn: 1942-2644

International Journal On Advances in Security

ICQNM, SECURWARE, MESH, DEPEND, INTERNET, CYBERLAWS

issn: 1942-2636

International Journal On Advances in Software

ICSEA, ICCGI, ADVCOMP, GEOProcessing, DBKDA, INTENSIVE, VALID, SIMUL, FUTURE COMPUTING, SERVICE COMPUTATION, COGNITIVE, ADAPTIVE, CONTENT, PATTERNS, CLOUD COMPUTING, COMPUTATION TOOLS, IMMM, MOBILITY, VEHICULAR, DATA ANALYTICS

issn: 1942-2628

International Journal On Advances in Systems and Measurements

ICQNM, ICONS, ICIMP, SENSORCOMM, CENICS, VALID, SIMUL, INFOCOMP

issn: 1942-261x

International Journal On Advances in Telecommunications

AICT, ICDT, ICWMC, ICSNC, CTRQ, SPACOMM, MMEDIA, COCORA, PESARO, INNOV

issn: 1942-2601