Proceedings of the IASTED International Conference
Signal Processing, Pattern Recognition and Applications (SPPRA 2013)
February 12 - 14, 2013 Innsbruck, Austria

USING WIZARD OF OZ TO COLLECT INTERACTION DATA FOR VOICE CONTROLLED HOME CARE AND COMMUNICATION SERVICES

Stephan Schlögl, Gerard Chollet, Pierrick Milhorat,
Jirasri Deslis, Jacques Feldmar
Signal and Image Processing Department
Télécom ParisTech
Paris, France
stephan.schlogl@telecom-paristech.fr

Jerome Boudy
Electronics and Physics Department
Télécom SudParis
Paris, France
jerome.boudy@it-sudparis.eu

Markus Garschall, Manfred Tscheligi
Center for Usability Research and Engineering
Vienna, Austria
vAssist@cure.at

ABSTRACT

This research aims at providing **V**oice controlled **Assist**ive (vAssist) Care and Communication Services for the Home to seniors suffering from fine-motor problems and/or chronic diseases. The constantly growing life expectancy of the European population increasingly asks for technological products that help seniors to manage their activities of daily living. In particular, we require solutions which offer interaction paradigms that fit the cognitive abilities of elderly users. Natural language-based access can be seen as one way of increasing the usability of these services. Yet, the construction of robust language technologies such as Automatic Speech Recognition and Natural Language Understanding does require sufficient domain specific interaction data. In this paper we describe how we plan to obtain the relevant corpus data for a set of different application scenarios, using the Wizard of Oz (WOZ) prototyping method. Using a publicly available WOZ tool we discuss how the integration of existing language technologies with a human wizard may help in designing a natural user interface for seniors and how such has the potential to underpin an iterative user-centred development process for language-based applications.

KEY WORDS

Ambient Assisted Living; Wizard of Oz; Prototyping; Language Technologies;

1 Introduction

In 2010 Europe (EU-27) counted 87 million people (17.38% of its population) older than 65 years of age and current projections suggest that by 2060 we will have less than two people of working age (15-64 years) for every person beyond 64 [8]. One result of this development may be a re-assessment of existing retirement systems. Another, however, is an increased engagement in products and services that help to support the life of our ageing society. While overall people may remain healthy, increased age sooner or later leads to physical and/or cognitive problems. Here Ambient Assisted Living (AAL) solutions aim to offer support. One aspect of AAL that requires particular attention is the prevalence of fine-motor problems and/or chronic diseases. European statistics show that currently 1.2 million senior citizens suffer from Parkinson's[1] and 630.000 from multiple sclerosis[2]. Hence, solutions should not only offer advanced services that help mastering health-related difficulties, but also provide adapted interaction paradigms that allow for the compensation of age-related restrictions.

In this paper we present the vAssist[3] project, which focuses on supporting seniors suffering from fine-motor problems and/or chronic diseases by applying multilingual (French, German, Italian) natural speech interaction for the use of home care and communication services. In order to augment existing services with speech-based interaction capabilities, such as Automatic Speech Recognition (ASR), Natural Language Understanding (NLU) and Text-to-Speech Synthesis (TTS), domain-specific language corpora are required. One method for obtaining initial corpus data is Wizard of Oz (WOZ), where a human wizard simulates the functions of an intelligent system. This paper presents our efforts of adapting an open-source WOZ prototyping platform to fit the needs of the vAssist project. Section 2 starts with a description of the vAssist project, its goals, the services it aims to offer and their specific use cases. Section 3 follows with an introduction to WOZ prototyping, and describes the WOZ prototyping platform that we plan to use for collecting initial language corpora. Section 4 elaborates on how we plan to adapt this platform so that it allows to use a wizard in collaboration with existing language technologies. Section 5 describes the current status of this undertaking and discusses future work, and Section 6 concludes the paper with a short summary.

[1] http://www.parkinsonsawareness.eu.com/campaign-literature/prevalence-of-parkinsons-disease/
[2] http://www.who.int/mental_health/neurology/Atlas_MS_WEB.pdf
[3] http://vassist.cure.at/home/

DOI: 10.2316/P.2013.798-104

2 The vAssist Project

vAssist aims at providing multilingual voice controlled home care and communication services for seniors who suffer from chronic diseases and/or (fine-) motor skills restrictions. The goal of the project is to develop an adaptive communication interface, which allows for a simple and efficient interaction with elderly users. While vAssist's main focus lies on providing predominantly speech-based access to services, it also wants to offer Graphical User Interfaces (GUIs) for situations in which a rather traditional interaction paradigm may be more appropriate. Furthermore, in order to lower potential adoption barriers as well as to reduce service-delivery costs, vAssist aims at using hardware platforms that often already exist in users' homes (e.g. PCs, smart TVs, mobile phones, tablet PCs). The envisioned channel independent services are developed following a User-Centered Market-Oriented Design approach (UCMOD). End users are involved in all stages of development, assuring that iteratively developed services and business models are adapted to people's requirements and specific needs. While vAssist seeks to address a broad audience by offering user-specific natural speech controlled interfaces it does not aim at the construction of yet another platform for service and interface integration. The goal of the project is rather to provide specific modules in order to enhance existing services with speech intelligence. Hence, currently available platforms and initiatives (e.g. universAAL[4]) are considered in the technical design and the exploitation strategy of the project.

2.1 The vAssist Services

The vAssist services aim at enhancing already existing software applications through integrating an addition natural language-based interaction modality. Applications that are supported in this way include communication tools (e.g. audio and video calling, messaging), search and organisation tools (e.g. contact management, calendar, internet search), as well as health tools (e.g. well-being diary, cognitive games). An explicit goal of vAssist is that, even though it will provide access to a set of different applications, it should be experienced as a single integrated product. When released, vAssist should therefore resemble a virtual 'butler' similar to the famous 'Memex' envisioned by Bush in his 1945 article [5]. Almost 70 years after Bush's article we do see a great potential in using natural language as a means to augment the life of seniors. However, an important aspect that needs to be taken into account is the adaptation of the interaction style to the cognitive skills of this specific user group. For example, previous work [28] has shown that TTS quality can significantly influence the usability of voice interfaces. Hence, for the vAssist services we plan to use a clear, slowly speaking natural voice to confirm input and to initiate confirmation and clarification utterances. Also, a simple and common language will be used and technical terms explicitly be avoided. This should lead to better comprehension as well as to an overall increased user satisfaction. Furthermore, an optional virtual (or physical) keyboard will be provided, allowing seniors to use a physical input modality in those cases where they prefer to do so. The following sections will provide a general overview of the type of services vAssist will offer.

[4]http://universaal.org/

2.1.1 Audio/Video Call

One goal of vAssist is to offer quick and easy direct access to contacts. Users will be able to audio/video call their medical advisor, their friends as well as any emergency contact. In the case of contact conflicts due to name duplication or missing details, the vAssist service will actively ask for clarification. Also, with time it will dissolve relationships between contacts, so that if a user asks to call his/her daughter, vAssist will be able to make the reference to the right entry in the phone book. From a technical perspective vAssist may support the connection to a mobile phone, a land-line phone as well as various Voice over IP (VoIP) solutions such as Skype[5] and Google Talk[6].

2.1.2 Email, SMS, MMS

Similar to the Audio/Video call service, vAssist will also offer the possibility to send messages. Message formats that will be supported include email, text message (SMS) as well as multi-media message (MMS). These message services may also be linked to scheduled appointments or reminders so that, for example, a user can be sent a picture - either via MMS or via email - of a medicine he/she should be taking. As with the other vAssist services, users will be able to compose and schedule messages via voice input. An adjustable touch keyboard will, however, be available in the case this type of input modality is preferred.

2.1.3 Contact management

Based on the existing contact application installed on a user's home device (e.g. PC, mobile phone, smart TV, etc.) vAssist wants to offer functions that allow for inserting, editing, deleting and retrieving private, medical and emergency contact information using natural spoken language. In order to keep the interaction simple, users will only be asked a for a minimum amount of information (i.e. a contact's first and last name). However, additional details such as different phone numbers, audio and video call preferences, email addresses, or a contact's birthday may be added if desired. When searching for an entry, a user will be able to simply say a contact's name and the details will

[5]http://www,skype.com
[6]http://www.google.com/talk/

be displayed or read out aloud. From there edits and adaptations may be conducted. In case the vAssist service finds duplicate names or other inconsistencies such as a missing detail, it will actively ask for clarification. The focus of the vAssist contact management service is to offer simple but powerful contact management, which should help a user administer any number of contacts without requiring cognitively demanding tasks such as searching through long lists of entries or comparing information details.

2.1.4 Calendar/Reminder

Seniors are usually confronted with a certain amount of periodic health related appointments. They may need to see their GP (General Practitioner) on a regular basis, they may have scheduled appointments with a physiotherapist or might be seeing a dietician. Furthermore, they often need to take a varying mix of medicine. Different dates and times on people's agendas can easily lead to confusions and mix-ups. The vAssist calendar service wants to help seniors organising their activities of daily living. The service will be designed so that users can store, edit and delete entries in an existing calendar program using voice input or a touch keyboard that is adjustable to different sizes. Users will be able to set reminders for both their appointments as well as their daily medicine intakes. We furthermore plan to expand some of the functions and make them accessible to formal caregivers such as a user's GP. Envisioned application scenarios for this extension include the remote adaptation of medicine dose rates or the rescheduling of appointments.

2.1.5 Internet Search

In addition to communication services, vAssist also plans to provide access to information. While especially health related subjects will be covered, users will generally be able to ask for information regarding any topic of interest. Presented content will be tailored to the user profile and will take into account information selected by the professional caregiver team. Results will be read out aloud and/or displayed on the screen of the device that was used to initiate the search request. The search feature also aims at providing an overarching functionality that connects all vAssist services. As such it catches voice commands that cannot be associated with distinct application functionalities such as sending an email or calling a friend. The goal of this approach is to improve the user experience and to emphasize the idea of vAssist being an integrated service rather than a compilation of single applications.

2.1.6 Well-being Diary

Continuous feedback on a patient's well-being is an important source of information for doctors. While health diaries have commonly been used in medical research (cf. [9]) they also prove beneficial in daily clinical practice for monitoring symptoms of health and efficacy of symptom responses [4]. For the case where patients are asked to keep a health diary themselves, mechanisms should be in place so that it is easy for them to record the relevant data. In order to offer such a simple functionality, vAssist will feature a diary assistant. Seniors will be able to talk to the system in order to record their well-being and, if requested, send a summary of this data to their caregivers (e.g. their GP). Furthermore, it will be possible to schedule a reminder, so that the daily diary entry is not forgotten.

2.1.7 Cognitive Games

Our cognitive abilities decrease with increasing age, potentially leading to slower mental processing and memory loss [7]. Cognitive stimulation is a therapy approach that is used with people suffering from early stage dementia to slow this cognitive decline. While the medical community still disagrees on the effectiveness of this method, recent international studies have produced promising results (cf. [30]). In order to offer similar possibilities to people at home, vAssist will feature a variety of cognitive games. Users will be able to play these games using voice and/or touch input. Games will be adapted to the personal cognitive abilities of a user and offer different difficulty levels. The user will also be able to authorise vAssist to transmit achieved scores to a health professional in order to check whether an adjustment of the game is needed.

3 Wizard of Oz

In order to design voice interfaces for the above described services we plan to use Wizard of Oz as a prototyping method. Just like applications based on a Graphical User Interface (GUI), software using speech and natural language-based interaction paradigms needs to be tested early in the design process. Whereas low-fidelity prototypes assessing GUI applications such as sketches and wireframes can be built relatively quickly and inexpensively, the development of prototypes evaluating the use of language technologies can be both cost and time intensive. For example, in order to be able to obtain a basic understanding of an envisioned interaction, these systems may require several hours of recorded speech, valid transcriptions, an implemented framework of rules as well as a solid error-recovery strategy. Obtaining this type of resources usually requires several person months of work. One technique that was proven to be an efficient alternative for testing and designing applications that involve language technologies is Wizard of Oz (WOZ) [6].

WOZ constitutes a prototyping method that is used by researchers and designers to obtain feedback on functionalities that would otherwise require significant resources to be implemented. In a so called 'WOZ experiment' a human 'wizard' mimics the functions of a system, either entirely or in part, which makes it possible to evaluate potential user experiences and interaction strategies without the need for building a fully functional product first. Borrowing its name from Baum's famous novel *The Wonderful Wizard of Oz* [1], the WOZ method was introduced by Kelley [14] who argues that human simulation can be an efficient empirical method for developing user-friendly natural language applications. He supports this argument by showing that only a small number of experiments (i.e. in his study he used 15 participants) and an iterative development methodology (i.e. results were integrated iteratively after every experiment run) are needed to capture most of the vocabulary and dialogue structures required to successfully operate a calendar program using natural language input. Since then WOZ prototyping has been used in a variety of different settings within as well as beyond the area of natural language processing. Gould et al. [10], for example, explored the possibilities of the 'Listening Typewriter', while Oviatt [19] investigated combined speech and handwriting-based interaction. More recent examples include Rajman et al. [22] who used WOZ to test multi-modal information retrieval, Mäkelä et al. [17] who simulated a virtual doorman and Bradley & Benyon [2] who simulated interactions with a web-based social companion. Researchers utilize WOZ in order to design dialogues, explore interaction strategies and collect language corpora. Results may lead to more realistic and rich dialogue models and can improve the naturalness of an interaction. In addition WOZ experiments are used to foster our understanding of emotions and help to explore social aspects of human-agent relations.

3.1 The WebWOZ Prototyping Platform

WOZ experiments usually require tools that help to mask the actions of the human wizard and assure that a test participant thinks he/she is interacting with a system. Often these tools are built for the demands of a specific experiment setting. In order to save costs it is, however, possible to adapt an existing WOZ tool so that it fits the given context. One publicly available solution that may qualify for this type of adaptation is WebWOZ[7] [26].

WebWOZ is a web-based open-source Wizard of Oz prototyping platform that allows for testing interaction scenarios which employ one or more Language Technology Components (LTCs). The integration of these LTCs is achieved through the use of web services. Currently it integrates ASR from Google using HTML-based Speech Input[8], on-the-fly Machine Translation (MT) services from Microsoft[9] and Google[10], and TTS provided by the Muse Speech Technology Research Platform[11]. In addition it supports the use of pre-recorded audio and video files made accessible through a web server. Depending on the application scenario those services can be turned on and off as well as be used in combination. WebWOZ is developed in Java and uses a relational database (e.g. MySQL[12]) to store experiment and logging data. In order to run experiments it is recommend to use an up-to-date web browser, which is able to adequately interpret recent HTML5 commands.

In terms of experiments WebWOZ allows for designing a variety of experiment settings. Different scenarios from classic monolingual text-to-text to multi-lingual speech-to-speech interactions are supported. From a wizard's perspective, possible tasks can reach from pure dialogue management to augmenting LTC output. That is, in WebWOZ a wizard may act as a substitute for a working dialogue manager, linking a test participant's input with an appropriate response by choosing from a set of pre-defined answer possibilities. Alternatively, however, he/she could also be focusing on enhancing the quality of a single LTC by augmenting its output. Examples may comprise choosing from an n-best list of recognition results, or the post-editing of output produced by an MT service. Table 1 lists the different features currently supported by WebWOZ.

Table 1. WebWOZ Feature List

Integrated Language Technologies	
ASR	HTML Speech Input
MT	Microsoft Translate
	Google Translate
TTS	Muse Speech Technology Research Platform
	Pre-recorded Audio Files
Supported Scenario Settings	
	Text-to-Text
	Text-to-Speech
	Speech-to-Text
	Speech-to-Speech
Supported Wizard Tasks	
	Choosing from a set of prepared utterances
	Selecting from a n-best list
	Directly augmenting output of a LTC

3.2 Platform Architecture

The WebWOZ platform has been implemented using the GOOGLE WEB TOOLKIT[13] which supports the construction of web interfaces using the JAVA programming language. The core of the platform is represented by a shared database that is responsible for exchanging input and output data between wizards and test participants. It stores both, static user data as well as dynamic logging data. A

[7] https://github.com/stephanschloegl/WebWOZ

[8] http://lists.w3.org/Archives/Public/public-xg-htmlspeech/2011Feb/att-0020/api-draft.html

[9] http://msdn.microsoft.com/en-us/library/ff512419.aspx

[10] http://code.google.com/apis/language/translate/v1/getting_started.html

[11] http://muster.ucd.ie/content/muse-speech-technology-research-platform

[12] http://www.mysql.com/

[13] https://developers.google.com/web-toolkit/

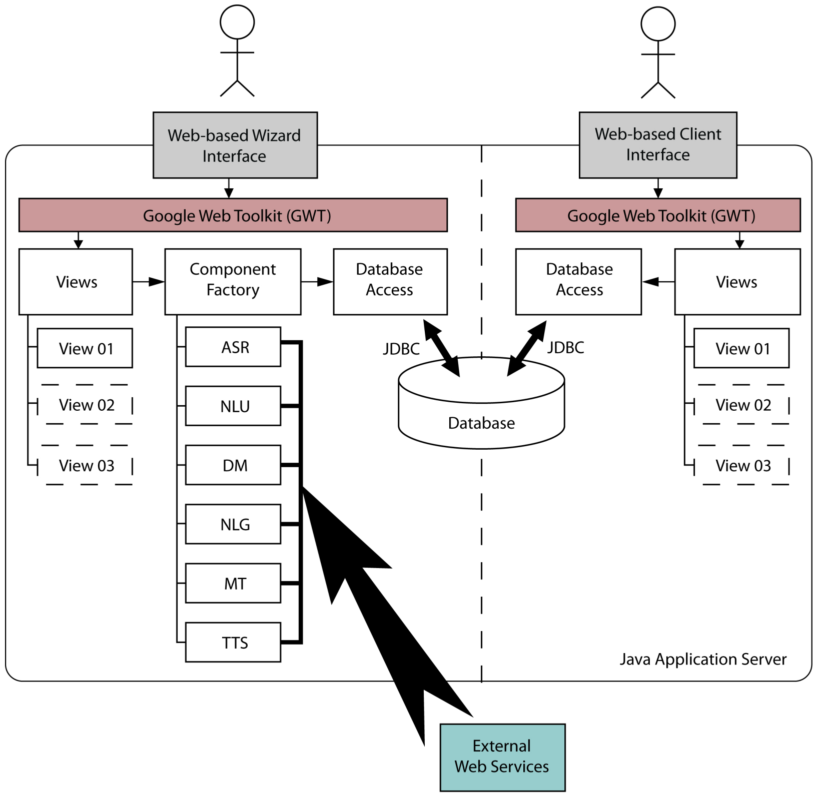

Figure 1. Architecture of the WebWOZ Wizard of Oz Prototyping Platform.

single table is used as a shared event stream. Wizard and client (i.e. test participant) interfaces are based on a modular software architecture which consists of three different types of classes. View classes are responsible for the visual layout and therefore allow for future variations and adaptations of the interfaces without influencing the business logic of the platform. Component classes are used to encapsulate various language technologies, which not only permits to develop components as separate entities but also allows for hosting them externally and accessing them through standardised web services. Finally, database access classes act as proxies providing methods for storing and retrieving data. Figure 1 shows a schematic illustration of the WebWOZ architecture and its main classes.

3.3 Integration of Language Technologies

In terms of language technology the default set-up in WebWOZ enables the wizard to choose from a number of pre-configured language components to be used in an experiment. It is possible to choose which components to turn on and which to turn off. Additional component settings that are currently supported include the language for ASR and TTS as well as the language pair, i.e. the source and the target language, for an MT component. If an experiment setting requires a more controlled set-up in which MT or TTS (or both) need to be held consistent, WebWOZ further offers the possibility to add pre-defined translations as well as pre-recorded audio.

4 Platform Adaptation for vAssist

In order to better support the different use cases of the vAssist project additional language components will be integrated with the WebWOZ platform. A gradual integration of components is planned so that by the end of next year all the modules of a spoken dialogue system (i.e. ASR, NLU, Dialogue Management, Natural Language Generation, TTS) are available and can be combined. The following sections describe the different components that will be integrated over the coming month.

4.1 Automatic Speech Recognition

In order to be able to switch between technology-based and wizard-based (i.e. simulated) speech recognition (or use the human wizard to augment technology results), we plan to integrate WebWOZ with the JULIUS[14] ASR system. JULIUS is an open-source large vocabulary continuous speech recognition (LVCSR) decoder based on n-grams and context-dependent Hidden Markov Models (HMM) [16] [15] (Note: current work aims at replacing the state-of-the-art HMM with parametrized ones [21] [29],

[14]http://julius.sourceforge.jp/en_index.php

allowing for a better control over the recognition process and more flexibility). Similar to other HMM-based toolkits (e.g. HTK[15]) JULIUS requires an initial training period to be operational. Earlier we have described a set of services for which vAssist plans to offer voice-based interfaces (cf. Section 2.1). In preparation for these services acoustic models for French (Note: French is our starting language with additional models for German and Italian to be implemented throughout the course of the project) have been trained on the ETAPE evaluation campaigns corpora [11]. N-grams were computed from the 'Le Monde' archives; years 2004-2006 (Note: Le Monde is a popular French newspaper). Next we plan to simulate these services in a series of experimental studies where we will take JULIUS recognition results and insert them into the WebWOZ event stream. The human wizard may then be able to adapt them and/or perform the relevant dialogue management tasks. Generated log-files of the wizard's actions will be used to train both the ASR system as well as the follow-up components, such as the semantic interpreter and the dialogue manager.

4.2 Semantic Interpretation

For years researchers have been working on technologies that are able to interpret natural language input (e.g. [14]). Jurčíček [13] has approached this problem by implementing a semantic interpreter based on work by Brill [3], which applies a transformation-based part-of-speech tagger to natural language utterances. Provided an optimisation function $f(o)$, an utterance-based training corpus that is annotated with semantic frames (structural representation of meaning), the learning algorithm searches for the transformation which optimises $f(o)$, and applies it to the whole corpus, then iterates again, until the value of $f(o)$ cannot be increased beyond a defined threshold. This ordered set of rules is applied sequentially to all the input utterances throughout the decoding time of the algorithm. One of the major factors defining the performance of such an algorithm is that it is fully deterministic and thus the training data needs to cover the entire range of possibilities. That is, the system cannot infer from incomplete or erroneous data. While, future work may add stochastic properties to the system and therefore learn methods from less data, at the beginning of such a process WOZ represents a viable method for data collection. Hence, in vAssist the wizard will initially be used to semantically tag utterances, generating the required data for the optimisation function. Later, the wizard's task changes to monitoring and overwriting the actions of the semantic interpreter, which consequently may be used to improve the accuracy of the stochastic model.

[15] http://htk.eng.cam.ac.uk/

4.3 Multiple-depth Semantic Analysis

Although semantic frames can be used to represent the meaning of an utterance as a flat set of instantiated slots, it is inherently a multi-level structure. For instance, if we consider a 'point in time' slot whose value may be described as precise moment in time based on a minute resolution. A valid approach for obtaining this information may be to request a precise clock-time. However, what if the user provides the time on a day scale; maybe because the application does not make sense with minute units, or maybe because it seems more natural this way? From a dialogue management perspective, the system has to maintain a mechanism that is able to cope with such partial pieces of information and convert them to acceptable ones. This is what we call multi-level semantic interpretation. The basic algorithm outputs semantic frames within which slots of unknown depth are annunciated. In a second stage, slots are combined or converted according to a hand-crafted hybrid grammar. In terms of building a technology component that is able to perform this type of multi-depth semantic analysis we again see WOZ as a powerful prototyping and simulation method, based on which machine learning algorithms may be derived. That is, a human wizard can be used to interpret and refine the output of a first level semantic interpreter and these optimisations can then be used to drive the development of the relevant component. Up to this point we have implemented a basic component using PROLOG predicates. It uses separate files to store rewriting rules and 'engine predicates' and offers an interface that can be accessed using the C programming language. In a next step it is planned to use WOZ experimentation in order to collect the relevant data that will allow us to turn this basic component into a multi-depth semantic analysis module.

4.4 Dialogue Management

The main goal of vAssist is to provide access to backend databases using spoken language requests. Also, services should be easily integrated with existing applications and allow for future expansion. In order to meet these demands we require a dialogue management component that is flexible and easy to extend. One potential solution for this task may be found in the DISCO dialogue manager. The DISCO dialogue management engine implements a shared plan theory [12] [25] [23] [24]. According to this theory, dialogues consist of one to several sub-dialogues (or sub-tasks) which are hierarchically combined so that a task/dialogue is represented as a set of actions or dialogue acts, either performed by an external entity, i.e. the user, or by a virtual agent. The structure of a dialogue is defined in task models which are based on different sets of pre-defined rules. The application of such rules converts database request constraints expressed as logic formulas over requested/optional inputs into an XML-like task model description. At this point, human expertise is still needed to add arbitrary priorities to some key nodes of this XML structure. That is, the current model seems biased towards specific sub-dialogues as the dialogue manager reacts to parametrised dialogue acts according to the task model(s) currently loaded, the current task focus and the current plan. Using WOZ experimentation we, however, believe that we can collect sufficient data to build a more flexible algorithm and make the interaction more natural.

4.5 Text-to-Speech Synthesis

The used TTS component directly influences the user experience of a spoken dialogue system. Previous research [18] has, for example, shown that users react to characteristics of computer-synthesised speech similar to how they react to characteristics of the human voice, and that the design of this aspect of an interaction therefore may influence a user's perception of a system's attractiveness, informativeness, and credibility. In order to obtain some initial feedback on potential voices to be used with the vAssist services we plan to integrate WebWOZ with the MARY Text-to-Speech platform[16]. So far MARY [20] [27] has been used to build a French synthesizer based on recordings of a Canadian-French voice talent. While this certainly serves as baseline to obtain some initial user feedback alternative solutions such as ACAPELA[17] may be used in the future.

5 Current State and Future Prospects

At this point the vAssist services are in the early stages of development. The technology components described above have only been tested in isolation and their functionalities are still limited. The coming month, however, will be dedicated to integrating the different components with the WebWOZ Wizard of Oz prototyping platform, and we are confident that by the time SPPRA 2013 will be taking place, we will be able to present first results and a working prototype. First attempts of inserting JULIUS ASR output into the event stream of WebWOZ have already proven successful. We believe that the integration of the DISCO dialogue management engine as well as the MARY Text-to-Speech platform will soon show similar advancements. With respect to the semantic interpretation module, an initial grammar has been defined. Next a set of basic WOZ studies, which will be conducted alongside the integration of the technology components, will be used to identify missing elements. After that, more targeted experimentations should allow us to progress towards a component that is capable of providing multiple depths of semantic analysis. In summary, we believe that our proposed prototyping approach that combines WOZ experimentation with technology evaluations can be seen as an effective procedure for informing the design and development of the type of domain-specific spoken dialogue system vAssist aims to offer.

[16]http://mary.dfki.de/

[17]http://www.acapela-group.com/text-to-speech-interactive-demo.html

6 Conclusion

This paper has reported on work in progress that aims at building the vAssist Voice Controlled Home Care and Communication Services for seniors suffering from chronic diseases and/or (fine-) motor skills restrictions. As part of this agenda we have discussed our plan of adapting a publicly available WOZ prototyping platform to fit the purpose of an integrated analysis tool. We have argued that WOZ is not only an efficient prototyping method for collecting language resources, but also a valuable evaluation instrument for the design and development of speech and natural language-based user interfaces. The approach of adapting an existing WOZ tool, so that it gradually allows for an interplay between a human wizard and various language technology components, furthermore supports an integrated user-centred engineering process for speech interfaces. So fart he WebWOZ platform has been integrated with the JULIUS ASR system and first test runs of this setup have shown that the approach is feasible. Over the next couple of month we plan to gradually integrate the rest of the components outlined in Section 4, so that we will be able to flexibly augment the output of technology components, such as the result of the semantic analysis, with the actions of the human wizard. Additional insights of employing this prototyping approach should be available in time for the SPPRA 2013 conference.

Acknowledgements

The research presented in this paper is conducted as part of the vAssist project (AAL-2010-3-106), which is partially funded by the European Ambient Assisted Living Joint Programme and the National Funding Agencies from Austria, France and Italy.

References

[1] L. F. Baum. *The wonderful wizard of oz*. George M. Hill, 1900.

[2] J. Bradley and D. Benyon. Wizard of oz experiments for companions. In *Proc. of the 23rd BCS annual Conf. on Human-Computer Interaction*, pages 313–317, 2009.

[3] E. Brill. Transformation-based error-driven learning and natural language processing: A case study in part-of-speech tagging. *Computational Linguistics*, 21(4):543–565, 1995.

[4] M. E. Burman. Health diaries in nursing research and practice. *Journal of Nursing Scholarship*, 27(2):1547–5069, 1995.

[5] V. Bush. As we may think. The Atlantic, 1945.

[6] N. Dahlbäck, A. Jönsson, and L. Ahrenberg. Wizard of oz studies - why and how. In *Proc. of the ACM Int. Workshop on Intelligent User Interfaces*, pages 193–200, 1993.

[7] I. J. Deary, J. Corley, A. J. Gow, S. E. Harris, L. M. Houlihan, R. E. Marioni, L. Penke, S. B. Rafnsson, and J. M. John. Age-associated cognitive decline. *British Medical Bulletin*, 92(1):135–152, 2009.

[8] European Union. Active ageing and solidarity between generations. In *Eurostat, Statistical books*, 2011.

[9] C. B. Freer. Health diaries: a method of collecting health information. *Journal of the Royal College of General Practitioners*, 30:279–282, 1980.

[10] D. Gould and T. Hovanyecz. Composing letters with a simulated listening typewriter. *Communications of the ACM*, 26(4):295–308, 1983.

[11] G. Gravier, G. Adda, N. Paulson, M. Carré, A. Giraudel, and O. Galibert. The ETAPE corpus for the evaluation of speech-based TV content processing in the French language. In *Proc. of the 8th LREC Int. Conf. on Language Resources and Evaluation*, 2012.

[12] P. Hanson and C. Rich. A non-modal approach to integrating dialogue and action. In *Proc. of the 6th AAAI Conference on Artificial Intelligence*, pages 3–8, 2010.

[13] F. Jurčíček, F. Mairesse, M. Gasić, S. Keizer, B. Thomson, K. Yu, and S. Young. Transformation-based learning for semantic parsing. In *Proc. of Interspeech*, pages 2719–2722, 2009.

[14] J. Kelley. An empirical methodology for writing user-friendly natural language computer applications. In *Proc. of the 2nd ACM SIGCHI Conf. on Human Factors in Computing Systems*, pages 193–196, 1983.

[15] A. Lee. *The Julius book*. 2010.

[16] A. Lee and T. Kawahara. Recent development of open-source speech recognition engine julius. In *Proc. of the Annual Summit of the Asia-Pacific Signal and Information Processing Association*, pages 131–137, 2009.

[17] M. Mäkelä, E.-P. Salonen, M. Turunen, J. Hakulinen, and R. Raisamo. Conducting a wizard of oz experiment on a ubiquitous computing system doorman. In *Proc. of the Int. Workshop on Information Presentation and Natural Multimodal Dialogue*, pages 9–11, 2001.

[18] C. Nass and K. M. Lee. Does computer-generated speech manifest personality? an experimental test of similarity-attraction. In *Proc. of the ACM SIGCHI Conference on Human Factors in Computing Systems*, pages 329–336, 2000.

[19] S. L. Oviatt, P. R. Cohen, M. Fong, and M. Frank. A rapid semi-automatic simulation technique for investigating interactive speech and handwriting. In *Proc. of the 2nd Int. Conf. on Spoken Language Processing*, pages 1351–1354, 1992.

[20] S. Pammi, M. Charfuelan, and M. Schröder. Multilingual voice creation toolkit for the MARY TTS platform. In *Proc. of the 6th LREC Int. Conference on Language Resources and Evaluation*, 2010.

[21] M. Radenen and T. Artieres. Contextual hidden markov models. In *Proc. of the Int. Conf. on Acoustics, Speech and Signal Processing*, pages 2113–2116, 2012.

[22] A. Rajman, M. Ailomaa, A. Lisowska, M. Melichar, and S. Armstrong. Extending the wizard of oz methodology for language-enabled multimodal systems. In *Proc. of the 5th Int. Conf on Language Resources and Evaluation*, pages 2531–2536, 2006.

[23] C. Rich. Building task-based user interfaces with ANSI/CEA-2018. *IEEE Computer*, 43(8):20–27, 2009.

[24] C. Rich and C. L. Sidner. Collaborative discourse, engagement and always-on relational agents. In *Proc. of the AAAI Fall Symposium*, pages 91–96, 2010.

[25] C. Rich, C. L. Sidner, and N. Lesh. COLLAGEN: Applying collaborative discourse theory to human-computer interaction. *AI Magazine*, 22(4):15–26, 2001.

[26] S. Schlögl, G. Doherty, N. Karamanis, and S. Luz. WebWOZ: a wizard of oz prototyping framework. In *Proc. of the ACM EICS Symposium on Engineering Interactive Systems*, pages 109–114, 2010.

[27] M. Schröder and J. Trouvain. The german text-to-speech synthesis system MARY: A tool for research, development and teaching. *International Journal of Speech Technology*, 6(4):365–377, 2003.

[28] M. A. Walker, C. A. Kamm, and D. J. Litman. Towards developing general models of usability with PARADISE. *Natural Language Engineering*, 6(3-4):363–377, 2000.

[29] A. D. Wilson and A. F. Bobick. Parametric hidden markov models for gesture recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(9):884–900, 1999.

[30] B. Woods, A. E. Spector, L. Prendergast, and M. Orrell. Cognitive stimulation to improve cognitive functioning in people with dementia. *Cochrane Database of Systematic Reviews*, 4, 2005.