



vAssist : building the personal assistant for dependent people : Helping dependent people to cope with technology through speech interaction

Hugues Sansen, Jean-Louis Baldinger, Jérôme Boudy, Gérard Chollet, Pierrick Milhorat, Stephan Schlögl

► To cite this version:

Hugues Sansen, Jean-Louis Baldinger, Jérôme Boudy, Gérard Chollet, Pierrick Milhorat, et al.. vAssist : building the personal assistant for dependent people : Helping dependent people to cope with technology through speech interaction. HEALTHINF 2014 : 7th International Conference on Health Informatics, Mar 2014, Angers, France. pp.490 - 495, 10.5220/0004912504900495 . hal-01263486

HAL Id: hal-01263486

<https://hal.science/hal-01263486>

Submitted on 27 Jan 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

vAssist: Building the personal assistant for dependent people

Helping dependent people to cope with technology through speech interaction

H. Sansen¹, J-L. Baldinger², J. Boudy², G. Chollet³, P. Milhorat³, S. Schlögl⁴

¹SHANKAA SARL, Stella-Plage, France

²Electronics and Physics Department, Télécom SudParis, Evry, France

³CNRS-LTCl, TSI Dept., Télécom ParisTech, Paris, France

⁴Dept. Management Communication & IT, MCI Management Center Innsbruck, Innsbruck, Austria

hugues@shankaa.com, jean-louis.baldinger@telecom-sudparis.eu, jerome.boudy@telecom-sudparis.eu,
gerard.chollet@telecom-paristech.fr, pierrick.milhorat@telecom-paristech.fr, Stephan.Schloegl@mci.edu

Keywords: AMBIENT ASSISTED LIVING, NATURAL LANGUAGE PROCESSING, VIRTUAL AGENTS, SPOKEN DIALOGUE

Abstract: Modern ICT solutions are capable of assisting dependent people at home and therefore able to replace the physical presence of a caregiver. However, the success of such solutions depends on an intuitive access to services. By proposing a speech-operated system and devices that facilitate this voice-based interaction, vAssist aims at a solution that corresponds to a virtual butler. The goal is to build a system with whom elderly users can interact naturally and even build up a social connection. Integrating modern language technology with a human-operated call center should allow for coping with current imperfect solutions and consequently offer the necessary reliability and user experience. vAssist is planned to be launched for German, Italian and French.

1 INTRODUCTION

Once in a while we all feel technology illiterate; irrespective of our personal and/or professional background. Despite this sporadic helplessness it may nevertheless seem difficult to imagine the everyday struggle of those people who did not grow up in our techno-centric culture, or those who slowly but surely require additional help operating technology due to emerging age-related restrictions. Unfortunately, our ‘advanced’ society is also less and less capable of caring for its aging population, resulting in caregiving costs so high that we increasingly have to rely on technologies in order to support everyday living; technologies to which we and in particular our elderly people are not used to.

The goal of an assistive service or technology should therefore be to hide complexity and increase simplicity, while at the same time providing a user experience that emulates real social life as close as possible. We demand an experience where natural speech processing is used to allow for intellectual activities and communications with 3D avatars, compensating for a person’s otherwise decreasing

social interactions. An accompanying advantage of using natural speech as an interaction modality is furthermore that it can help with difficulties elderly people have when reading or operating Graphical User Interfaces (GUI) on small screens.

However, in order for speech processing and intelligent dialog management to become key aspects of everyday life, we need to develop ‘smart’ objects for our home environments. While these objects are required to create the necessary physical link to hosted products and services, they should also homogeneously integrate with other, already existing household artifacts. Only if we achieve a seamless integration we will be able to offer to users a truly natural interaction environment – an environment where there is no need for bulky microphones and uncomfortable Bluetooth headsets. In chapter 2, we present our Speech recognition approach, in chapter 3 vAssist ecosystem and in chapter 4 the architecture of the project.

1. THE VASSIST PROJECT

vAssist¹ is a European research project funded by the Ambient Assisted Living Joint Program² which 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 that allows for a simple and efficient interaction with elderly users. While vAssist's main focus lies on providing predominantly speech-based access to services, GUIs will also be available, in particular for situations in which a rather traditional interaction paradigm is more appropriate (e.g. interactions in public places). 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).

2. THE VOICE INTERFACE

Users who are not familiar with the small user interface of a smartphone and the amount of content it shows, may be disoriented by the number of steps that are required to access a specific bit of information. Even if speaking to a machine may not seem natural, early user tests have shown a great acceptance of the concept despite its still primitive capabilities. To cope with the imperfection of the automatic system, vAssist has currently included a 'Wizard of Oz'-type service where a human replaces the machine to train the automatic dialog manager as long as possible.

2.1 vAssist Wizard of Oz

In the initial stages of the deployment of a cloud based spoken dialog system, it is desirable to experiment with a limited number of clients in order to check the acceptance of the application, collect speech data and make sure that the limitations of the speech recognition technology will not interfere with our goal. Therefore the dialog system will be hosted in a call center where operators will be available to monitor the discussions with clients. These operators will be able to listen to the spoken interactions and have access to the output of the speech recognizer.

The operator has the ability to interfere if a misunderstanding of the speech recognizer leads to improper responses.

Actions of these operators will direct the dialog in ways that may not have been planned initially. Spoken dialogs are being recorded for further adaptation of the required speech recognition resource (acoustic and language models) (Schlögl et al., 2013a).

2.2 The Spoken Dialog System

A Spoken Dialog System (SDS) is the integration of a dialog system whose core component is the Dialog Manager (DM) connected to speech input and output modalities (Milhorat et al., 2012). Such a software engages the user in an exchange of information in order to access back-end services, store user data in an organized and reusable way or simulate a conversational partner. A way to characterize a dialog system is with its input and output modalities available for interaction. This may include speech, text, gestures, or possibly a touch screen. In the case of speech, a further distinction can be made between command-and-control, keyword-based or natural language-based interaction. The first one may not be considered as "interacting" with an SDS since it only represents the association of well-defined spoken commands to actions to be undertaken by the system. A keyword-based system extends those associations by the context of the interaction as maintained by the DM. The interpretation of user utterances is based on single-word pattern detection. Keywords need to be carefully selected to avoid overlaps between intents. This paradigm therefore does not allow for subtle nuances with respect to the input that can be understood. Moreover it is very sensitive to speech recognition errors.

In the case of vAssist, the system (Milhorat et al., 2013a) is communicating with the user via natural speech, thus it does not require the user to learn a set of available commands nor a set of keywords. Instead, the natural language understanding components (Milhorat et al., 2013b) attempt to extract a meaning representation out of the word-level utterance using the external context, the internal context and some previously learned knowledge. The obtained representation (e.g. a semantic frame) is interpreted by the DM which calls back-end services and produces outputs to the user, such as requests for more information, or the meaningful represented results of service calls.

¹ <http://vassist.cure.at/>

² <http://www.aal-europe.eu/>

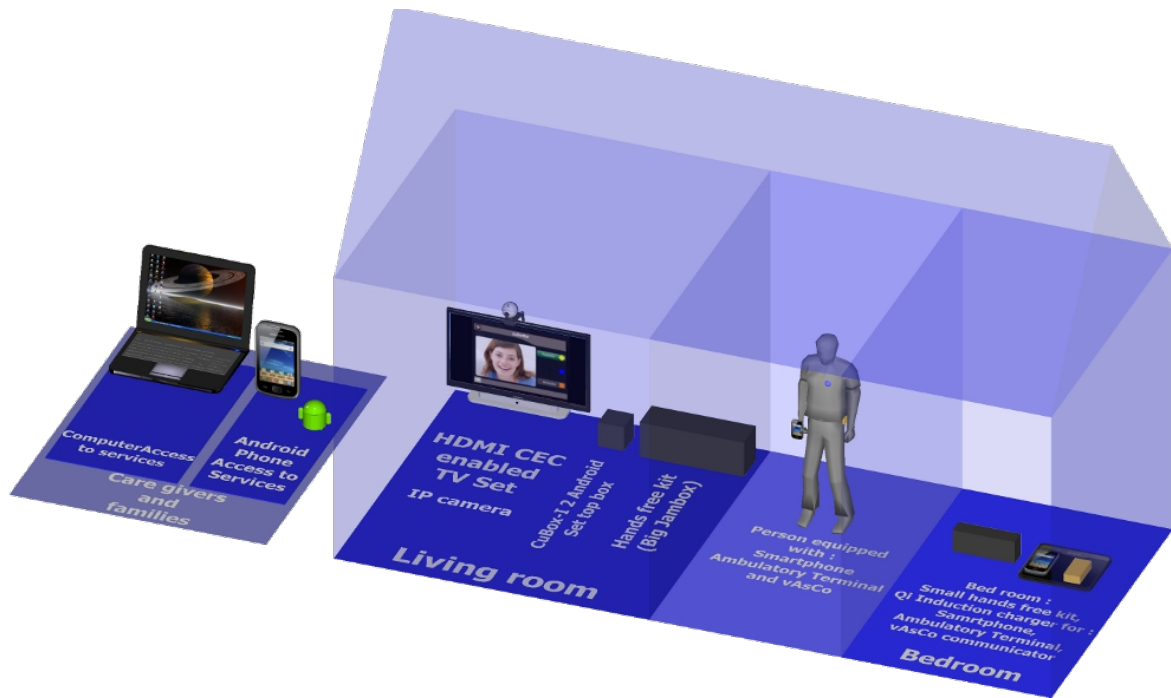


Figure 1: vAssist device ecosystem

As mentioned previously, speaking to a piece of technology, using natural language, may be disturbing; even to some extent uncomfortable. However, early trials of the simulated Wizard of Oz-based system (Schlögl et al., 2013b) produced encouraging feedback. The users' learning curves using the system proved to be steep. They highlighted the easiness of using the SDS (compared to a graphical interface) but pointed out the slow pace of the dialog progress. This was caused by several factors: (1) the slow communication with remote services, (2) the DM processing time and (3) the speed of the human operator (i.e. the wizard); the latter usually consuming the majority of the time. From the user point of view the SDS is an expert for the application. One can ask the expert to perform any task that lies within the scope it has been trained for; just as if it was a "real" full reasoning entity. Ideally, the system should be able to deflect any topic a user tries to introduce that is unrelated to its abilities. It is actually a complex task that needs to be tackled and which is still not entirely solved. Thus the user does not need any a priori know-how to use the services. By the push of a button, the agent is activated and ready to answer a request from the user, engaging in a personalized and contextualized dialog to primarily refine the intent of the caller, then perform some service calls and report back the results to the user.

3. THE VASSIST ECOSYSTEM

vAssist comes with a variety of wearable and fixed devices i.e. a smartphone, a set-top box connected to a TV set, a fall detection system called Ambulatory Terminal (AT) which analyzes a user's activity and his/her cardiac pulse rate, a wearable hands-free communication kit (vAsCo), and several fixed hands-free communication kits. All devices are interconnected and placed in a user's home. The smartphone is used as a wearable Internet hub whereas the set-top box serves as a fixed one, connecting to vAssist remote services. Overall vAssist has been designed based on an open architecture so that it may include additional systems and components. For example, one could implement voice controlled home automation through the set-top box and integrate devices such as tablets and Smart TVs. Also, the remote vAssist service platform is designed so as to include additional, pluggable services, whether they are part of the current vAssist core development activities or proposed by third parties.

3.1 The smartphone: a mobile hub

vAssist offers GUI-based Android smartphone applications, such as a DailyCare and a Pillbox app. Most of the time, however, the smartphone is used as a data communication hub for vocal interaction

(i.e. VUI) or to connect to other devices such as the Ambulatory Terminal (AT). Hardware connections are established using the Bluetooth 4.0 Low Energy standard³ which consequently helps to improve the life span of built-in batteries. Hands free kits communicate via the standard Advanced Audio Distribution Profile (A2DP), the Hands-Free Profile (HFP) and the Headset Profile (HSP) 1.1. The connection with vAssist remote services is maintained through Wi-Fi when the user is at home, and through a mobile data connection when he/she is out of the home. The smartphone location tracking system is used to turn the mobile data channel on or off, depending on whether it is needed or not.

3.2 The set-top box: a static hub

The vAssist set-top box is based on the Android mobile operating system (similar to the smartphone) and built based on a Cubox-i2 client offered by SolidRun⁴. It supports almost the same vAssist embedded services as the smartphone. The GUI applications which were specifically designed to run on the smartphone client were re-written using PhoneGap⁵. This choice was made in order to offer the applications both on the smartphone and through the web browser. The set-top box is then connected to a TV Set which supports HDMI and HDMI/CEC. The latter allows to fully control the TV (e.g. power on/off, channel selection, etc.). The set-top box furthermore supports the use of a webcam and a microphone, and will come with a Zigbee dongle⁶ in order to be prepared for voice-controlled home automation.

3.3 Voice interaction in the home

Hands-free kits are situated in a number of rooms (e.g. bed room, living room, kitchen, etc.) in the user's home where they allow for high volume vocal interaction with vAssist services. In the living room a hands-free kit can be replaced by the TV set. Hands-free kits are connected to the smartphone and to the set-top box. Currently Jawbone Jamboxes⁷ are used but they may easily be replaced by other, similar products.

3.4 VasCo: the vAssist Communicator

Currently vAsCo is not more than a hands free kit (i.e. Supertooth Buddy⁸) connected to the smartphone. Future versions, however, will take into account features such as the interoperability with hearing aids (without replacing them), a simplified charging method, the possibility to include the features of the vAssist Ambulatory Terminal, and the possibility of customization using different kinds of covers and skins.

3.5 The Ambulatory Terminal (AT): fall detection and heart rate measurement

The Ambulatory Terminal (AT) is an important feature of vAssist as it provides a certain feeling of security for elderly users. From 2007 to 2009, a research team at the University Hospital Center in Grenoble⁹ has conducted studies with different ATs, showing that these kinds of device can significantly contribute to a patient's feeling of safety and consequently fight potential distress situations such as falls and cardio-vascular problems. Hence the vAssist AT device (Baldinger et al., 2004) features two main functionalities:

(1) *Fall detection* based on accelerometers sensors and more generally the patient's actimetry level computed within the AT using the body tilt and movement rate;

(2) *Vital information* based on a patient's heart rate measured by the PPG sensor and a noise-robust pulse rate (in beats per minute) computation algorithm, making the pulse measurement fully continuous along the elderly person's activities. The AT device can furthermore communicate with the DailyCare data communication hub using a Bluetooth Low Energy connection.

Special attention is currently given to making the AT more autonomous in terms of energy consumption and smaller in order to increase overall user acceptance.

³ <http://www.bluetooth.com/Pages/Low-Energy.aspx>

⁴ <http://cubox-i.com/>

⁵ <http://phonegap.com/>

⁶ <http://www.zigbee.org/>

⁷ <https://jawbone.com/speakers/jambox>

⁸ <http://www.supertooth.net/EN/BUDDY/#page=page-1>

⁹ <http://www.social-sante.gouv.fr/personnes-agees.762/dependance-solidarite.1876/rapport-de-vincent-rialle-sur-les.5708.html>

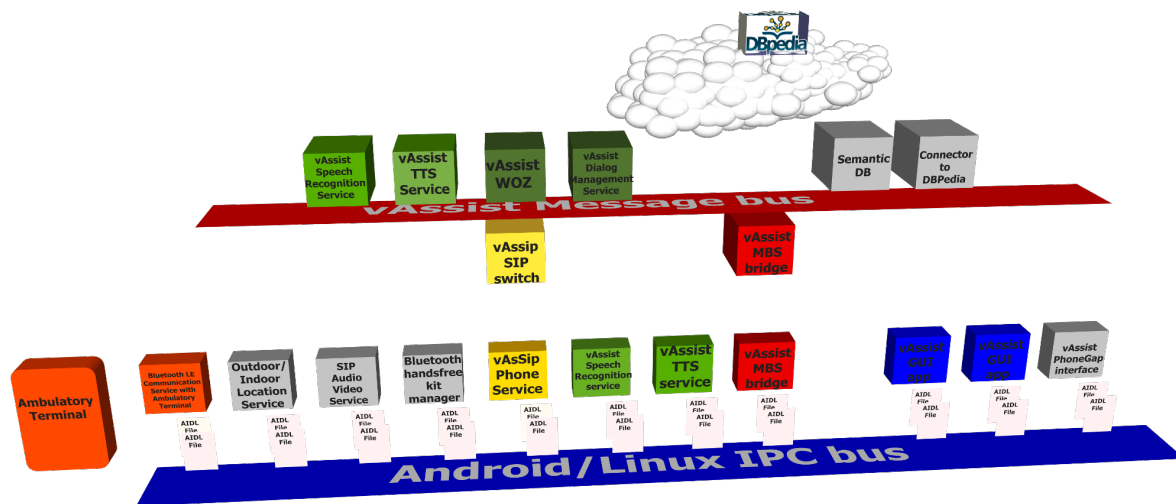


Figure 2: vAssist architecture

3.6 vAssist extensibility

The goal of vAssist is to offer a versatile framework designed to easily integrate future devices. For example, voice-controlled home automation can be integrated via the ZigBee dongle on the set-top box. Here the user is able to initiate the command from his/her smartphone. The command is analyzed by the remote speech analyzer whose result is sent on to all devices that have subscribed to the service. In the given case, the set-top box interprets the result and sends it to the ZigBee dongle, which consequently may change the light setting.,

Also, the location of the user can be used as an additional piece of contextual information. So it is for example possible to identify the exact light bulb that should be switched on or off, based on where the user is currently located in the room. As more and more health care and well-being communication devices become available on the market, the potential extensibility of vAssist is also increasing, opening up additional application scenarios for assisted living.

3.7 Qi Induction: A universal charging solution for devices

Connecting micro USB cables to devices in order to charge their battery can sometimes be a tedious task, even for light-fingered people. For elderly and dependent people it may be impossible. Hence, in order to better facilitate this task all vAssist devices come with QI inductive charging. Ideally, a charger capable of charging three devices

simultaneously should be used, i.e. the ambulatory terminal, a smartphone and the vAsCo. However, due to unavailability of this mat in Europe we currently experiment with the Zens Dual Wireless Charger¹⁰.

4. VASSIST ARCHITECTURE

The overall vAssist architecture relies on message buses both on the server as well as on the device side. A device based service maintains the connection with the server for all vAssist applications that subscribe to a topic. This architecture allows for a flexible construction of services that do not require a constant Internet connection. Furthermore it support the start of device applications via server-side application requests.

¹⁰ <http://www.makezens.com/products/zens-dual-wireless-charger/>

4.1 The remote servers and the message bus architecture

As the user will have at least two entry points for interacting with vAssist services, the Android smartphone and the Android set-top box (potentially more in the future due to increasing developments with respect to the 'Internet of Things'), synchronization is an important aspect i.e. an action on the smartphone must be reflected on the set-top box and vice versa. Hence, all vAssist services are connected to a backbone built on a message bus to which every user has to subscribe.

4.2 Embedded services and applications

vAssist provides a variety of embedded services that are used to track the user context. This includes an indoor location system, a user behavior learner that alerts in case a user does not follow his/her usual activity, a service connecting to the AT, a permanent audio connection to the remote dialog service, a SIP phone, a daily care application as well as cognitive games. All those services and applications inter-operate via the message bus built on top of the Android standard IPC bus. The embedded message bus integrates a bridge service that performs and maintains the connections to the remote Message Bus for all vAssist applications that subscribe to remote services, whether they are running or not. This mechanism allows the remote services to start applications on the device or on the set-top box when needed.

5. CONCLUSIONS

The authors started their vision of building a voice interacting personal butler back in the 1980's. Around the same time, Apple released a video of the Knowledge Navigator¹¹ which synthesized what a personal assistant could be. vAssist is an attempt to bring this technology to those who need it the most i.e. dependent people. Early vAssist testings have shown a great acceptance of voice interaction among targeted users. To further increase the acceptance of such cognitive prosthesis, the consortium has put effort into streamlining technology usage by hiding complexity and facilitating acceptance on various levels (from setting up the system to charging

component batteries). Finally, the system as a whole offers an open architecture so that it can be integrated with a wider ecosystem of already existing services.

6. ACKNOWLEDGMENTS

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.

7. REFERENCES

- Baldinger J.L., J. Boudy, B. Dorizzi, J.-P. Levrey, R. Andreao, C. Perpère, F. Delavault, F. Rocaries, C. Dietrich, A. Lacombe, "Tele-surveillance System for Patient at Home: the MEDIVILLE system", *Conference ICCHP 2004, Paris, July 2004*.
- Caon D., Simonnet T., Sendorek P., Boudy J., Chollet G. 2011. vAssist: The Virtual Interactive Assistant for Daily Home-Care. In *pHealth'11, 8th International Conference on Wearable Micro and Nano Technologies for Personalised Health*. Lyon, France.
- Chollet G., Amehraye A., Razik J., Zouari L., Khemiri H., Mokbel C., 2010. Spoken Dialogue in Virtual Worlds. In *Development of Multimodal Interfaces : Active Listening and Synchrony*, Esposito A. (Ed.). Springer Verlag, LNCS-5967, pp. 423-443.
- Milhorat P., Istrate D., Boudy J., Chollet G., 2012. Hand-free Speech-Sound Interactions at Home. In *EUSIPCO. Budapest*.
- Milhorat P., Schlögl S., Boudy J., Chollet G., 2013a. What if Everybody could do it ? A framework for easier Spoken Dialog System Design. In *Symposium on Engineering Interactive Computing Systems. ACM SIGCHI, London, UK*.
- Milhorat P., Schlögl S., Chollet G., Boudy J., 2013b. Multi-step Natural Language Understanding. In *SIGDial Meeting on Discourse and Dialogue. Metz, France*.
- Schlögl S., Chollet G., Milhorat P., Deslis J., Feldmar J., Boudy J., Garschall M., Tscheligi M., 2013a. Using Wizard of Oz to collect Interaction Data for Voice Controlled Home Care and Communication Services. In *International Conference on Signal Processing, Pattern Recognition and Applications. SPPRA, Innsbruck, Austria*.
- Schlögl S., Chollet G., Garschall M., Tscheligi M., Legouverneur G. 2013. Exploring Voice User Interfaces for Seniors. In *International Conference on Pervasive Technologies Related to Assistive Technologies. PETRA, Island of Rhodes, Greece*.

¹¹ http://www.dailymotion.com/video/xkvhq7_apple1987_tech