Toward a Virtual Assistant for Vulnerable Users:
Designing Careful Interaction

Ramin Yaghoubzadeh
Sociable Agents Group, CITEC
Bielefeld University; PO Box 100131
33501 Bielefeld, Germany
ryaghoub@techfak.uni-bielefeld.de

Stefan Kopp
Sociable Agents Group, CITEC
Bielefeld University; PO Box 100131
33501 Bielefeld, Germany
skopp@techfak.uni-bielefeld.de

Abstract

The VASA project develops a multimodal assistive system mediated by a virtual agent that is intended to foster autonomy of communication and activity management in older people and people with disabilities. Assistive systems intended for these user groups have to take their individual vulnerabilities into account. A variety of psychic, emotional as well as behavioral conditions can manifest at the same time. Systems that fail to take them into account might not only fail at joint tasks, but also risk damage to their interlocutors. We identify important conditions and disorders and analyze their immediate consequences for the design of careful assistive systems.

1 Introduction

In 2001, the World Health Organization consolidated previous taxonomies of somatic and mental functions and the everyday needs of human beings into the comprehensive International Classification of Functioning, Disability and Health *ICF* (WHO, 2001). Older people, as well as people with impairments, often need support from others to satisfy those basic needs, among which are activities related to self-care, to mobility, but also to communication and management of the daily activities and the social environment. For many older people, a catastrophic event, most often either a fall or the passing of their spouse, leads to their sudden loss of autonomy and subsequent submission into stationary care. In the latter case, the loss of their day structure is frequently the intermediate cause. The same effect can be observed for many disabled people of all ages who must make a transition from assisted living to stationary care. Here, specialized systems that assist in preserving autonomy in a spectrum of daily need fulfillment can potentially be of great benefit.

The present paper introduces the VASA project ("Virtual Assistants and their Social Acceptability"), which in cooperation with a health-care foundation examines how both older patients and people with various impediments, congenital or acquired, both in stationary and assisted living, can be provided with technical assistance to maintain autonomy for as long as possible. Importantly, we are not focusing on physical assistance, but on supporting a person's capability for organizing a social environment (WHO ICF d9205: 'Socializing') and managing the day structure generally (d230: 'Carrying out daily routine'). These two tasks turned out to be crucial in our analysis with the health care personnel. We thus aim to develop an assistive system for (1) managing daily routine and weekly appointments with the user, and (2) accessing a video call interface for contacting acquaintances or care personnel (d360: 'Using communication devices and techniques').

But how should such a system meet its user, and what criteria should guide the system interface design? Research has shown that older users are far more likely to employ a 'social' conversation style with a system (Wolters et al., 2009). The VASA project explores the use of a "virtual assistant", an humanoid conversational agent that features natural-language and touchscreen input and human-like communicative behavior (speech, gesture; see Fig. 1 for the current running prototype).
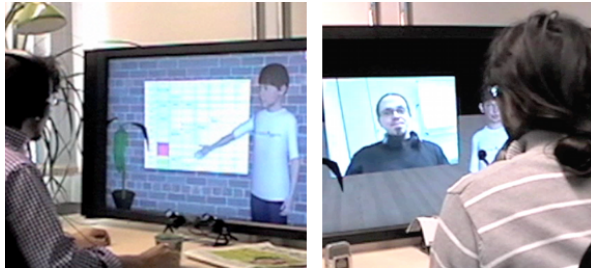
13

Figure 1: The VASA system. Left side: natural-language calendar management; right side: video call interface.

In this paper we review work on related systems for older people and people with disabilities. We then argue that beside the general goal of maximizing usability for this specific user group, there is an enhanced *vulnerability* of these users that calls for special care in interaction design; we substantiate this view by an analysis of potential mental conditions of the prospective users along with discussions of what requirements arise from them.

2 Related work

Generally, assistive systems are driven by task reasoning systems as well as components for human-computer interaction, which can be specialized for older or disabled persons. Modern systems that attempt to provide a "natural" interaction are being developed and evaluated, including touch-screen and haptic interfaces and interfaces capable of understanding and generating natural language, all of them providing an immediacy between communicative intentions and their execution that makes them suitable especially for users without technical expertise, with reduced sensorimotor skills or reduced capability for learning new interaction paradigms, as is frequently the case with older or impaired persons.

The performance of such systems in terms of suitable operation in interaction, successful task completion, and user-reported satisfaction, has been subject to systematic evaluation under controlled conditions: The performance of speech recognition systems has been compared between base-line users and people with varying degrees of dysarthria (breathiness, dysfluencies, involuntary noises). Off-the-shelf speech recognition systems have higher failure rates with dysarthric speakers (Raghavendra et al., 2001). Mildly and moderately dysarthric

speakers can attain a recognition accuracy of 80% in dictation systems, breath exercises and phonation training improve performance (Young and Mihailidis, 2010). Vipperla et al. (2009) compared speech recognition for younger and older users, reporting an 11% baseline increase in word error rates for the latter group, attributed to both acoustic and linguistic causes. The Stardust project succeeded in very high single-word recognition rates on small dictionaries in patients with severe dysarthria, enabling them to control their environment by voice (Hawley et al., 2007). Fager et al. (2010) implemented a multimodal prototype system that combined ASR with a word prediction model and a capability to enter an initial letter, leading to an accuracy of $> 80\%$; noting that other conditions, such as a reduced visual field or ataxia, had to be addressed with technical solutions for each individual. Jian et al. (2011) designed a system for direction giving for seniors, suggesting specific design guidelines addressing typical perceptive, motor, attentive and cognitive impairments of older users. The evaluation of their multimodal system (speech and touch/image) led to positive results with respect to effectivity, efficiency and user-reported satisfaction.

3 Careful Interaction with Vulnerable People: Analysis

The more autonomously assistive systems act, the higher the potential negative effects they can consequentially cause. This is especially true for robotic systems, since their extension into the physical world entails possible harmful effects if proper reasoning or safety precautions should be breached by unanticipated events. But even without physical manipulation, real damage can still be done. This might be due to misunderstandings, leading to wrong assumptions in the system, and hence to actions being performed on behalf, but actually to the detriment, of the user. It might, however, also be due to the wrong things being communicated, or communicated in an inappropriate manner, leading to unnecessary negative appraisal, discomfort, or triggering of a negative psychic condition in the user. While unlikely to cause damage in an interaction with the average healthy interactant, this issue is of the utmost importance for many potential user groups.

Frail or potentially unstable users are arguably among those who can derive the greatest benefits from easily accessible assistive systems, enabling them to perform tasks which they might else not, or no longer, perform, thus preserving their autonomy. However, they are at the same time affected by a multitude of possible cognitive, psychic and emotional conditions and behavioral anomalies that can occur simultaneously. Each of these conditions entails special constraints for interactive systems, either for the interaction channels, for the contents, or for both. Several factors have been accounted for in existing systems: Reduced perceptive faculty (vision, hearing), reduced motor abilities (ataxia), and attention and memory impairments, mitigated by best-practice rules (Jian et al., 2011). Attempts to account for users with mild dementia have been made, such as in the ISISEMD project. Avoiding a deep hierarchy of dialogue structures and providing extra information (repetition, paraphrase) rather than maximum parsimony are paramount in cases of impaired memory and abstraction faculty, whereas people with learning difficulties need a system that operates without extensive training (of the user).

For systems that strive to provide long-term support to a specific person, adaptation to that person is of vital importance – by employing user models that are adapted either manually or using learning algorithms. System behavior should be adapted both in the content provided as well as the form it is provided in, to enable a working relationship that is both effective and pleasant for the user (Yaghoubzadeh and Kopp, 2011). This alone however is insufficient; since the vulnerability of the actual clientele in VASA is considerable, each of the encountered mental conditions has to be analyzed and additional dialogue constraints be enforced before autonomous interactions can be permitted. There is a variety of such factors that have not yet been comprehensively addressed, but might cause critical damage to some interactants if not considered. The following section captures the most frequently encountered phenomena, which were identified in dialogue with care personnel:

- **Depression and Bipolar Disorder:** Roughly ten percent of the population suffer from depression at some point in their lives. Depres-

sion increases the risk for suicide ten- to twentyfold (Sadock et al., 2007). Bipolar disorder manifests in episodic effects, where sensations of racing thoughts and heightened activity (mania) and listlessness and social passivity (severe depression) alternate or occur simultaneously; depressive relapses in particular are points of vulnerability (Hill and Shepherd, 2009). There are successes in detecting depressive states from facial and voice cues at $> 80\%$ rate (Cohn et al., 2009). A good practice is to employ mitigation strategies when breaking bad news to the user (Brown and Levinson, 1987; Fraser, 1980), e.g. by presenting obligations as options (Williams, 2011), or presenting the "bad news" simultaneously with "good news". We provide for discussion another requirement for interactive systems in this case: *The system must not produce ambiguously interpretable answers* – consider a catastrophic answer of "okay" as an affirmative response to a wrongly parsed utterance that was actually an expression of intent for suicide, a frequent phenomenon with risk patients (Kelly, 2009).

- **Borderline Personality Disorder:** This type of disorders, characterized by emotional instability, can lead to anxiety, social insecurity and depression, but also inappropriate outbursts of anger. Anger management techniques are employed to inhibit the expression of such anger (Swaffer and Hollin, 2009). An assistive system should be able to *cope with impulses of anger*, and as a bare minimum interrupt the interaction and offer to resume it at a later point. The EmoVoice system, for instance, can classify emotional features in natural language with good rates (Vogt et al., 2008), and could be used to identify anger.

- **Epilepsy:** Patients with acquired brain injuries frequently suffer from epilepsy. Even short (petit mal) epileptic seizures can lead to temporary absence and periods of confusion and disorientation (APA, 2000). In such a situation, the patient may utter irrational sentences or be silent altogether. An assistive system should be able to *detect these irrational deviations from*

the course of conversation, and fill the user in again, abort the conversation, or call for help.

- **Panic:** Proneness to panic attacks can result from a multitude of afflictions and is hard to predict. In the event of a panicking interactant, the system should not take steps that could further exacerbate the situation. According to literature (Gournay and Denford, 2009), panic attacks are generally unable to do any real harm and subside quickly. Therefore, passivity from the system's side, in a neutral mode, is the minimal appropriate behavior. Panicking people are most likely not able to perform in interaction as successfully as usual – systems that should still be operable by a user in this situation must provide minimalistic shortcuts to essential features (i.e. a "panic button" for emergencies).

- **Anxiety:** Special care must be taken in the design of systems aimed at people with social anxiety. Interactants might be hesitant to open a conversation even with an artificial system. The system could take the initiative by simply opening with a short utterance about the task domain (Williams, 2011).

- **Phobias and Impulse Control Disorders:** Phobic disorders and obsessive-compulsive disorders can be triggered by environmental cues (Gournay and Denford, 2009). User interfaces have to take this into account, and *avoid presenting stimuli that could act as potential triggers* (e.g. people with an insect phobia should neither be presented with pictures of insects, nor their verbal mention). The same precautions are valid in the case of addictions.

Any interactive system, and in particular systems that do not only provide information but can also be made to perform tasks autonomously on behalf of the user, must be designed with all possible afflictions of all possible users in mind, not only as a wise legal precaution, but also as an ethical obligation to the designer. We argue that, quite unlike the 'best practices' of user interface design, there is no degree of optionality to the implementation of the above constraints and countermeasures, but that it must be performed with all musterable diligence. Some constraints are especially hard to meet in open-world systems (e.g. with free Internet access), since the contents presented are harder to predict.

Note that the set of conditions presented above is by no means comprehensive. For instance, we have, for now, altogether omitted an incorporation of autism-spectrum disorders or of functional psychoses such as schizophrenia, paraphrenia and paranoia – which are not uncommon in the older population (Ashton and Keady, 2009).

4   Summary

The VASA project is developing a multimodal natural-language agent-mediated assistance system for older people and patients with disabilities for enhancing their autonomy in the everyday tasks of communication and activity management. The clientele is afflicted with a variety of cognitive, psychic, and emotional conditions that have to be dealt with with extreme care and entail a necessity for specific safety mechanisms which will be implemented for VASA in coordination with the care personnel. We attempted to identify common conditions of older and impaired patients that should be considered and resolved in any assistive system (or indeed any autonomous interactive system) that might communicate with them. Factors that could lead to a detrimental outcome of such an interaction include depression, emotional instability, disorientation, panic, anxiety and phobia. Some constraints on the design rationale for such systems can provide a mitigation of those risks: avoiding ambiguity in the system's utterances, coping with anger, irrationality and panic by employing appropriate system responses, capability for system-side initiative, and preventing inadvertent stimulation of disorders. Since the field of potential interactants for generic assistive systems is vast, as any inspection of a larger health-care institution will show, more discussion in the research community should aim at establishing a stable ontology of their special needs and the ramifications for the design of careful assistive systems.

Acknowledgments

References

American Psychiatric Association. 2000. *Diagnostic and Statistical Manual of Mental Disorders DSM-IV-TR, Fourth Edition.* American Psychiatric Publishing, Inc., Arlington, VA.

Peter Ashton and John Keady. 2009. Mental disorders of older people. Newell & Gournay (eds.), *Mental Health Nursing: an Evidence-Based Approach*, 341–370. Churchill Livingstone, Philadelphia, PA.

Paul Brown and Stephen Levinson. 1987. *Politeness. Some Universals in Language Usage.* Cambridge University Press, Cambridge.

Jeffrey F. Cohn, Tomas Simon Kruez, Iain Matthews, Ying Yang, Minh Hoai Nguyen, Margara Tejera Padilla, Feng Zhou, and Fernando de la Torre. 2009. Detecting Depression from Facial Actions and Vocal Prosody. *Proceedings of the 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops (ACII 2009)*, 1–7. IEEE, Amsterdam.

Susan K. Fager, David R. Beukelman, Tom Jakobs, and John-Paul Hosom. 2010. Evaluation of a Speech Recognition Prototype for Speakers with Moderate and Severe Dysarthria: A Preliminary Report *Augmentative and Alternative Communication*, 26(4):267-277.

Bruce Fraser. 1980. Conversational mitigation. *Journal of Pragmatics*, 4:341–350.

Kevin Gournay and Lindsay Denford. 2009. Phobias and Rituals. Newell & Gournay (eds.), *Mental Health Nursing: an Evidence-Based Approach*, 207–224. Churchill Livingstone, Philadelphia, PA.

Mark S. Hawley, Pam Enderby, Phil Green, Stuart Cunningham, Simon Brownsell, James Carmichael, Mark Parker, Athanassios Hatzis, Peter ONeill, and Rebecca Palmer. 2007. A speech-controlled environmental control system for people with severe dysarthria. *Medical Engineering & Physics*, 29(5):586–593.

Robert Gareth Hill and Geoff Shepherd. 2009. Disorders of Mood: Depression and Mania. Newell & Gournay (eds.), *Mental Health Nursing: an Evidence-Based Approach*, 165–185. Churchill Livingstone, Philadelphia, PA.

Cui Jian, Nadine Sasse, Nicole von Steinbüchel-Rheinwall, Frank Schafmeister, Hui Shi, Carsten Rachuy, and Holger Schmidt. 2011. Towards effective, efficient and elderly-friendly multimodal interaction. *Proceedings of the 4th International Conference on Pervasive Technologies Related to Assistive Environments (PETRA 2011)*, article 45, 1–8. ACM, New York, NY.

Sarah Kelly. 2009. Suicide and Self-Harm. Newell & Gournay (eds.), *Mental Health Nursing: an Evidence-Based Approach*, 187–206. Churchill Livingstone, Philadelphia, PA.

Parimala Raghavendra, Elisabet Rosengren, and Sheri Hunnicutt. 2001. An investigation of different degrees of dysarthric speech as input to speaker-adaptive and speaker-dependent recognition systems. *Augmentative and Alternative Communication*, 17(4):265-275.

Benjamin J. Sadock, Harold I. Kaplan, and Virginia A. Sadock. 2007. *Kaplan & Sadock's Synopsis of Psychiatry: Behavioral Sciences/Clinical Psychiatry.* Lippincott Williams & Wilkins, Philadelphia.

Tracey Swaffer and Clive R. Hollin. 2009. Anger and Impulse Control. Newell & Gournay (eds.), *Mental Health Nursing: an Evidence-Based Approach*, 267–289. Churchill Livingstone, Philadelphia, PA.

Ravichander Vipperla, Maria Wolters, Kallirroi Georgila, and Steve Renals. 2009. Speech input from older users in smart environments: Challenges and perspectives. *HCI (6): Universal Access in Human-Computer Interaction, Intelligent and Ubiquitous Interaction Environments*, LNCS 5615:117–126. Springer, Heidelberg.

Thurid Vogt, Elisabeth André, and Nikolaus Bee. 2008. EmoVoice - A framework for online recognition of emotions from voice. *Proceedings of the 4th IEEE tutorial and research workshop on Perception and Interactive Technologies for Speech-Based Systems (PIT 2008)*, 188–199. Springer, Heidelberg.

Val Williams. 2011. *Disability and discourse : analysing inclusive conversation with people with intellectual disabilities.* Wiley-Blackwell, Chichester, West Sussex / Malden, MA.

World Health Organization. 2001. *International Classification of Functioning, Disability and Health: ICF*, WHO, Geneva, Switzerland.

Maria Wolters, Kallirroi Georgila, Johanna D. Moore, and Sarah E. MacPherson. 2009. Being Old Doesn't Mean Acting Old: How Older Users Interact with Spoken Dialog Systems. *ACM Transactions on Accessible Computing (TACCESS)*, 2(1):1–39.

Ramin Yaghoubzadeh and Stefan Kopp. 2011. Creating familiarity through adaptive behavior generation in human-agent interaction. *Proceedings of the 11th International Conference on Intelligent Virtual Agents (IVA 2011)*, LNCS(LNAI) 6895:195–201. Springer, Heidelberg.

Victoria Young and Alex Mihailidis. 2010. Difficulties in Automatic Speech Recognition of Dysarthric Speakers and Implications for Speech-Based Applications Used by the Elderly: A Literature Review. *Assistive Technology*, 22(2):99–112.