

# Speech and Gesture Interaction in an Ambient Assisted Living Lab

**Dimitra Anastasiou**  
SFB/TR8 Spatial Cognition,  
Languages Science,  
University of Bremen,  
Germany  
anastasiou@uni-  
bremen.de

**Cui Jian**  
SFB/TR8 Spatial  
Cognition,  
University of Bremen,  
Germany  
ken@informatik.uni-  
bremen.de

**Desislava Zhekova**  
Department of Linguistics,  
Indiana University, USA  
dzhekova@  
indiana.edu

## Abstract

In this paper we describe our recent and future research on multimodal interaction in an Ambient Assisted Living Lab. Our work combines two interaction modes, speech and gesture, for multiple device control in Ambient Assisted Living environments. We conducted a user study concerning multimodal interaction between participants and an intelligent wheelchair in a smart home environment. Important empirical data were collected through the user study, which encouraged further developments on our multimodal interactive system for Ambient Assisted Living environments.

## 1 Introduction

Multimodal interaction has been gaining more and more importance in various application systems and domains. On one hand, it is considered as an encouraging way to improve the effectiveness and efficiency of interaction in general, and on the other hand, to increase user satisfaction in a more natural and intuitive manner (Jaimes & Sebe, 2007; Oviatt, 1999).

Meanwhile, the domain of Ambient Assisted Living (AAL), although very significantly and increasingly well researched in recent literature (Steg et al., 2006; Fuchsberger, 2008; Wichert & Eberhardt, 2011), has not been enriched with profuse advanced multimodal technologies so far.

Therefore, generally as well as particularly in assistive environments, multimodal applications are not only preferred, but also often the only solution for people who cannot master their everyday tasks by themselves. These multimodal applications are showing their necessities of compensating for the various visual, perceptual, sensory, cognitive, and motoric impairments of senior and/or disabled people.

We focus on speech and gesture as two intuitive modalities which can be combined to compensate for physical and/or cognitive limitations; speech interaction for those who have motor disabilities and gesture for those with speech impairments. A Wizard-of-Oz (WoZ)-controlled user study concerning multimodal interaction between participants and an intelligent wheelchair took place in our AAL lab. A drawback in the design of gesture-based user interfaces today is the lack of experience and empirical data about which gestures are required for which activities. Our goal in the presented user study is to collect empirical speech and gesture data of natural dialogue in Human-Robot Interaction (HRI).

This paper is laid out as follows: in section 2 we present the most relevant work on speech interaction in assistive environments (2.1), spatial gestures in assistive environments (2.2) and speech-gesture interaction (2.3). Section 3 gives an introduction to our Ambient Assisted Living Lab and the intelligent wheelchair. Section 4 describes the current speech interaction with the wheelchair in our assistive environment/smart home. Section 5

reports on a user study focusing on the speech-gesture interaction within this environment. Conclusion and future work follow in section 6.

## 2 Related Work

One of the main goals of AAL is to alleviate and compensate for the disabilities of its inhabitants. The latter are often predisposed to constant and permanent increase of their inability to orally express themselves and/or adequately employ elementary motoric functions. Thus, efficient and dynamic interaction of the AAL modalities should be targeted in order to balance the lack of either eloquence and/or movement. In case further deficiencies are present, additional modalities that can account for those deficiencies should be considered. For example, if a person experiences any kind of speech disorder, the speech modality can be exchanged with a typed-text interface. Yet, in the following subsections we will concentrate on related work about two common interaction modalities, speech and gesture, and the way they can interact with each other.

### 2.1 Speech interaction in assistive environments

As one of the most important interaction modalities in assistive environments, there has been a significant amount of research on speech interaction regarding different kinds of motivation and technology-dependent approaches.

Some studies are focusing on gathering objective and subjective evidence for motivating and supporting further development on speech interaction. Takahashi et al. (2003) collected dialogue examples and conducted a recognition experiment for the collected speech; Ivanecky et al. (2011) found that the set of the commands for the house control is relatively small (usually around 50).

At the same time, other research concerning general-purpose speech-enabled dialogue systems has also been reported. Goetze et al. (2010) described technologies for acoustic user interaction in AAL scenarios, where they designed and evaluated a multimedia reminding and calendar system. The authors carried out an automatic speech recognition (ASR) performance study having as training set both male and female speakers of different age and hearing loss. The

results showed that the ASR performance was lower for older persons and for female. Moreover, Becker et al. (2009) carried out experiments in an assistive environment using voice recognition and pointed out that “the speech interface is the easiest way for the user to interact with the computer-based service system”.

Furthermore, much effort has been put into considering the special requirements of assistive environments and developing the accordingly adapted interactive systems. Krajewski et al. (2008) described an acoustic framework for detecting accident-prone fatigue states according to prosody, articulation and speech quality related speech characteristics for speech-based human computer interaction (HCI). Moreover, Jian et al. (2012) studied, implemented, and evaluated the speech interface of a multimodal interactive guidance system based on the most common elderly-centered characteristics during interaction within assistive environments.

### 2.2 Spatial gestures in assistive environments

We coin the term “spatial” to describe gestures that often iconically represent spatial concepts (Rauscher et al., 1996). Alibali (2005: 307) states: “gestures contribute to effective communication of spatial information”. She added that “speakers tend to produce gestures when they produce linguistic units that contain spatial information, and they gesture more when talking about spatial topics than when talking about abstract or verbal ones”. Kopp (2005) has shown that gestures have sufficient specificity to be communicative of spatial information. Spatial gestures belong to *representational* gestures, which according to McNeill’s (1992) taxonomy can be deictic, iconic, or metaphoric. Kita (2009: 145) stated: “representational gestures (...) that express spatial contents (...) reflect the cognitive differences in how direction, relative location and different axes in space are conceptualized and processed”.

As far as spatial gestures in assistive environments is concerned, Nazemi et al. (2011) conducted an experiment, where test subjects in middle age were asked to make gestures with the *WiiMote* to scroll, zoom, renew, and navigate in a relational database. The results showed that in complex tasks, participants employed more and various gestures. Neßelrath et al. (2011) designed a gesture-based system for context-sensitive interaction with a

smart kitchen. Users had to solve interaction tasks by controlling appliances in a smart home. Recently Marinc et al. (2012) presented a demonstrator that uses *Kinect* to recognize pointing gestures for device selection and control. When a device is selected, a graphical user interface (GUI) is shown on a screen to inform the user that the interaction has started. A hand movement to the left stops the HCI.

### 2.3 Speech-gesture interaction

Concerning speech accompaniment of gestures, Chovil (1992), among others, stated that speakers frequently use gesture to supplement speech. McNeill (1992) pointed out that speech and gesture must cooperate to express a person's meaning and Goldin-Meadow (2003) stated that speech-associated gestures often convey information that complements the information conveyed in the talk they accompany and, in this sense, are meaningful. Similarly, Kendon (2004) suggests that gestures enrich the speech, helping the interlocutor to easily express concepts that will otherwise be complex to explain through speech only. McNeill (2000) points out that gestures and the synchronous speech are semantically and pragmatically co-expressive.

Specifically concerning the relationship between spatial gestures and speech, Kita (2000) stated that a possible function of gesture is that gesture may help speakers to package spatial information into units suitable for verbal output. Moreover, Hostetter & Alibali (2005) regarded individual differences in gesture and found that the gesture rate was highest among individuals who had a combination of high spatial skill and low verbal skill.

Furthermore, research has been carried out towards a grammar of gesture, in other words the relationship of gestures within a multimodal grammar. Fricke (2009) claimed that in German spoken language, co-speech gestures can be structurally integrated as constituents of nominal phrases, and can semantically modify the nucleus of the nominal phrases. Hahn & Rieser (2010) looked at the types of gestures co-occurring with noun phrases and their function, semantic values, and how these values interface with a natural language expression.

In addition, the employment of gesture to improve the semantic analysis of the dialogues in AAL has

gained considerable attention in the research community in the last decade. In particular the effect of gesture on the improvement of co-reference resolution (the process of determining if two phrases in a discourse refer to the same real-world entity) has been examined in a variety of studies. Eisenstein and Davis (2006) consider various gesture features and delineate their importance for the co-reference process. Chen et al. (2011) show that when the pronominal mentions are typed and simultaneously a pointing gesture is used, the co-reference performance improves for personal and deictic pronouns. Co-reference in spoken dialogues has proven to be much more different than the one we encounter in written texts. As Strube and Müller (2003) point out, a big number of the pronouns used in spoken dialogue have non-noun phrase (NP) antecedents or no antecedents at all, which can prove to be a challenge for the semantic analysis of dialogue in AAL. The TRAINS93 corpus study of Byron and Allen (1998) shows that about 50% of the pronouns that are used in the corpus have antecedents that are non-NP-phrases. Thus, co-reference resolution for dialogue can highly benefit from the additional information that various modalities and more specifically gesture can provide.

## 3 Our Ambient Assisted Living Lab and the Intelligent Wheelchair

The Bremen Ambient Assisted Living Lab (BAALL) comprises all necessary conditions for trial living intended for two persons. This lab is a smart home suitable for the elderly and people with disabilities. It has an area of 60m<sup>2</sup> and contains all standard living areas, i.e. kitchen, bathroom, bedroom, and living room. It has intelligent adaptable household appliances and furniture for compensating for special limitations, e.g. separate kitchen cabinets can be moved up and down. The lab looks like a normal apartment and the technological infrastructure is discreet, if visible at all.

In the lab mobility assistance is provided through an *Intelligent Wheelchair* as well as an *Intelligent Walker*. For our studies we use the autonomous wheelchair/robot which is equipped with two laser range-sensors, wheel encoders, and an onboard computer; the wheelchair has a spoken dialogue

interface that allows to navigate to predefined destinations and to control devices in the lab.

The goal of the smart environment with mobility assistants and smart furniture is to evaluate new ambient assisted living technologies regarding their everyday usability. Users can interact through various interaction modes, such as a head joystick, a touch screen, and natural language dialogue.

In this paper we focus on the natural language dialogue and on contact-free, touchless, and not pen-based gestures in interaction with the wheelchair and smart furniture.

Figure 1 shows a smart appliance, i.e. the kitchen cabinet, which is moving down, so that it can be reachable for the wheelchair user.



Figure 1. Kitchen cabinets moving down

## 4 Speech Interaction in our Lab

Since the users of an AAL environment are typically untrained persons, elderly persons or persons possibly with physical or cognitive deficits, the user-centered analysis and adaptation of specific AAL-related application scenarios are necessary for developing a speech-enabled interactive dialogue system for our environment. In the following subsections we first describe the speech-related functionalities for the targeted users in our smart home (4.1) and then report on our recent work at the grammar level for improving the common problems caused by the automatic speech recognition (4.2).

### 4.1 Speech-related functionalities

According to the various assistance possibilities currently provided in our AAL environment, each

of the speech supportive functionalities can be classified based on the following three levels:

- An **explicit elementary action** on behalf of a simple dialogic utterance is used to ask for a specific assisting service to control each device in the AAL environment, such as “turn on the kitchen light”, “close the door of the bathroom” or “drive me to my bed”, etc.
- An **implicit composite action**, which can be uttered by simple or longer sentences, is used to converse with the dialogue system to trigger a set of explicit elementary actions regarding a predefined yet dynamically adaptable planning component. A typical utterance of such is “where is my pizza?”, which can then result in a sequence of actions including driving the user to the kitchen, opening all the doors on the path, showing the location of the pizza either orally or using other already implemented hardware supports (e.g. blinking light).
- A **context sensitive negotiating action**, which can be uttered during a clarification situation, should be used on the top of the explicit and implicit actions according to the situated context. Our AAL environment is in fact a multi-agent environment, which involves necessary dialogic interaction with other agents and their activities with respect to the possible temporal and spatial conflicts. For example, if a user wants to bake a pizza and the system detects that the oven is being used, the system would inform the user about it, then the user should be allowed to say “then take me to the oven when it’s available”.

In order to support the above three speech-enabled dialogic activities, a general dialogue system framework, the Diaspace Adaptive Information State Interaction Executive (DAISIE, cf. Ross & Bateman, 2009), is investigated and accordingly being extended.

DAISIE is a tightly coupled information state-based (see Larsson & Traum, 2000) dialogue backbone that fuses a formal language based dialogue controller, which provides a complex yet easily reusable plug-in mechanism for domain specific applications. Figure 2 depicts the general architecture of DAISIE.

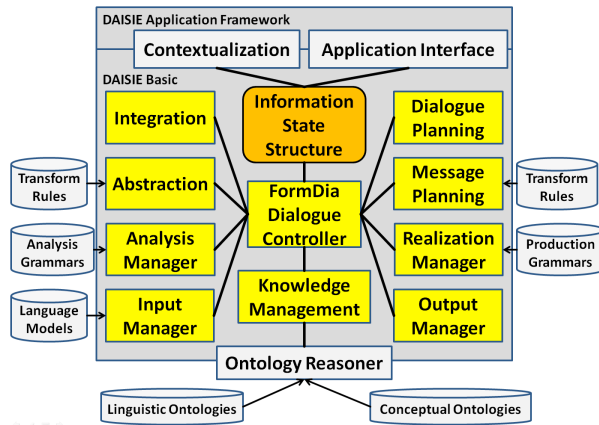


Figure 2. DAISIE with its plug-in components

According to the requirements of an operational dialogue system, the DAISIE architecture consists of a set of principle processing components, which are classified into the following three functionality groups:

- *The DAISIE Plug-ins* include a range of common language technology resources, such as speech recognizers/synthesizers, language parsers/generators, etc.
- *The DAISIE Basic System* integrates all the DAISIE plug-ins with the information state structure, knowledge monument component and the formal language based dialogue controller into a basic functional dialogue system.
- *The DAISIE Application Framework* specifies the application dependent components and provides a direct interface between the concrete domain application and the DAISIE basic system.

An instantiation of DAISIE is being developed and tested, which includes the implementation ranging over all levels of linguistic and conceptual representation and reasoning, as well as the adaptation of the current hybrid unified dialogue model (cf. Shi et al., 2011) to the AAL environment application regarding the listed three speech-related activities and possibly further modalities, such as gesture (see section 5).

## 4.2 Foot-syllable Grammar

Reliable speech recognition is a key factor for the success of a dialogue system in our AAL

environment. Currently, the expression stratum of the language system is modeled with two components in the *DAISIE* Framework: a speech recognizer for understanding spoken text and speech synthesizer for producing it. We use VoCon<sup>1</sup> as our speech recognizer, which takes a restriction grammar to know which commands the AAL and the wheelchair can undertake.

A foot-syllable restriction grammar was developed for optimized performance (Couto Vale & Mast, 2012). This grammar has a three-level structure starting at the lowest level with the syllable (S), an intermediate structure named *foot* (F), and the clause (C). A foot is a rhythmic unit in the compositional hierarchy of spoken language, which contains syllables as its parts and which is part of a curve (Halliday & Matthiessen, 2004). In German, it is composed by one stressed syllable and its adjacent unstressed ones. Below we present a segment of the grammar:

```

Foot-Syllable Grammar
<Foot1> : <IndIBegin> <kYStrong> <CEWeak> ;
<IndIBegin> : <InWeak> <dIWeak> | <IWeak> <nIWeak> ;
<kYStrong> : 'kY !pronounce("kY") ;
<CEWeak> : CE !pronounce("CE") | C$ !pronounce("C$") ;
<IWeak> : I !pronounce("I") | $ !pronounce("$") ;
<InWeak> : In !pronounce("In") | $n !pronounce("$n") ;
<nIWeak> : nI !pronounce("nI") | n$ !pronounce("n$") ;
<dIWeak> : dI !pronounce("dI") | d$ !pronounce("d$") ;

```

The foot-syllable grammar (FS) was contrasted with a foot-word (FW) and a phrase-word grammar (PW) in speech recognition effectiveness. The foot helped enforce corpus-based restrictions on syntactic structures and the syllable gave fine control over phonological variation. We conducted an evaluation study and our results have shown that, for complex highly flexible natural language dialogue situations such as human-robot interaction in AAL, a restriction grammar such as our foot-syllable grammar outperforms the other two approaches: 51,81% (FS) of correctly recognized utterances versus 26,67% (FW) and 6,67% (PW). We argue that using phonological units, such as syllables and foot units, provides a better way to achieve high recognition performance than phrases and words in both development cost and effectiveness.

<sup>1</sup> <http://www.nuance.com/for-business/by-product/automotive-products-services/vocon3200/index.htm>, 19/03/2012

## 5 Speech-Gesture User Study

A user study was conducted in our lab in November-December 2011. This user study included a real-life everyday scenario of a human user using a wheelchair to navigate in their environment by means of speech and gesture. The goal was to observe whether people would gesture and how, and what they would say if they used a wheelchair in their domestic environment. The study took place in BAALL and 20 German participants (students) took part in the study (mean age 25). Older users were not considered as participants in this study, as various tests, such as OsteoArthritis screening, neuropsychological tests, memory tests etc. would have to be taken in order to make sure that the elderly are physically able of performing gestures. Furthermore, it is difficult to bring seniors to the lab due to their physical condition. Elderly users might also be digitally intimidated by such technology. Although the tested group and the prospective user group are divergent, our user study primarily focuses on the collection of empirical gesture-speech data through the interaction of participants with technical devices in a smart home and thus does not distinguish between participants based on their age.

The participants were asked to act as if they were dependent on the wheelchair called *Rolland*. They had to navigate with *Rolland* to carry out daily activities (wash their teeth, eat something, read a book). They were informed in advance about the goal of the study, i.e. the collection of speech-gesture data and the video recording. The participants used a Bluetooth head-set and their activities were recorded by two digital IP cameras placed in BAALL, and also an SLR camera on *Rolland*'s back. Through audio and video streaming an experimenter (WoZ) selected through a GUI the destination point of *Rolland*. It is important to note again that we are interested into collecting various empirical spoken commands and gestures produced by the participants in their interaction with the wheelchair during the experimental run. Thus both the wheelchair navigation and *Rolland*'s speech feedback were WoZ-controlled. During most of the tasks the user was sitting on the wheelchair, but in one task the wheelchair drove autonomously without the user, as differences in gesture may change based on the

recipient (see discussion in Rimé & Schiaratura, 1991). Technical problems appeared in 8 sessions out of 20, when *Rolland* did not drive to the desired destination. The reasons for this are outside the scope of our research and of this paper. We evaluated 12 sessions regarding speech, but all 20 sessions regarding gesture.

As far as the results of this study are concerned, we collected 317 spoken commands in total. Many different language variants were uttered in order to carry out the same task. For example, four distinct utterances which were produced when participants were sitting on the bed and asked *Rolland* to come to them follow:

- i) “Rolland, komm her”  
(*Rolland, come here*)
- ii) “Rolland, <break9secs> Rolland,  
<break3secs> komm her”  
(*Rolland, Rolland, come here*)
- iii) “Rolland, komm bitte zum Bett, hier wo ich sitze”  
(*Rolland, please come to the bed, here where I am sitting*)
- iv) “Rolland, fahr zum Bett”  
(*Rolland, drive to the bed*)

Thus many context-sensitive utterances appeared in the collected data; for example, in the first two utterances above the participants did not use the name of a landmark (*bed*) in their command. Moreover, in the second example above we see that the participant waited for a backchannel feedback from the wheelchair and then uttered the actual command (*come here*).

From the study also the attitude, e.g. politeness, and expectations of the participants against the robot were measured. The style, volume of utterance, waiting time for the wheelchair to react as well as the sentence structure and lexical content were measurement factors. For example, male participants used more direct style with imperative sentences than female and included the name of their wheelchair in their command.

Concerning gestural frequency during the user study, in 7 sessions out of 20 participants employed at least one gesture during a session. In 6 sessions participants used deictic/pointing gestures and in 1 an iconic gesture (rubbing hands under the tap to represent washing hands). In 2 of the 7 sessions participants gestured more than once, while in the remaining 5 sessions, they gestured

once. The participants gestured mostly when something happened out of order, e.g. the wheelchair drove to a wrong place or stopped too far from the participant. Particularly in the bathroom, the wheelchair could not drive very close to the washbasin (predefined destination) and thus many participants gestured so that the wheelchair moves closer.

Two exceptions on gestural types and frequency were the following: in one case a male participant used often iconic gestures “for fun”, e.g. representing that he holds a gun. Another participant (female) gestured constantly using pointing gestures during all activities that she carried out. These cases can be attributed to personal influences, e.g. the user’s personality (see Rehm et al., 2008).

The gestures are annotated with the tool ANVIL (Kipp et al. 2007), a free video annotation tool that offers multi-layered annotation. The gesture annotation conventions for gesture, form, space, and trajectory, which are based on the practice of N. Furuyama (see McNeill, 2005: 273-278), are followed.

As far as co-reference is concerned, in all 317 commands there were several instances that the participants employed in their utterances. Yet, all of them were either references to the participant him/herself or to the wheelchair:

- i) “*Rolland, drehe dich bitte um*”  
(*Rolland, turn around please*)
- ii) “*Fahr mich bitte zum Badezimmer*”  
(*Drive me to the bathroom please*)

For those cases, the participants did not use gestures. We assume that the rare use of co-reference in our user study is due to the fact that participants almost never had to refer to the same entity again. Once a command was uttered, the WoZ executed the required action and the participants could move further to the next task they had in their agendas. Thus, once an entity (i.e. the sofa, the bathroom, the kitchen) has been introduced, the participants never needed to refer back to that same entity again.

Last but not least, nobody of the participants realized that the experiment was WoZ, believing that the wheelchair moved based on their own commands.

## 6 Conclusion and Future Work

Our lab is equipped with intelligent adaptable household appliances and furniture for compensating for special limitations; this lab can be used as an experiment area for many user studies with different purposes.

In the presented user study we collected empirical speech and gesture data of natural dialogue in HRI. By making the speech-gesture interaction between users and robot more natural, intuitive, effective, efficient, and user-friendly, assistive environments will become more appropriate in the real world used by seniors and/or *seniors to be*, people actively planning their future.

A planned second user study handles selection and control of objects in a smart environment. In this study objects such as television, lights, electronic sliding doors, etc. will be remotely controlled by the experimenter (WoZ). The participants will be requested to select objects and control their position and level (higher, louder, etc.). A condition tested here will be the presence/lack of ambiguity. Participants will be asked to “open a door” or “turn on a light” having many doors and lights available in the lab. In addition, the wheelchair will be intentionally driven by an experimenter to a wrong destination or stopped on its way to a destination point. This adjustment has been made considering the results in the conducted study that participants gestured more when something went wrong.

A third study is planned in order to identify which spatial gestures are universal and which are *locale-dependent*. Within the field of localization, *locale* is a combination of language and culture. The criteria of *locale* selection are countries with i) big geographic distance, ii) strong cultural differences, iii) diversity of gestures based on literature evidence, and iv) typologically different languages. This study is necessary to investigate the differences in speech-gesture interaction between the German and other *locales*.

Last but not least, a small scale follow-up study with elderly people will take place in a nursing home. There the elderly could be requested to perform gestures that have already been collected in our previous studies in order to evaluate them depending on their skills and preferences.

The collected data from the user studies stored in a corpus will be examined based on the speech-

gesture alignment concerning their semantics, their temporal arrangement, and their coordinated organization in the phrasal structure. Later an extension for the semantics of gesture types will be added to the Generalized Upper Model (GUM) in order to anchor spatial gestures into a semantic spatial representation. GUM (Bateman et al., 2010) is a linguistically motivated ontology for the semantics of spatial language of German and English. New GUM categories will be created for gestures, when the linguistic ones are not applicable and/or sufficient.

## Acknowledgments

We gratefully acknowledge the support of the Deutsche Forschungsgemeinschaft (DFG) through the Collaborative Research Center SFB/TR 8 Spatial Cognition. We also thank Daniel Vale, Bernd Gersdorf, Thora Tenbrink, Carsten Gendorf, and Vivien Mast for their help with the user study.

## References

- M. Alibali. 2005. Gesture in spatial cognition: expressing, communicating, and thinking about spatial information. *Spatial Cognition and Computation*, 5(4):307-331.
- J. Bateman, J. Hois, R. Ross, and T. Tenbrink. 2010. A linguistic ontology of space for natural language processing. *Artificial Intelligence*, 174(14): 1027-1071.
- E. Becker, Z. Le, K. Park, Y. Lin and F. Makedon. 2009. Event-based experiments in an assistive environment using wireless sensor networks and voice recognition. *Proceedings of the International conference on Pervasive technologies for assistive environments (PETRA)*.
- D.K. Byron and J.F. Allen. 1998. Resolving demonstrative pronouns in the TRAINS93 corpus. *New Approaches to Discourse Anaphora: Proceedings of the 2<sup>nd</sup> Colloquium on Discourse Anaphora and Anaphor Resolution (DAARC2)*, 68-81.
- L. Chen, A. Wang and B. Di Eugenio, B. 2011. Improving pronominal and deictic co-reference resolution with multi-modal features. *Proceedings of the SIGDIAL Conference*, 307-311.
- N. Chovil. 1992. Discourse-oriented facial displays in conversation. *Research on Language and Social Interaction*, 25:163-194.
- D. Couto Vale and V. Mast. 2012. Customizing Speech Recognizers for Situated Dialogue Systems. *Proceedings of the 15th International Conference on Text, Speech and Dialogue*.
- J. Eisenstein and R. Davis. 2006. Gesture improves coreference resolution. *Proceedings of the Human Language Technology Conference of the NAACL*, 37-40.
- E. Fricke. 2009. Multimodal attribution: How gestures are syntactically integrated into spoken language. *Proceedings of GESPIN: Gesture and Speech in Interaction*.
- V. Fuchsberger. 2008. Ambient assisted living: elderly people's needs and how to face them. *Proceedings of the 1<sup>st</sup> ACM International Workshop on Semantic Ambient Media Experiences*, 21-24.
- S. Goetze, N. Moritz, J.E. Appell, M. Meis, C. Bartsch and J. Bitzer. 2010. Acoustic user interfaces for ambient-assisted living technologies. *Inform Health Soc Care*, 35(3-4):125-143.
- S. Goldin-Meadow. 2003. *Hearing gesture: How our hands help us think*. Cambridge, MA: Harvard University Press.
- F. Hahn and H. Rieser. 2010. Explaining speech gesture alignment in MM dialogue using gesture typology. P. Lupowski and M. Purver (Eds.), *Proceedings of the 14<sup>th</sup> Workshop on the Semantics and Pragmatics of Dialogue (SemDial)*, 99-111.
- F.M.A.K. Halliday and C.M.I.M. Matthiessen. 2004. *An introduction to functional grammar*. 3<sup>rd</sup> Edition. Edward Arnold, London.
- A. Hostetter and M. Alibali. 2005. Raise your hand if you're spatial—Relations between verbal and spatial skills and gesture production. *Gesture*, 7(1): 73-95.
- J. Ivanecky, S. Mehlhase and M. Mieskes, M. 2011. *An Intelligent House Control Using Speech Recognition with Integrated Localization*. R. Wichert and B. Eberhardt, B. (Eds.), 4. AAL Kongress. Berlin, Germany.



- A. Jaimes and N. Sebe. 2007. Multimodal human-computer interaction: A Survey, Computational Vision and Image Understanding. Elsevier Science Inc., New York, USA, 116-134.
- C. Jian, F. Schafmeister, C. Rachuy, N. Sasse, H. Shi, H. Schmidt and N.v. Steinbüchel. 2012. Evaluating a Spoken Language Interface of a Multimodal Interactive Guidance System for Elderly Persons. Proceedings of the International Conference on Health Informatics.
- A. Kendon. 2004. Gesture: Visible action as utterance. Cambridge: Cambridge University Press.
- S. Kita. 2000. How representational gestures help speaking. McNeill, D. (Ed.), Language and gesture, Cambridge, UK: Cambridge University Press, 162-185.
- S. Kita. 2009. Cross-cultural variation of speech-accompanying gesture: A review. Language and Cognitive Processes, 24(2): 145-167.
- M. Kipp, M. Neff and I. Albrecht. 2007. An annotation scheme for conversational gestures: How to economically capture timing and form. Language Resources and Evaluation Journal, 41: 325-339.
- S. Kopp. 2005. The spatial specificity of iconic gestures. Proceedings of the 7<sup>th</sup> International Conference of the German Cognitive Science Society, 112-117.
- J. Krajewski, R., Wieland and A. Batliner. 2008. An acoustic Framework for detecting Fatigue in Speech based Human Computer Interaction. Proceedings of the 11<sup>th</sup> International Conference on Computers Help People with Special Needs, 54-61.
- S. Larsson and D. Traum. 2000. Information State and Dialogue Management in the TRINDI Dialogue Move Engine Toolkit. Natural Language Engineering. Special Issue on Best Practice in Spoken Language Dialogue Systems Engineering, 323-340.
- A. Marinc, C. Stocklów and S. Tazari, S. 2012. 3D Interaktion in AAL Umgebungen basierend auf Ontologien. Proceedings of AAL Kongress.
- D. McNeill. 1992. Hand and Mind: What Gestures reveal about Thought. University of Chicago Press.
- D. McNeill. 2000. Introduction. McNeill, D. (Ed.), Language and gesture. Cambridge: Cambridge University Press.
- D. McNeill. 2005. Gestures and Thought. University of Chicago Press.
- K. Nazemi, D. Burkhardt, C. Stab, M. Breyer, R. Wichert and D.W. Fellner. 2011. Natural gesture interaction with accelerometer-based devices in ambient assisted environments. R. Wichert and B. Eberhardt, B. (Eds.), 4. AAL-Kongress, Springer, 75-84.
- R. Neßelrath, C. Lu, C.H. Schulz, J., Frey, and J. Alexandersson. 2011. A gesture based system for context-sensitive interaction with smart homes. R. Wichert and B. Eberhardt, B. (Eds.), 4. AAL-Kongress, 209-222.
- S. T. Oviatt. 1999. Ten myths of multimodal interaction. Communications of the ACM. ACM New York, USA, 42(11): 74-81.
- F.H. Rauscher, R.M. Krauss and Y. Chen. 1996. Gesture, speech, and lexical access: The role of lexical movements in speech production. Psychological Science, 7: 226-230.
- M. Rehm, N. Bee and E., André. 2008. Wave like an Egyptian: accelerometer-based gesture recognition for culture specific interactions. Proceedings of the 2<sup>nd</sup> British HCI Group Annual Conference on People and Computers: Culture, Creativity, Interaction, 1:13-22.
- B. Rimé and L. Schiaratura. 1991. Gesture and speech. Fundamentals of nonverbal behavior. Studies in emotion & social interaction, 239-281.
- J. R. Ross and J. Bateman. 2009. Daisie: Information State Dialogues for Situated Systems. Proceedings of Text, Speech and Dialogue, 5729/2009, 379-386.
- H. Shi, C. Jian and C. Rachuy. 2011. Evaluation of a Unified Dialogue Model for Human-Computer Interaction. International Journal of Computational Linguistics and Applications, 2.
- H. Steg, H. Strese, C. Loroff, J. Hull and S. Schmidt. 2006. Europe is facing a demographic

challenge ambient assisted living offers solutions.  
VDI/VDE/IT, Berlin, Germany.

- M. Strube and C. Müller. 2003. A machine learning approach to pronoun resolution in spoken dialogue. Proceedings of the 41<sup>st</sup> Annual Meeting on Association for Computational Linguistics, 1:168-175.
- S. Takahashi, T. Morimoto, S. Maeda and N. Tsuruta. 2003. Dialogue experiment for elderly people in home health care system. Proceedings of the 6<sup>th</sup> International Conference on Text, Speech and Dialogue, 418-423.
- R. Wichert and B. Eberhardt, B. (Eds.). 2011. Ambient Assisted Living. 4. AAL-Kongress, Springer.