# Better Choice? Combining Speech and Touch in Multimodal Interaction for Elderly Persons

Cui Jian[1], Hui Shi[1], Nadine Sasse[2], Carsten Rachuy[1], Frank Schafmeister[2], Holger Schmidt[3] and Nicole von Steinbüchel[2]

[1]*SFB/TR8 Spatial Cognition, University of Bremen, Enrique-Schmidt-Straße 5, Bremen, Germany*

[2]*Medical Psychology and Medical Sociology, University Medical Center Göttingen, Waldweg 37, Göttingen, Germany*

[3]*Neurology, University Medical Center Göttingen, Robert-Koch-Str. 40, Göttingen, Germany*

Keywords: ICT, Ageing and Disability, Elderly-centered Design, Multimodal Interaction, Elderly-friendly Interface, Formal Methods, System Evaluation.

Abstract: This paper presents our work on developing, implementing and evaluating a multimodal interactive guidance system that features spoken language and touch-screen input for elderly persons. The development foundation of the system comprises two systematically designed and empirically improved aspects: a set of development guidelines for elderly-friendly multimodal interaction according to common ageing-related decline of important human abilities, and a hybrid dialogue modelling approach with a formal method triggering and agent-based management for the elderly-centered multimodal interaction. To evaluate the minutely developed and implemented system, an experimental study was conducted with thirty-three elderly persons and empirical data were analyzed by applying an adapted version of a general evaluation framework, which provided overall positive analysis results and validated our effort to develop an effective, efficient and elderly friendly multimodal interaction.

## 1 INTRODUCTION

As the demographic development shows, the amount of elderly people is constantly growing in modern societies (Lutz et al., 2008). These persons often suffer from age-related decline or impairment of sensory, perceptual, physical and cognitive abilities. This poses particular challenges to the application of technical systems nowadays, which are getting more and more commonly implemented in the daily routines of elderly persons.

Meantime, attention is increasingly focused on the technical systems with multimodal interfaces, which provide the users with multiple modes of interaction with a system; therefore they improve the quality of human-system communication concerning effectiveness, efficiency and user-friendliness (cf (Jaimes and Sebe, 2007)).

Thus, in order to maximize the usability and user experience of technical systems for elderly persons, research on multimodal interaction for this specific user group is increasingly gaining more interest during the last decade. Various emerging technologies have been considered, such as advanced speech enabled interface (Krajewski et al., 2008), brain-signal interface (Mandel et al., 2009), visual input via digital camera (Goetze et al., 2012); also, a large contribution is being made to "Ambient Assistive Living", the concept for developing age-adjusted and care-friendly living environments (cf. (Rodríguez et al., 2011)).

This paper presents a multimodal interactive system that can provide elderly persons with both spoken language and touch-screen input modalities. It has been particularly developed and implemented for the elderly focussing on two important aspects: 1) a set of development guidelines for multimodal interactive systems with respect to the basic design principles of conventional interactive systems and the most common age-related characteristics; and 2) a hybrid dialogue modelling and management approach that combines the advanced finite state based generalized dialogue model and the classic agent based dialogue theory; it supports a flexible and context-sensitive, yet tractable and controllable multimodal interaction with a formal language based



Jian C., Shi H., Sasse N., Rachuy C., Schafmeister F., Schmidt H. and von Steinbüchel N..
Better Choice? Combining Speech and Touch in Multimodal Interaction for Elderly Persons .
DOI: 10.5220/0004244800940103
In *Proceedings of the International Conference on Health Informatics* (HEALTHINF-2013), pages 94-103
ISBN: 978-989-8565-37-2

development framework. The resulting system has been continuously improved by a series of evaluative studies of our previous work concerning the development foundation and different modalities (cf. (Jian et al., 2011), (Jian et al., 2012)). In order to perform a further evaluation of spoken language and touch-screen combining input modalities, as well as the assessment of the complete multimodal interactive system concerning its effectiveness of task performance, efficiency of interface interaction and user acceptance by elderly persons, a supplementary experimental study was conducted with 33 elderly. The data were analysed by applying an adapted version of the general evaluation framework PARADISE (Walker et al., 1997). The Results are briefly described in this paper.

The rest of the paper is organized as follows: section 2 briefly introduces the design guidelines for multimodal interactive systems for elderly persons and the hybrid approach for multimodal interaction management; section 3 presents the multimodal interactive guidance system and section 4 describes the experimental study on evaluating the modality combining spoken language and touch-screen; the results are analysed and discussed in section 5. Finally, section 6 concludes the reported work and outlines the direction of our future activities.

# 2 THE DEVELOPMENT FOUNDATION

The theoretical and technical foundation of our work comprises two aspects:

- A set of design guidelines for an elderly-centered multimodal interactive system;
- A formal method and agent based hybrid modelling approach for dialogue management.

They are both systematically designed with respect to their suitability for the application, and continuously improved by our previous empirical studies (cf. (Jian et al., 2012)).

## 2.1 Design Guidelines of Multimodal Interactive Systems for Elderly Persons

Physical and cognitive decline is almost universal in the elderly. According to (Birdi et al., 1997), these age-related characteristics should be considered while developing interactive systems for the elderly. Therefore, based on the common design principles for conventional interactive systems and the ageing-related empirical findings, we defined a set of design guidelines for multimodal interaction with respect to the decline of important human perceptual and cognitive functions. These guidelines have been implemented into our multimodal interactive guidance system, evaluated by our previous empirical studies, and then improved on the basis of their results.

The final set of improved design guidelines were summarized in (Jian et al., 2012). For reasons of brevity we report empirical findings regarding the decline of the seven most common human abilities, accordingly followed by the most important elements implemented and improved during our system development:

**Visual Perception** worsens for most people with age (Fozard, 1990). Physically the size of the visual field is decreasing and the peripheral vision can be lost. It is more difficult to focus on objects up close and to see fine details, including rich colours and complex shapes that make images hard or even impossible to identify. Rapidly moving objects are either causing too much distraction, and/or become less noticeable. This decline concerns most with the graphical user interface. Based on the suggested guidelines, only simple and clear layout was constructed without overlapping items; 12-14 sized sans-serif fonts were chosen for all displayed texts. Simple and high contrast colours without fancy visual effects were used and placed aside; regularly shaped rectangles and circles were selected for comfortable perception and easy identification.

**Speech Ability** declines while ageing in the way of being less efficient for pronouncing complex words or longer sentences, probably due to reduced motor control of tongue and lips (Mackay and Abrams, 1996). (Moeller et al., 2008) confirmed that, elderly-centered adaptation of speech-enabled interactive components can improve the interaction quality to a satisfactory level. Therefore, the vocabulary and grammar for our speech recognizer and analyser were constructed with preferably short and easy wording in daily life communication; dialogue strategies were also adjusted to many elderly-specific situations.

**Auditory Perception** declines to 75% between the age of 75 and 79 year olds (cf. Kline and Scialfa, 1996). High pitched sounds are hard to perceive; complex sentences are difficult to follow (Schieber, 1992). Therefore, text and acoustic output were both provided as system responses. Style, vocabulary and structure of the sentences were intensively revisedregarding brevity. A low-pitched yetvigorous male voice was used for the speech synthesis.

**Motor Abilities** decline generally due to loss of physical activities while ageing. Complex fine motor activities are more difficult to perform, e.g. to grab small or irregular targets (cf. Charness and Bosman, 1990); conventional input devices such as a computer mouse are less preferred by elderly persons as good hand-eye coordination is required (Walkder et al., 1997). Taking these findings into account, a touch screen was chosen as the haptic interface; Regularly shaped, sufficiently sized and well separated interface elements were constructed; pressing instead of clicking or dragging was decided to be the only action in order to avoid otherwise frequently occurring errors.

**Attention and Concentration** drop while ageing. Elderly persons either become more easily distracted by details and noise, or find other things harder to notice when concentrating on one thing (Kotary and Hoyer, 1995); they show great difficulty with situations where divided attention is needed (McDowd and Craik, 1988). Thus, fancy irrelevant images or decorations were removed. Unified font, colours, sizes of interface elements were used throughout the interaction. Simple animations for notifying changes were constructed, giving sufficiently clear feedback to the user.

**Memory Functions** decline differently. Short term memory holds fewer items with age and working memory becomes less efficient (Salthouse, 1994). Semantic information is normally preserved in long term memory (Craik and Jennings, 1992). Guided by these facts, the quantity of displayed items was restricted to no more than three, regarding the average capacity of short term memory of elderly persons; sequentially presented items were intensively revised to assist orientation during interaction. Context sensitive cues were presented with selected colours: green for items concerning persons, yellow for items concerning rooms, etc.

**Intellectual Reasoning Ability** does not decline much during the normal ageing process. (Hawthorn, 2000) believed that crystallized intelligence can assist elderly persons to perform better in a stable well-known interface environment. Therefore, consistent layout, colours and interaction styles were used throughout the interaction. Changes on the interface can only happen on data level.

## 2.2 The Hybrid Approach for Interaction Management

The hybrid dialogue modelling approach combines the finite-state-based generalized dialogue models with the classic agent-based dialogue management theories. This section

- introduces the basic theory of this approach,
- presents the adapted instance model for multimodal interaction in elderly persons by applying the hybrid approach;
- describes a formal language based development toolkit, which is then used to support the implementation of the instance model and its integration into our multimodal interactive guidance system for achieving an effective, flexible, yet formally controllable multimodal interaction management.

### 2.2.1 The Theory

The development of the hybrid dialogue modelling approach benefited from existing researches on these two important interaction management theories:

The **generalized dialogue models**, which are constructed with recursive transition networks (RTN) at the illocutionary level. These networks can abstract dialogue models by describing discourse patterns as illocutionary acts, without reference to any direct surface indicators (cf. (Alston, 2000));

The **classic agent-based management** method: information state update based management theories (cf. (Traum and Larsson, 2003)), which focus on the modelling of discourse context as the attitudinal state of an intelligent agent. This method shows a powerful way to handle dynamic information for a context sensitive dialogue management.

However, these two well-accepted methods have their own limitations. On one hand, the generalized dialogue models are based on finite state transition models, which are criticized for their inflexibility of dealing with dynamic information exchange; on the other hand, the information state update models are usually very difficult to manage and extend.

Therefore, we designed a hybrid dialogue modelling approach by extending the generalized dialogue model with conditions and information state update rules added into finite-state transitions.

### 2.2.2 The Interaction Model

In order to manage multimodal interaction for elderly persons, an adapted hybrid dialogue model was constructed and evaluated by our previous work (cf. (Shi et al., 2011)). The accordingly improved version consists of four hybrid dialogue schemas: the initiating schema, the user's action schema, the system's response schema and the user's response schema, regarding the four general transitions during interaction.

Each interaction is initiated with the schema *Dialogue(S, U)* (cf. Figure 1), by the initialization of the system's start state and a greeting request. In Dialogue(S, U) the system initiates a dialogue with a request move (i.e. *S.request*), which cause the initialization of the dialogue context using the update rule INIT.

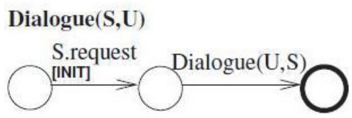


Figure 1: The initiating model.

The dialogue continues with the user's instruction, request for a certain information or restart action, leading to the system's further response or dialogue restart, respectively, while updating the information state with the attached update rules (cf. *Dialogue(U, S)* in figure 2).

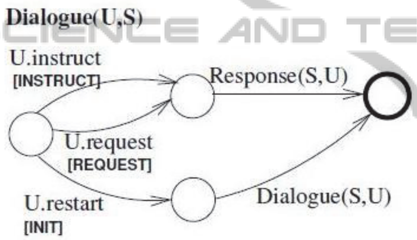


Figure 2: The user's actions model.

After receiving user input, the system tries to generate an appropriate response according to its current knowledge base and information state (cf. *Response(S, U)* in figure 3). This can be informing the user with requested data, rejecting an unacceptable request with or without certain reasons, providing choices for multiple options, or asking for further confirmation of taking a critical action, each of which triggers transitions to other hybrid models.

Finally, the user can accept or reject the system's response, or even ignore it by simply providing new instructions or requests, triggering further state transitions as well as information state updates (cf. *Response(U, S)* in Figure 4).

Besides the improvement performed with respect to the specific interaction data of the elderly subjects in our previous studies, the decline of physical and cognitive abilities of elderly persons, especially memory function, concentration and fluid reasoning ability should be considered as well. Therefore, for the improvement of the current hybrid dialogue model we also included the following features to assist the elderly during the multimodal interaction:

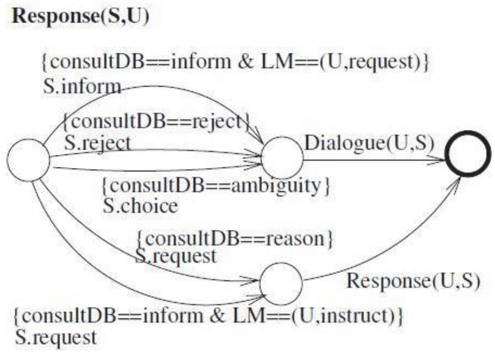


Figure 3: The system's response model.

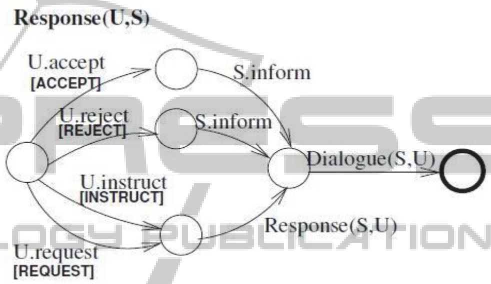


Figure 4: The user's response model.

- Relevant dialogue history information, such as the latest utterance, was added into the current information state and provided in case of speech recognition problems.
- Context sensitive information, which is kept in the current information state, is designed to be either directly presented after each interaction pace, or included within dialogue utterance, in order to ease the common problems caused by the declining memory function.
- Additional context information is provided with specific information state update rules in extreme cases, e.g. if the automatic speech recognition problems become too interfering, messages containing possibly recognized context will be presented.
- Instead of keeping rich transition alternatives at the illocutionary level, the hybrid model was kept as compact and intuitive as possible.

### 2.2.3 A Development Framework to Support the Hybrid Dialogue Modelling Approach

The structure of a hybrid dialogue model is in fact a typical finite state transition model. This feature enables any hybrid dialogue model to be formally specified as a set of machine readable codes, e.g.

using mathematically well-founded formal language, Communicating Sequential Processes (CSP) (cf. (Roscoe, 1997)) in the formal methods and computer science community. Furthermore, the CSP program is also supported by well-established model checkers, which provides the rich possibilities of validating the concurrent aspects and increasing the tractability of the specified model (cf. (Hall, 2002)).

Thus, in order to support the development of hybrid dialogue models using the formal language CSP and its integration into a practical multimodal interactive system, we designed FormDia, the Formal Dialogue Development Toolkit (cf. Figure 5).

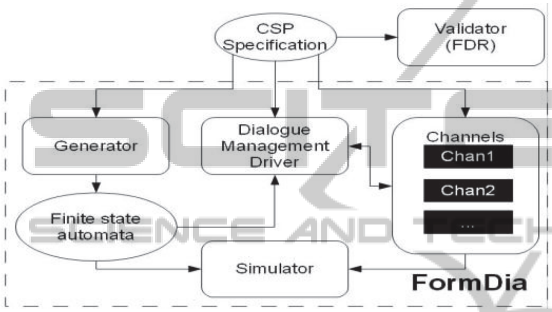


Figure 5: The Structure of the FormDia Toolkit.

Theoretical and technical details about FormDia can be found in (Shi and Bateman, 2005). In general, the FormDia Toolkit supports the implementation and integration of a hybrid dialogue model into an interactive system with the four components:

**Validator**: after a hybrid dialogue model is specified with CSP, it can be validated by an external model checker: the Failures-Divergence Refinement tool or FDR (Broadfoot and Roscoe, 2000), for validating and verifying concurrency of state automata.

**Generator**: with the validated CSP specification, machine readable finite state automata can be generated by the Generator.

**Simulator**: with the generated finite state automata and the communication channels, dialogues scenarios can be simulated via a graphical interface, which visualizes dialogue states as a directed graph and provides a set of utilities for primary testing.

**Dialogue Management Driver**: finally the dialogue model is integrated into an interactive system via the dialogue management driver.

Therefore, FormDia enables an intuitive design of hybrid dialogue models with formal language, automatic validation of the related functional properties, easy simulation and verification of specified interaction situations, and a straightforward integration into a practical interactive system.

# 3 SYSTEM DESCRIPTION

Based on the development foundation introduced in the previous section, we developed a general Multimodal Interactive Guidance System for Elderly Persons (MIGSEP).

## 3.1 System Introduction

MIGSEP runs on a portable touch-screen tablet PC and will serve as the interactive media, which is intended to be used by an elderly or handicapped person seated in an autonomous electronic wheelchair that can automatically carry its users to desired locations within complex environments. The user should interact with MIGSEP with spoken-language and touch-screen combining input modality to find the desired target.

## 3.2 System Architecture

The architecture of MIGSEP is illustrated in figure 6. The *Generalized Dialogue Manager* was developed using the introduced adapted hybrid dialogue model and the FormDia toolkit. It functions as the central processing unit of the entire system and supports a formally controllable and extensible, meantime flexible and context-sensitive multimodal interaction management. An *Input Manager* receives and interprets all incoming messages from the *GUI Action Recognizer* for GUI input events, the *Speech Recognizer* for natural language understanding and the *Sensing Manager* for other possible sensor data. An *Output Manager* on the other hand, handles all outgoing commands and distributes them to the *View Presenter* for presenting visual feedbacks, the *Speech Synthesizer* to generate natural language responses and the *Action Actuator* to perform necessary motor actions, such as sending a driving request to the autonomous electronic wheelchair. The *Knowledge Manager*, constantly connected with the *Generalized Dialogue Manager,* uses a *Database* to keep the static data of certain environments and the *Context* to process the dynamic information exchanged with the users during the interaction.

All components of MIGSEP are closely connected via XML-based communication channels and each component can be treated as an open black box which can be accordingly modified or extended for concrete domain specific use, without affecting other components in the MIGSEP architecture. It provides a general open platform for both theoretical researches and empirical studies on single- or multimodal interaction that can relate to different application domains and scenarios.

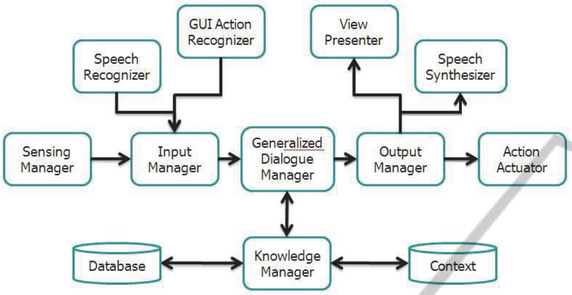


Figure 6: The architecture of MIGSEP.

### 3.3 Interaction with the System

The current instance of MIGSEP was implemented as a guidance system used by elderly persons for the application domain of hospital environments. Figure 7 shows a user interacting with MIGSEP.

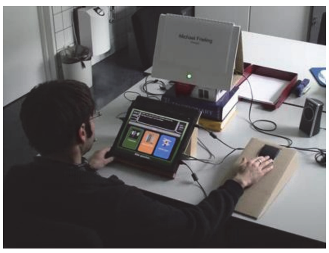

Figure 7: A user is interacting with MIGSEP.

This MIGSEP system consists of a button device for triggering a “press to talk” signal, a green lamp to signalize the “being pressed and ready to talk” state, and the tablet PC, on which the MIGSEP system is running and the interface is displayed. The MIGSEP interface simply consists of two areas:

**Function-area** contains the function button “start” on the top left for going to the start state, the function button “toilet” below it regards the basic needs of elderly persons, and the text area for displaying the system responses in the middle;

**Choice-area** displays information entities as single cards that can be selected, with a scrollbar indicating the position of the current displayed cards and a context sensitive coloured bar showing the current concerned context if necessary.

Figure 8 shows a sample of spoken language and touch-screen combined interaction dialogue between MIGSEP and a user who would like go to the cardiology department, to a doctor named Wolf.

**S: <shows 3 cards, green for persons, yellow for rooms and blue for departments>**
**hello, where do you want to go?**
**U: I want to go to Wolf in the cardiology department.**
**S: <shows a blue bar texted with cardilogy on the top, and 2 green cards, with photos, titles and names in the middle>**
**I found two persons working in the cardiology department with the name Wolf, to which do you want to go?**
**U: <presses the second card>**
**S: <enlarges the second green card, shows yes/no button>**
**Do you want to go to Mrs. Dr. Diana Wolf?**
**U: <press the yes button>**
**S: <resizes the second green card and shows the 3 cards in the start menu again>**
**OK, I have saved this goal. Where else do you want to go?**

Figure 8: A sample interaction with MIGSEP.

## 4 EXPERIMENTAL STUDY

To evaluate how well the MIGSEP system can assist elderly persons by using a modality combining spoken language and touch-screen, an experimental study was conducted with the department of Medical Psychology and Medical Sociology in Göttingen.

### 4.1 Participants

33 elderly persons (m/f: 19/14, mean age of 70.7, standard deviation 3.1), all German native speakers, participated in the study. They all had to pass the mini-mental state examination (MMSE), a screening test to assess the cognitive mental status (cf. (Folstein et al., 1975)). A test score between 28 and 30 indicates slightest decline versus normal cognitive functioning. Our participants showed an average score of 28.9 (std.=.83).

### 4.2 Stimuli and Apparatus

Visual stimuli were presented via a green lamp and a graphical user interface on the screen of a portable tablet PC; audio stimuli were also generated by the MIGSEP system and played via two loudspeakers at a well-perceivable volume. All tasks were given as keywords on the pages of a calendar-like system.

There were two types of input possibilities, which could be freely chosen: the spoken language, activated if the button was being pressed and the green lamp was on; and the touch-screen action, directly performable on the touch-screen display.

The same data set contains virtual yet sufficient information about personnel, rooms and departments in a common hospital, was used in the experiment.

During the experiment each participant was accompanied by the same investigator, who introduced the system and gave well-defined instructions at the beginning, and provided help if necessary during the trail (which was very rare).

An automatic internal logger of the MIGSEP system was used to collect the real-time system internal data, while the windows standard audio recorder program kept track of the whole dialogic interaction process.

A questionnaire, focusing on the user satisfaction with MIGSEP with respect to the spoken language and touch-screen combining input modality, was especially designed for this study. It contains 6 questions concerning the quality of the combined modality compared to a single modality, the feasibility, the advantages, the usability, the appropriateness and the preference. This questionnaire was answered by each participant via a five point Likert scale.

### 4.3 Procedure

Each participant had to undergo four phases:

**Introduction**: a brief introduction was given to the participants, so that they could get the basic idea and an overview of the experiment.

**Learning and Pre-tests**: the participants were instructed how to interact with MIGSEP using the spoken natural language and the touch-screen input. In order to minimize the learning or bias effect with respect to the use of one modality, we introduced a cross-over procedure, 16 participants out of 33 had to first use the touch-screen input and then the spoken language, the other 17 used spoken language first and then the touch-screen input. All of the participants had to perform 11 tasks concerning their navigation procedures in a hospital in order to reach a certain aim. Each modality and each task contained incomplete yet sufficient information about a destination the participant should select. For example they had to drive to “room 2603”, to “Sonja Friedrich”, or to “room 1206 or room 2206 with the name OCT-Diagnostics”. Tasks were fulfilled or ended, if the goal was selected or the participant gave up trying after six minutes.

**Testing**: After performing 22 tasks with both modalities, each participant was asked to freely choose between spoken language and touch-screen input modality to perform again 11 tasks; they contained similar information as in pre-tests (varied only on data level) and were performed under the same conditions.

**Evaluation**: After all tasks were run through, each participant was asked to fill in the questionnaire for evaluation.

## 5 RESULTS AND ANALYSIS

According to the Paradise framework (Walker, et al., 1997), the performance of an interactive system can be measured via the effectiveness, efficiency of the system and the user satisfaction. Therefore, these three aspects were analysed.

### 5.1 Effectiveness of the System

To find out how effective the elderly were assisted by the MIGSEP system with the combining modality, statistical method “Kappa coefficient” is used. However, in the classic Paradise framework the Kappa method was originally used to evaluate a spoken dialogue system.

Therefore, in order to be able to calculate the Kappa coefficient with respect to the multimodal interaction with the MIGSEP system, we first had to develop an adaptation of the original attribute value matrix, which still contained all information that was exchanged during the multimodal interaction between MIGSEP and participants. For this reason, we introduce the concept of an Attribute Value Tree (AVT) (cf. the example in Figure 9).

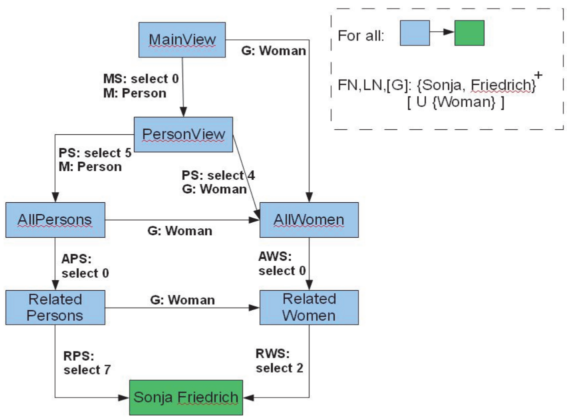


Figure 9: An Attribute Value Tree.

An AVT is defined as a finite state transition diagram, which contains all the expected correct way, either touch-screen input or spoken language command, as the state transitions from the start state to the target state. As the AVT for the task “go to a person named Sonja Friedrich” illustrated in Figure 9, any correct interaction should go from the state mainView, then e.g. to PersonView by selecting the first card (MS: select 0), or performing the spoken language command “I want to go to a person” (M: Person), or go to AllWomen state by simply saying “I want to go to a woman” from the MainView, etc.

An AVT contains the expected data set of a task, and therefore functions similarly as the original attribute value matrix, yet with the possibility of recording multimodal interaction exchange.

Thus, 11 AVTs were created for the 11 tasks respectively and by combining the actual data recorded during the experiment with the expected attribute values in the AVTs, we can construct the confusion matrices for all tasks. E.g., table 1 shows the confusion matrix for the task ”drive to a person named Sonja Friedrich”, where ”M” and ”N” denote whether the actual data match with the expected attribute values in the AVTs. E.g. there were 25 correctly selected actions in the PersonSelect (PS) state; and the spoken language command regarding the first name (FN) was misrecognized by the system for 6 times. Note that, because of the width of the text, not all attributes of this confusion matrix can be shown in this example.

Table 1: The confusion matrix for the task “drive to a person named Sonja Friedrich”.

| | PS | | MS | | ... | | FN | | |
|---|---|---|---|---|---|---|---|---|---|
| Data | M | N | M | N | M | N | M | N | sum |
| PS | 25 | 0 | | | | | | | 25 |
| MS | | | 14 | 0 | | | | | 14 |
| ... | | | | | ... | ... | | | ... |
| FN | | | | | | | 62 | 6 | 68 |

The data for all confusion matrices were merged and a total confusion matrix for all the data of the 11 performed tasks was created.

Given the total confusion matrix, the Kappa coefficient was calculated with

$$\kappa = \frac{P(A)-P(E)}{1-P(E)}, \text{ (Walker, et al., 1997)}$$

In our experiment, $P(A) = \frac{\sum_{i=1}^{n} M(i,\ M)}{T}$ is the proportion of times that the actual data agree with the attribute values, and $P(E) = \sum_{i=1}^{n}(\frac{M(i)}{T})^2$ is the proportion of times that the actual data are expected to be agreed on by chance, where M(i, M) is the value of the matched cell of row i, M(i) the sum of the cells of row i, and T the sum of all cells.

Thus, we could calculate the Kappa Coefficient of the total confusion matrix $\kappa=0.91$, suggesting a highly successful degree of interaction between the MIGSEP and the participants using the spoken language and touch-screen combining modality.

## 5.2 Efficiency of the System

In order to find out how efficiently the participants were assisted using the combining modality, quantitative data of every single interaction during the testing were automatically logged. Results are summarized in Table 2, with respect to four important aspects for efficiency analysis.

Table 2: Efficiency of the system for each participant and each task.

| | Average | Standard deviation |
|---|---|---|
| User turns | 7.4 | 3.6 |
| Sys turns | 7.4 | 3.6 |
| ASR error times | 0.3 | 0.4 |
| Elapsed time (s) | 48.7 | 20.0 |

The average 7.4 user turns and 7.4 system turns per participant per task have shown a very satisfying efficiency of the system, because the average basic turn numbers, which can be inferred with the theoretically shortest solution, are 2.9 user turns and 2.9 system turns for the only spoken language input, and 5.6 user turns and 5.3 system turns for the touch-screen input. The standard deviation 3.6 even indicates that, some of the participants are solving tasks using the approximately shortest solutions.

However, as observed the average turn numbers are a bit higher than the average number for the shortest solution, with a further insight into the detailed data, two reasons can be concluded:

- 4 participants were using only touch-screen input to interact with the system, which significantly increased the total turn numbers.
- By combining spoken language and touch-screen inputs, many participants first used the touch-screen to sort out the rough direction for each task and then used spoken language instructions to find the target, which however inevitably increased the turn numbers, yet clearly indicates their intention of avoiding the possible problems caused by the automatic speech recognition. This is also reflected by the very good average ASR error rate: 0.3, no ASR error occurred during the interaction.

Meanwhile, the average elapsed time for each task and participant (48.7 seconds) is considered as very short as well, because even with the shortest solution using spoken language commands, merely 48.7 seconds were used for 5.8 user interaction paces (2.9 user turns + 2.9 system turns), which is averagely maximum 8.4 seconds for each turn, this even includes long sentences uttered by the system for over 10 seconds. Although the standard deviation 20.0 is a bit high, this is caused by the same participants, especially the one who was using only touch-screen and doing brute-force searching, and used averagely 135.8 seconds for each task.

### 5.3 User Satisfaction of the System

Regarding the user satisfaction of the system, we analysed the subjective data coming from the evaluation questionnaire concerning the interaction with the system with the combining modality. The results are summarized in Table 3, underlining very good user experiences with the combining modality.

Table 3: Data concerning subjective user satisfaction.

| | Mean | Standard deviation |
|---|---|---|
| Better than single modality? | 4.4 | 1.1 |
| Easier solving tasks? | 4.0 | 1.3 |
| Showing advantages? | 4.5 | 1.0 |
| Usable to use combi-modality? | 4.1 | 1.5 |
| Prefer to use combi-modality? | 4.4 | 1.3 |
| Not confusing? | 4.5 | 0.9 |
| Overall | 4.3 | 1.0 |

However, the scores of easier solving tasks and the usability of the combining modality were a bit lower than the others and the corresponding standard deviations were also higher. It is again mainly due to the extreme cases, where the participants only used touch-screen input and had made unpleasant impression of using only touch-screen, and therefore gave comparably lower score in the questionnaire.

## 6 CONCLUSIONS AND FUTURE RESEARCH

In this paper we reported our work on multimodal interaction for elderly persons by focusing on the following two important aspects:

- The summary of our systematically designed and empirically improved foundation for developing and implementing the elderly-centered multimodal interaction;
- The evaluation of the spoken language and touch-screen combined input modality of a multimodal interactive guidance system for the elderly by applying an adapted well-established evaluative framework.

Results of the evaluation showed a very high degree of effectiveness, efficiency and user satisfaction of our system, specifically by using the spoken language and touch-screen combining input modality. This confirmed our theoretical and technical foundation, approaches and frameworks on developing effective, efficient and elderly-friendly multimodal interactive systems.

The reported work continued the pursuit of our goal towards building effective, efficient, adaptive and robust multimodal interactive systems and frameworks for elderly in ambient assistive living environments. Further studies are needed to investigate the reported extreme cases. Corpus-based supervised and reinforcement learning techniques will be applied to support and improve the formal language driven and agent-based hybrid modelling and management approach. More relevant research and experiments on assisting elderly in navigating through complex buildings are also being conducted.

## ACKNOWLEDGEMENTS

We greatly acknowledge the support of the German Research Foundation through the Collaborative Research Center SFB/TR 8 Spatial Cognition, the department of Medical Psychology and Medical Sociology and the department of Neurology of the University Medical Center Göttingen.

## REFERENCES

Alston, P. W., 2000, *Illocutionary acts and sentence meaning.* Cornell University Press.

Birdi, K., Pennington, J., Zapf, D., 1997. Aging and errors in computer based work: an observational field study. In *Journal of Occupational and Organizational Psychology*. pp. 35-74.

Broadfoot, P., Roscoe, B., 2000. Tutorial on FDR and Its Applications. In *K. Havelund, J. Penix and W. Visser (eds.), SPIN model checking and software verification.* Springer-Verlag, London, UK, Volume 1885, pp. 322.

Charness, N., Bosman, E., 1990. Human Factors and Design. In *J.E. Birren and K.W. Schaie, (eds.), Handbook of the Psychology of Aging.* Academic Press, Volume 3, pp. 446-463.

Craik, F., Jennings, J., 1992. Human memory. In *F. Craik and T.A. Salthouse, (eds.), The Handbook of Aging and Cognition.* Erlbaum, pp. 51-110.

Folstein, M., Folstein, S., Mchugh, P., 1975. “mini-mental state”, a practical method for grading the cognitive state of patients for clinician. In *Journal of Psychiatric Research.* Volume 12, 3, pp. 189-198.

Fozard, J. L., 1990. Vision and hearing in aging. In *J. Birren, R. Sloane and G. D. Cohen (eds), Handbook of*

*Metal Health and Aging.* Academic Press, Volume 3, pp. 18-21.

Goetze, S., Fischer, S., Moritz, N., Appell, J. E., Wallhoff, F., 2012. Multimodal Human-Machine Interaction for Service Robots in Home-Care Environments. In *Proceedings of the 1st International Conference on Speech and Multimodal Interaction in Assistive Environments*. The Association for Computer Linguistics, pp. 1-7.

Hall, A., Chapman, R., 2002. Correctness by construction: Developing a commercial secure system. In *IEEE Software.* Vol. 19, 1, pp. 18-25.

Hawthorn, D., 2000. Possible implications of ageing for interface designer. In *Interacting with Computers.* pp. 507-528.

Jaimes, A., Sebe N., 2007. Multimodal human-computer interaction: A survey. In *Computational Vision and Image Understanding. Elsevier Science Inc.,* New York, USA, pp. 116-134.

Jian, C., Scharfmeister, F., Rachuy, C., Sasse, N., Shi, H., Schmidt, H., Steinbüchel-Rheinwll, N. v., 2011. Towards Effective, Efficient and Elderly-friendly Multimodal Interaction. In *PETRA 2011: Proceedings of the 4th International Conference on PErvasive Technologies Related to Assistive Environments. ACM,* New York, USA.

Jian, C., Scharfmeister, F., Rachuy, C., Sasse, N., Shi, H., Schmidt, H., Steinbüchel-Rheinwll, N. v., 2012. Evaluating a Spoken Language Interface of a Multimodal Interactive Guidance System for Elderly Persons. In *HealthInf 2012: Proceedings of the International Conference on Health Informatics. SciTepress,* Vilamoura, Algarve, Portugal.

Kline, D. W., Scialfa, C. T., 1996. Sensory and Perceptual Functioning: basic research and human factors implications. In *A. D. Fisk and W. A. Rogers. (eds.), Handbook of Human Factors and the Older Adult,* Academic Press.

Kotary, L., Hoyer, W. J., 1995. Age and the ability to inhibit distractor information in visual selective attention. In *Experimental Aging Research.* Volume 21, Issue 2.

Krajewski, J., Wieland, R., Batliner, A., 2008. An acoustic framework for detecting fatique in speech based human computer interaction. In *Proceedings of the 11th International Conference on Computers Help People with Special Needs*. Springer-Verlag Berlin, Heidelberg, pp. 54-61.

Lutz, W., Sanderson, W., Scherbov, S., 2008. The coming acceleration of global population ageing. In *Nature*. pp. 716-719.

Mackay, D., Abrams, L., 1996. Language, memory and aging. In *J. E. Birren and K. W.Schaie (eds), Handbook of the psychology of Aging*. Academic Press, Volume 4, pp. 251-265.

Mandel, C., Lüth, T., Laue, T., Röfer, T., Gräser, A., Krieg-Brückner, B., 2009. Navigating a Smart Wheelchair with a Brain-Computer Interface Interpreting Steady-State Visual Evoked Potentials. In *Proceedings of the 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE Xplore, St. Louis, Missouri, United States, pp. 1118-1125.

McDowd, J. M., Craik, F. 1988. Effects of aging and task difficulty on divided attention performance. In *Journal of Experimental Psychology: Human Perception and Performance 14*. pp. 267-280.

Moeller, S., Goedde, F., Wolters, M., 2008. Corpus analysis of spoken smart-home interactions with older users. In N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odjik, S. Piperidis, and D. Tapias, (eds.), Proceedings of the Sixth International Conference on Language Resources Association. ELRA.

Rodríguez, M. D., García-Vázquez, J. P., Andrade, Á. G., 2011. Design dimensions of ambient information systems to facilitate the development of AAL environments. In *Proceedings of the 4th International Conference on PErvasive Technologies Related to Assistive Environments*. ACM Press, New York, United States, pp. 4:1-4:7.

Roscoe, A. W., 1997. In *The Theory and Practice of Concurrency,* Prentice Hall.

Salthouse, T. A., 1994. The aging of working memory. In *Neuropsychology 8,* pp. 535-543.

Schieber, F., 1992. Aging and the senses. In *J. E. Birren, R. B. Sloane, and G. D. Cohen, (eds.) Handbook of Mental Health and Aging,* Academic Press, Volume 2.

Shi, H., Bateman, J., 2005. Developing human-robot dialogue management formally. In Proceedings of Symposium on Dialogue Modelling and Generation. Amsterdam, Netherlands.

Shi, H., Jian, C., Rachuy, C. 2011. Evaluation of a Unified Dialogue Model for Human-Computer Interaction. In *International Journal of Computational Linguistics and Applications*. Bahri Publications, Volume 2.

Traum, D., Larsson, S., 2003. The information state approach to dialogue management. In *J.v. Kuppevelt and R. Smith (eds.), Current and New Directions in Discourse and Dialogue.* Kluwer, pp. 325-354.

Walkder, N., Philbin, D. A., Fisk, A. D., 1997. Age-related differences in movement control: adjust submovement structure to optimize performance. In *Journal of Gerontology: Psychological Sciences 52B,* pp. 40-52.

Walker, M. A., Litman, D. J., Kamm, C. A., Kamm, A. A., Abella, A., 1997. Paradise: a framework for evaluating spoken dialogue agents. In *Proceedings of the eighth conference on European chapter of Association for computational Linguistics,* NJ, USA, pp. 271-280.