EVALUATING A SPOKEN LANGUAGE INTERFACE OF A MULTIMODAL INTERACTIVE GUIDANCE SYSTEM FOR ELDERLY PERSONS

Cui Jian[1], Frank Schafmeister[2], Carsten Rachuy[1], Nadine Sasse[2], Hui Shi[1,3], Holger Schmidt[4] 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]*German Research Centre for Artificial Intelligence, University of Bremen, Enrique-Schmidt-Straße 5, Bremen, Germany*
[4]*Neurology, University Medical Center Göttingen, Waldweg 37, Göttingen, Germany*

Keywords: ICT and ageing, Elderly-friendly interaction, User centered design, Human-computer interaction, Spoken dialogue systems, Formal methods, Multimodal interaction.

Abstract: This paper presents a multimodal interactive guidance system for elderly persons for the use in navigating in hospital environments. We used a unified modelling method combining the conventional recursive transition network based approach and agent-based dialogue theory to support the development of the central dialogue management component. Then we studied and specified a list of guidelines addressing the needs of designing and implementing multimodal interface for elderly persons. As an important step towards developing an effective, efficient and elderly-friendly multimodal interaction, the spoken language interface of the current system was evaluated by an elaborated experiment with sixteen elderly persons. The results of the experimental study are overall positive and provide evidence for our proposed guidelines, approaches and frameworks on interactive system development while advising further improvements.

1 INTRODUCTION

Multimodal interfaces are becoming more and more common since the inspirational introduction by (Bolt, 1980). They are considered as a promising possibility to improve the quality of communication between users and systems and have significant impact on effectiveness and efficiency of interaction (cf. e.g. (Jaimes and Sebe, 2007)), they also enhance users' satisfaction and provide a more natural and intuitive way of interaction (cf. e.g. (Oviatt, 1999)).

Meanwhile, the demographic development towards more elderly keeps motivating the research of elderly-friendly interactive systems; there is a special focus on the multimodal communication channels, which can enhance interaction by taking age-related decline into special accounts (Holzinger, Mukasa and Nischelwitzer, 2008).

In this paper, we will present an interactive guidance system for elderly persons. It uses a unified dialogue modelling approach combining the classic agent based dialogue theories and a formal language supporting generalized recursive transition network based method to achieve a flexible and context-sensitive, yet formally tractable and controllable interaction. Furthermore, it is developed according to a number of elaborated guidelines regarding basic design principles of conventional interactive systems and most common elderly-centered characteristics. To evaluate this system with respect to its feasibility and acceptance by elderly, an experimental study was conducted, which was focused on the natural spoken language input interface of the system. However, the study also aimed at evaluation of the multimodal interactive guidance system as a whole, while regarding the essential criteria of the following aspects: the effectiveness of task success, the efficiency of executing tasks and the user satisfaction with the system.

The remainder of the paper is organized as follows: section 2 introduces the formal unified dialogue modelling approach which combines the

Jian C., Schafmeister F., Rachuy C., Sasse N., Shi H., Schmidt H. and von Steinbüchel N..
EVALUATING A SPOKEN LANGUAGE INTERFACE OF A MULTIMODAL INTERACTIVE GUIDANCE SYSTEM FOR ELDERLY PERSONS.
DOI: 10.5220/0003783800870096
In *Proceedings of the International Conference on Health Informatics* (HEALTHINF-2012), pages 87-96
ISBN: 978-989-8425-88-1

classic agent based approach and the recursive transition network based theory for building the discourse management of the multimodal interaction; section 3 presents a set of specific guidelines for designing multimodal interactive system for elderly persons; section 4 then describes the multimodal interactive guidance system, which is developed based on the unified dialogue modelling approach and the proposed set of design guidelines; in section 5 the experiment is described, and the results are analysed and discussed in section 6. Finally, in section 7 we will conclude and give an outline of future work.

2 A FORMAL UNIFIED DIALOGUE MODELLING APPROACH

As a typical recursive network based approach, generalized dialogue models were developed by constructing dialogue structures at the illocutionary level (Sitter and Stein, 1992). However, it is criticized for its inflexibility of dealing with dynamic information exchange. On the other hand, information state update based theories were deemed the most successful foundation of agent based dialogue approaches (Traum and Larsson, 2003), which provides a powerful mechanism to handle dynamic information and gains a context sensitive dialogue management. Nevertheless, such models are usually very difficult to manage and extend (Ross, Bateman and Shi, 2005).

Thus, a unified dialogue modelling approach was developed. It combines the generalized dialogue models with information state updated based theories. This approach is supported by a formal development toolkit, which is used to implement an effective, flexible, yet formally controllable dialogue management for multimodal interaction.

2.1 A Unified Dialogue Modelling Approach

Generalized dialogue models can be constructed with the recursive transition networks (RTN). They abstract dialogue models by describing illocutionary acts without reference to direct surface indicators (Alston, 2000). Figure 1 shows a simple generalized dialogue model as a recursive transition network diagram. It is initiated with an assertion from a person A, and responded by B with three possible actions: accept, agree or reject.

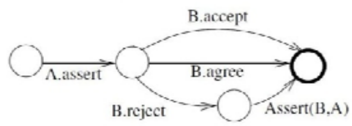

Figure 1: A generalized dialogue model as a simple RTN.

The generalized dialogue model above is a none-deterministic model, to build a feasible interaction model, deterministic behaviour should be assured for the interaction flow. Thus, conditional transitions are introduced to modify the above dialogue model (cf. figure 2). Let *checkAssert* be a method to check whether an assertion holds with B's knowledge and *a* an assertion given by A, if the assertion holds, B can agree with it; otherwise, B rejects it and initiates further discussion; if the assertion is not known by B, then B accepts it. Such conditional transitions can only be activated if the relevant condition is fulfilled. We call it the conditional RTN.

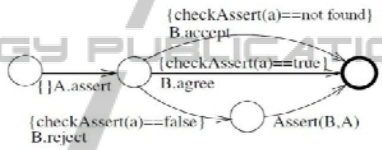

Figure 2: A generalized dialogue model as a simple deterministic RTN with conditional transitions.

Although the conditional RTN based generalized dialogue model defines a deterministic illocutionary structure, it does not provide the mechanism to integrate discourse information. Thus, information state based theory was integrated into our unified dialogue model by eliminating some typical elements, e.g. AGENDA for planning the next dialogue moves, because such information is already captured by the generalized dialogue model; furthermore it complements illocutionary structure with update rules, which is associated with the information state of current context, and can update the information state respectively if necessary. As a result, a unified dialogue model is constructed as shown in figure 3. Four update rules are added, so that the information state regarding context can always be considered and updated; e.g. the update rule ACCEPT is used to add a new assertion *a* into B's belief and refer it as known from then on.

Finally, we define a unified dialogue model as a deterministic recursive transition network built at the illocutionary level of interaction processes; its transitions can only be triggered by fulfilled conditions concerning the information state, and with the consequences of possible information state update according to a set of update rules.

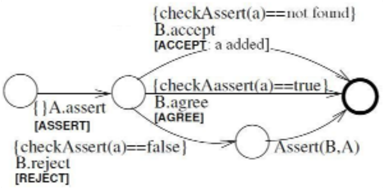

Figure 3: Unified dialogue model as a simple deterministic RTN with conditional transitions and update rules.

2.2 A Formal Language based Development Toolkit for Dialogue Modelling

Deterministic recursive transition networks can be illustrated as a typical finite state transition diagram (cf. figure 3), which provides the possibility of specifying the described illocutionary structure with mathematically well-founded formal methods, e.g., with *Communicating Sequential Processes* (CSP) in the formal methods community of computer science.

CSP can not only be used to specify finite state automata structured patterns with abstract, yet highly readable and easily maintainable logic formalization (cf. (Roscoe, 1997)), but it is also supported by well-established model checkers to verify the concurrent aspects and increasing the tractability (Hall, 2002). Thus, CSP is used to specify and verify the unified dialogue models (cf. the example in figure 4).

```
UDM = A.assert -> B.checkAssert ->
        ( B.accept -> UDM
        [] B.agree -> UDM
        [] B.reject -> AssertBA)
```

Figure 4: A sample CSP specification of the illocutionary structure of the unified dialogue model in figure 3.

In order to support the development of unified dialogue models within practical interactive systems, we provided FormDia, the Formal Dialogue Development Toolkit (cf. figure 5).

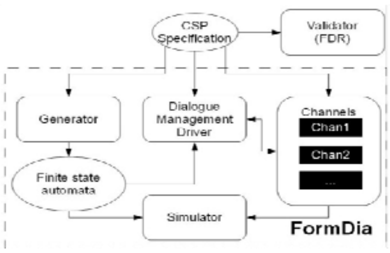

Figure 5: The Structure of the FormDia Toolkit (cf. (Shi and Bateman, 2005)).

To develop the unified dialogue model based management, FormDia toolkit can be used according to the following essential steps:

- **Validation**: the CSP specified structure of a unified dialogue model can be validated by using Failures-Divergence Refinement tool, abbrv. FDR (Broadfoot and Roscoe, 2000), which is a model checking tool for validating and verifying concurrency of state automata.
- **Generation**: according to the given CSP specification, finite state automata can then be generated by the FormDia Generator.
- **Channels Definition**: channels between the dialogue management and application/domain specific components can be defined. These channels are at first black boxes, which will later be filled with deterministic behaviour of concrete domain components.
- **Simulation**: with the generated finite state automata and the communication channels, dialogues scenarios are simulated via a graphical interface, which visualizes dialogue states as a directed graph and provides a set of utilities to trigger events and the dialogue state update for testing and verification.
- **Integration**: after the dialogue model is validated, tested and verified, it can be directly integrated into a practical interactive dialogue system via a dialogue management driver.

The FormDia toolkit shows a promising way for developing formally tractable and extensible interaction. It enables an intuitive design of dialogue models with formal language, automatic validation of related functional properties, and it also provides an easy simulation, verification for the specified dialogue models, and the straightforward integration within a practical interactive system. In addition, with the unified dialogue model, FormDia toolkit can even be used in multimodal interactive system.

3 DESIGN GUIDELINES OF MULTIMODAL INTERACTIVE SYSTEMS FOR ELDERLY PERSONS

Elderly persons often suffer from decline of sensory, perceptual, motor and cognitive abilities due to age-related degenerative processes. (Birdi, Pennington and Zapf, 1997) and (Morris, 1994) indicated that this decline should be considered while designing interactive systems for the elderly. Therefore, we defined a set of design guidelines for multimodal interaction with respect to the decline of the most common abilities. They are implemented and integrated into our multimodal interactive guidance system and tested by a pilot study. The results are described in (Jian, et al., 2011) and the improved guidelines are now presented as follows, regarding the decline of the seven most common abilities.

3.1 Visual Perception

Visual perception declines for most people with age (Fozard, 1990). Even in the early forties, many people find it more difficult to focus on objects up close and to see fine details. The size of the visual field is decreasing and leads to loss of peripheral vision. Rich colours and complex shapes make images hard or even impossible to identify. Rapidly moving objects are either causing too much distraction, or becoming less noticeable. To cope with these impairments, the following guidelines should be taken into account:

- Layouts of the user interface should be devised as simple and clear as possible, with few (if any) or no overlapping items.
- All texts should be large enough, suggesting simple fonts in the 12-14 point range.
- Strong contrast should be used with as few colors as possible; this also applies to simple and easily recognizable shape designs.
- Unnecessary and irrelevant visual effects and animation should be avoided.

3.2 Speech Ability

Elderly persons need more time to produce complex words or longer sentences, probably due to reduced motor control of tongue and lips (Mackay and Abrams, 1996). Furthermore, speech-related elderly-centered adaptation is necessary to improve the interaction quality to a sufficient level (Moeller, Goedde and Wolters, 2008). Based on these, the following aspects should be taken into account:

- Acoustic models specialized for the elderly should be used for speech recognizer.
- Vocabulary should be built with more definite articles, auxiliaries, first person pronouns and lexical items related to social interaction.
- Dialogue strategies should be able to cope with elderly specific needs such as repeating, helping and social interaction, etc.

3.3 Hearing Ability

Hearing ability declines to 75% with increasing age 75 and 79 year olds, (Kline and Scialfa, 1996). High pitched sounds are increasingly not percieved, as well as long and complex sentences becoming difficult to follow (Schieber, 1992). Therefore special attention should be paid to the following:

- Text displays can help when information is mis- or not heard.
- Synthesized texts should be intensively revised regarding style, vocabulary, length and sentence structures suitable for elderly.
- Low pitched voices are more acceptable for speech synthesis, e.g., female voices are less preferred than male ones.

3.4 Motor Abilities

Using a computer mouse has been problematic for many elderly persons as good hand-eye coordination is required (Walkder, Philbin and Fisk, 1997). It is difficult for them to position the cursor if the target is too small or too irregular to locate, and they have problems with control of fine movements (Charness and Bosman, 1990), especially when other cognitive functions are required at the same time. Thus, the following procedures are suggested:

- Direct interaction is recommended.
- All GUI items should be accessibly shaped, sized and well spaced from each other.
- Simple movements are recommended, such as clicking instead of dragging or drawing.
- Text input should be avoided or replaced with other simpler input actions.
- An undo function is needed to correct errors.
- Simultaneous multimodal input such as the combination of speech and other input should be avoided or replaced.

3.5 Attention and Concentration

Elderly individuals become more easily distracted by details or noise (Kotary and Hoyer, 1995). They display great difficulty maintaining divided attention, e.g. where attention must be paid to more than one aspect at the same time (McDowd and Craik, 1988). To cope with these constraints the following points are suggested:

- Only relevant images should be used.
- Items should not be displayed simultaneously.
- Unified or similar fonts, colors and sizes of displayed texts are recommended.
- Changes on the user interface should be emphasized in an obvious way.

3.6 Memory

Different memory functions decline at different degrees during ageing. Short term memory holds fewer items while ageing and more time is needed to process information (Hoyer and Rybash, 1992). Working memory also becomes less efficient (Salthouse, 1994). Semantic information is believed to be preserved in long term memory (Craik and Jennings, 1992). To compensate the decline of the different memory functions, the following points are suggested:

- Pure image items should be avoided or placed near relevant key words.
- Presented items should not exceed five, the average maximum capacity of short term memory of elderly people.
- Information should be categorized to assist storage into long term memory.
- Context sensitive information is necessary to facilitate working memory activities.

3.7 Intellectual Ability

Fluid intelligence does decline with ageing (Shaie, 1996), however, crystallized intelligence does not (Hawthorn, 2000); it can assist elderly people to perform better in a stable well-known interface environment. To reflect this on interface design, we suggest assuring the following points:

- Unified interface layout, where changes should only happen on data level.
- Semantically intuitive structure, where users should not be too surprised while traversing the interaction levels.
- Consistent interaction style, easing learning and assist elderly to master interface use.

4 MULTIMODAL INTERACTIVE GUIDANCE SYSTEM FOR ELDERLY PERSONS

The Multimodal Interactive Guidance System for Elderly Persons (MIGSEP) was developed for elderly or handicapped persons to navigate through public spaces. MIGSEP runs on a portable touch screen tablet PC. It serves as the interactive media designed for an autonomous intelligent electronic wheelchair that can automatically carry its users to desired locations within complex environments.

4.1 System Architecture

The architecture of MIGSEP is illustrated in figure 6. A Generalized Dialogue Manager is developed using the unified dialogue modelling approach.

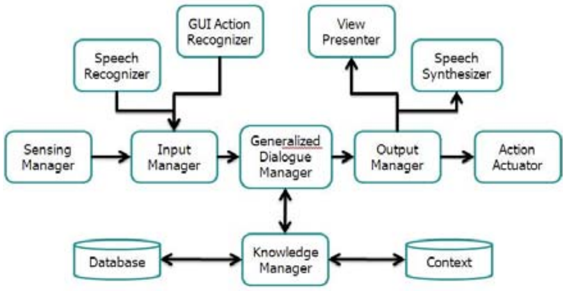

Figure 6: The architecture of MIGSEP.

It functions as the central processing unit and enables a formally controllable and extensible, meanwhile context-sensitive multimodal interaction. An *Input Manager* receives and interprets all incoming messages from *GUI Action Recognizer* for GUI inputs, *Speech Recognizer* for natural language understanding and *Sensing Manager* for other sensor data. An *Output Manager* on the other hand, handles all outgoing commands and distributes them to *View Presenter* for visual feedbacks, *Speech Synthesizer* to generate natural language responses and *Action Actuator* to perform necessary motor actions. *Knowledge Manager* uses *Database* to keep the static data of certain environments and *Context* to process the dynamic information exchanged with users during the interaction.

Although the essential components of MIGSEP are closely connected with each other via predefined XML-based communication mechanism, each of them is treated as an open black box and can be implemented or extended for specific use, without affecting other MIGSEP components. It provides a general platform for both theoretical researches and empirical studies on multimodal interaction.

4.2 The Unified Dialogue Model in MIGSEP

The current unified dialogue model (UDM) consists of four extended state transition diagrams.

The interaction is initiated with the diagram *Dialogue(S, U)* (cf. figure 7), by the initialization of the system's start state and a greeting-like request.

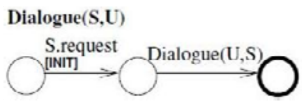

Figure 7: The initiate diagram.

The dialogue continues with user's instruction to a certain location, request for a certain information or restart action, leading to the system's further response or dialogue restart, respectively, as well as updating the information state with the attached update rules (cf. *Dialogue(U, S)* in figure 8).

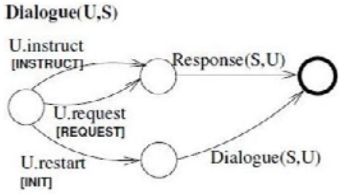

Figure 8: The transition diagram triggered by the user.

After receiving user's input, the system tries to generate an appropriate response according to its current knowledge base and information state (cf. *Response(S, U)* in figure 9). 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 different diagrams.

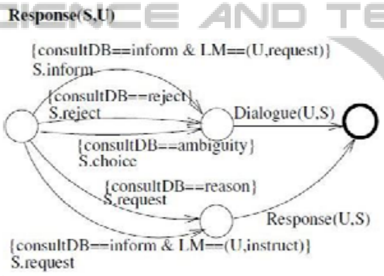

Figure 9: The system's response.

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 10).

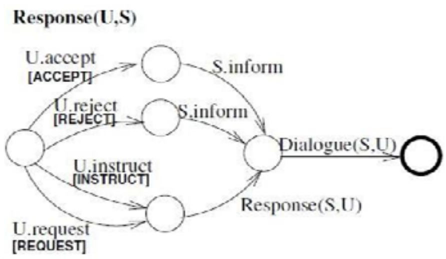

Figure 10: The user's response.

Using the FormDia toolkit, the UDM was developed as CSP specifications, and its functional properties have been validated and verified via FDR, as well as its conceptual interaction process using FormDia simulator. The tested specification was then used to generate corresponding machine-readable state transition automata and integrated into the *Generalized Dialogue Manager* of MIGSEP.

4.3 The Elderly-friendly Design Elements in MIGSEP

According to the design guidelines in the previous section, a set of elderly-centered design elements were implemented in MIGSEP. Specifically, the most essential elements are listed below:

- **Visual Perception**: 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, enabling comfortabe perception and easy recognition.
- **Hearing Ability**: both text and acoustic output are provided as system responses. Styles, vocabulary, structures of the sentences have been intensively revised. A low-pitched yet vigorous male voice is chosen for the synthesis.
- **Motor Functions**: regularly shaped, sufficiently sized and well separated interface elements were designed for easy access. Clicking was decided to be the only action to avoid otherwise frequently occurring errors caused by decline of motor and attentional functions. "Start" was provided as the only way of orientating oneself to avoid confusion.
- **Attention or Concentration**: fancy irrelevant images or decorations were avoided. Unified font, colors, sizes of interface elements were used for the entire interface. Simple animation notifying changes were constructed, giving sufficiently clear yet not distracting feedback to the user.
- **Memory Abilities**: all items are used with relevant keywords. The number of displayed items is restricted to no more than three, considering the maximum capacity of short term memory, the accessible size as well as the readable amount of information of the interaction items on a table PC. Logically well-structured and sequentially presented items were intensively revised to assist orientation during interaction. Context sensitive clues are given with selected colors.
- **Intellectual Ability**: consistent layout, colours and interaction styles are used. Changes on the interface happen only on data level.

4.4 Interaction with MIGSEP in Hospital Environments

We have implemented a MIGSEP system and set its application domain to hospital environments. Figure 11 shows a user interacting with it via speech modality.

Figure 11: A user is interacting with MIGSEP.

Figure 12 shows a sample dialogue between the MIGSEP system and a user who would like to be guided 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: to Dr. Wolf please.
S: <enlarges the second green card>
Do you want to go to Mrs. Dr. Diana Wolf?
U: Yes please.
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 12: Example of a dialogue with MIGSEP.

5 THE EXPERIMENTAL STUDY

To evaluate how well elderly is assisted by MIGSEP system, an experimental study was conducted.

5.1 Participants

Eighteen elderly persons (m/f: 11/7, mean age of 70.9, standard deviation (SD)=3.0), all German native speakers, took part in the study. They all had the mini-mental state examination (MMSE), which is a screening test to measure cognitive mental status (Folstein, Folstein and Mchugh, 1975). A test value between 28 and 30 indicates normal cognitive functioning, therefore, our participants showing 28.3 (SD=.86) were in the normal range.

5.2 Stimuli and Apparatus

As shown in figure 11, visual stimuli were given by the green lamp and the graphical user interface on the screen of a portable tablet PC; audio stimuli as complementary feedbacks were also generated by the MIGSEP system and presented via two loudspeakers at a well-perceivable volume. All tasks were given as keywords on the pages of a calendar-like system. The only input possibility was the spoken language instructions, activated if the button was being pressed and the green lamp was on.

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

During the experiment each participant was accompanied by only one investigator, who gave the introduction and well-defined instructions at the beginning, and provided help if necessary (which was very rare the case).

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

A questionnaire focusing on the user satisfaction was designed. It includes questions of seven categories: system behaviour, speech output, textual output, interface presentation, task performing, user-friendliness and user perspective. The questionnaire was completed by each participant by a five point Likert scale, where one represents the lowest appropriateness and five the highest.

5.3 Procedure

Each participant had to undergo four phases:

- **Introduction**: a brief introduction was given to the participants.
- **Learning**: they were instructed how to interact with the MIGSEP system using the button device and spoken natural language. After they made no more mistakes using the button device, a further introduction was given to the verbal and graphical feedbacks the system provides. Then they were asked to perform one or two sample tasks to gather more practical experiences with the system.
- **Testing**: Each participant had to perform eleven tasks, each of which contains incomplete yet sufficient information about a destination the participant should select. Each task was ended, if the goal was selected, or the participant gave up trying after six minutes.
- **Evaluation**: After all tasks were run through, each participant was asked to fill in the questionnaire for evaluation.

5.4 Questions and Methods

Altogether, there are three important questions to be focused and answered by the experiment:

- ”Can elderly use the MIGSEP system to complete the tasks?”
  A standard measurement method Kappa coefficient is used to assess the successfulness of the interaction between the participants and the system.
- “Can elderly persons handle the tasks with MIGSEP efficiently?”
  This shall be answered by the automatically logged data of every single interaction.
- “Do elderly find it comfortable to interact with MIGSEP?”
  This should be reflected in the results of the evaluation questionnaires.

6 RESULTS

6.1 Effectiveness of MIGSEP

To answer the first question, i.e., how well the MIGSEP system assists elderly persons to perform tasks, we used Kappa coefficient, which is a well-accepted method for measuring effectiveness of interaction (Walker, et al., 1997).

In order to apply this method, we needed to define the attribute value matrix (AVM), which had to contain all information that has to be exchanged between MIGSEP and the participants. E.g. table 1 shows the AVM for the task: ”Drive to a person named Michael Frieling.”, where the expected values of this task are also presented.

Table 1: An example AVM for the task “drive to a person name Michael Frieling”.

Attribute	Expected value
FN	Michael
LN	Frieling
G	Male

By combining the actual data recorded during the experiment with the expected attribute values in the AVMs, we can construct the confusion matrices for all tasks. E.g., table 2 shows the confusion matrix for the task ”drive to a person named Michael Frieling”, where ”M” and ”N” denote whether the actual data match with the expected attribute values in the AVMs. E.g. one participant selected a person with wrong first and last names.

Table 2: The confusion matrix for the task “drive to a person named Michael Frieling”.

	FN		LN		G		sum
Data	M	N	M	N	M	N	
FN	17	1					18
LN			17	1			18
G					18		18

Given one confusion matrix, the Kappa coefficient can be 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 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.

Therefore, we summarized the results of all the tasks and constructed one confusion matrix for all the data, and got that, $P(A) = 0.961$ and kappa coefficient $\kappa = 0.955$, which suggests a highly successful degree of interaction between the MIGSEP system and the participants.

6.2 Efficiency of MIGSEP

Regarding the efficiency of MIGSEP, quantitative data automatically logged during the experiments are summarized in table 3, with respect to user turns, system turns, ASR failed times (the frequency of the Automatic Speech Recognizer failing getting a parsable sentence), ASR error times (the frequency of the ASR wrongly recognizing utterances), user turns without ASR (user turns without being affected by the ASR related failures) and the elapsed time for each participant and each task.

From a dialogue system’s points of view, a very good overall performance efficiency is shown by averagely 4.1 user turns and 3.9 system turns per task for each participant, as the average basic turn numbers, which can be inferred by the shortest solution regarding the number of slots for each task to be filled, are 3 user turns and 3 system turns. In addition, if the ASR related failures and errors are excluded, the user turns would be only 1.9. This shows that almost each task was completed by each participant with only one complicated sentence. Furthermore, the user turns without ASR, which is lower than the theoretically minimum 2 user turns, even implied that with slightly wrong recognized sentence, the MIGSEP system was still able to find a solution to help elderly persons to complete tasks.

Table 3: Quantitative results calculated based on the recorded data concerning efficiency.

	Average	Standard deviation
User turns	4.1	1.8
Sys turns	4.0	1.7
ASR failed times	1.2	0.8
ASR error times	1.0	1.2
User turns without ASR	1.9	0.4
Elapsed time	61.0	23.6

On the other hand, the elapsed time for each task and each participant is considered as satisfying, with averagely 61.0 second for minimal 6 interaction paces (3 user turns +3 system turns), including the relatively long spoken utterance either by the system or the elderly participants. However, the standard deviation of 23.6 is a bit high, since two participants needed much longer time than the others. They encountered many problems with the automatic speech recognizer, which indicates the necessity for further analysis and improvement of the ASR.

6.3 User Satisfaction

Table 4: The assessment of subjective user satisfaction.

	Mean	Standard deviation
System behaviour	3.7	0.8
Speech output	4.5	0.5
Textual output	4.7	0.5
Interface presentation	4.6	0.4
Task performing	4.3	0.4
User-friendliness	4.6	0.4
User perspective	3.9	0.8
Overall	4.3	0.4

Overall, it shows a very good user satisfaction with the averagely score of 4.3 out of 5. Specifically, the speech and textual outputs are considered appropriately constructed with the score of 4.5 and 4.7; the interface is intuitive and easy to understand with the score of 4.6; the process to perform the task is quite feasible with the score of 4.3; and the system is rather user-friendly with the score of 4.6 out of 5.

However, the scores of system behaviour and user perspective were a bit lower than the others. It is mainly due to the problem of the automatic speech recognizer, which could trigger unexpected system responses, and therefore make the future use from the user perspective less attractive.

7 CONCLUSIONS AND FUTURE WORK

This paper presented our work on multimodal interaction for elderly persons from three essential perspectives: the modelling and development of multimodal interaction using a tool-supported, formally tractable and extensible unified dialogue modelling approach; the design and implementation of a multimodal interactive system according to a number of elderly-friendly guidelines regarding the basic design principles of conventional interactive interfaces and ageing centered characteristics. The multimodal interactive system was evaluated with eighteen elderly persons. The evaluation showed high effectiveness, high efficiency and a high satisfaction of the user with our system. These findings provide us with further evidence for our proposed guidelines, approaches and frameworks on system design and implementation.

The presented work served as part of a developmental process towards building an effective, efficient, adaptive and robust multimodal interactive framework for the elderly. Further study focussing on speech and touch screen combined modalities is being conducted. Moreover, corpus-based supervised and reinforcement learning techniques will be applied to improve the current dialogue model and gain more flexible interaction to compensate for the insufficient reliability of automatic speech recognizers. Our future research will continue with combining and experimenting emerging technologies in addition to speech, touch screen and visual modalities. Special attentions are also being paid to learning-based discourse modelling and management in advanced multimodal interactive systems for elderly persons.

ACKNOWLEDGEMENTS

We gratefully acknowledge the support of the Deutsche Forschungsgemeinschaft (DFG) through the Collaborative Research Center SFB/TR8, the department of Medical Psychology and Medical Sociology and the department of Neurology of the University Medical Center Göttingen, and the German Research Centre for Artificial Intelligence.

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.

Bolt, R. A., 1980. Put-That-There: Voice and Gesture at the Graphics Interface. In *Proceedings of the 7th International Conference on Computer Graphics and Interactive Techniques.* Seattle, USA, pp. 262-270.

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.

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

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

Holzinger, A., Mukasa, K.S., Nischelwitzer, A. K., 2008. Introduction to the special thematic session: Human-computer interaction and usability for elderly. In *Proceedings of the 11th International Conference on Computers Help People with Special Needs.* Springer Verlag, Berlin, Germany, pp. 18-21.

Hoyer, W. J., Rybash, J. M., 1992. Age and visual field differences in computing visual spatial relations. In *Psychology and Aging 7.* pp. 339-342.

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.

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.

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.

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.

Morris, J. M., 1994. User interface design for older adults. In *Interacting with Computers.* Vol. 6, 4, pp. 373-393.

Oviatt, S. T., 1999. Ten myths of multimodal interaction. In *Communications of the ACM.* ACM New York, USA, Vol. 42, No. 11, pp. 74-81.

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

Ross, J. R., Bateman, J., Shi, H., 2005. Using Generalized Dialogue Models to Constrain Information State Based Dialogue Systems. In *Symposium on Dialogue Modelling and Generation.*

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.

Shaie, K.W., 1996. Intellectual development in adulthood. In *J.E. Birren and K.W. Shaie, (eds.), Handbook of the psychology of aging.* Academic Press, Volume 4.

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

Sitter, S., Stein, A., 1992. Modelling the illocutionary aspects of information-seeking dialogues. In *Journal of Information Processing and Management.* Elsevier, Volume 28, issue 2, pp. 165-180.

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.