Explorative Studies on Multimodal Interaction in a PDA- and Desktop-based Scenario

Andreas Ratzka
IMIK – Department of Information Science
University of Regensburg
D-93040 Regensburg
Germany
andreas.ratzka@sprachlit.uni-regensburg.de

ABSTRACT

This paper presents two explorative case studies on multimodal interaction. Goal of this work is to find and underpin design recommendations to provide well proven decision support across all phases of the usability engineering lifecycle [1]. During this work, user interface patterns for multimodal interaction were identified [2, 3]. These patterns are closely related to other user interface patterns [4, 5, 6]. Two empirical case studies, one using a Wizard of Oz setting and another one using a stand-alone prototype linked to a speech recognition engine [7] were conducted to assess the acceptance of resulting interaction styles. Although the prototypes applied as well those interface patterns that increase usability by means of traditional interaction techniques and thus compete with multimodal interaction styles, multimodal interaction was preferred by most of the users.

Categories and Subject Descriptors

H.5.2 [**User Interfaces**]: Interaction Styles, Theory and Methods; H.1.2 [**User/Machine Systems**]: Human Factors; I.3.6 [**Methodology and Techniques**]: Interaction Techniques

General Terms

Design, Experimentation, Human Factors, Verification

Keywords

multimodality, user interface patterns, mobile computing

1. INTRODUCTION

1.1 User Interface Patterns for Multimodal Interaction

The context of this work is usability engineering for multimodal interaction. Previous work in this field focusses on artifacts, specification languages and processes [8, 9]. Other work, which also covers the representation of design knowledge, is limited to the phases of requirements analysis and work reengineering whereas the phases of screen design standards and detailed design are addressed only marginally [10, 11, 12].

Patterns provide design knowledge accross all phases of the development process. They are "three-part rules" relating together the context (where this pattern can be used), the problem to be solved and the solution for this problem. Patterns are based on a proven solution which can be found in (at least three) example applications [13, 14]. They are semi-formal rules: Their overall structure is formalised in the sense that patterns are composed of a uniform inventory of sections. But within these sections, natural language is used for detailed description.

Patterns originated in architecture in the late sixties [15, 13] but found their actual proliferation in object oriented programming [16] and software architecture [17] as well as in user interface and website design [4, 5, 18, 19, 6, 20].

As multimodal interaction is a relatively young field with still very little market penetration, eliciting patterns of successful system implementations seems to be almost impossible. Nevertheless, multimodal interaction can look back on almost thirty years of research such that recurring patterns of successful interaction techniques can be identified via literature mining [2]. In the context of this work, a dozen of patterns have been identified [3], two of which are illustrated in this work (view table 1).

These patterns do not stand alone but are in close relation to one another and even to patterns from other, GUI-only collections (view table 2 for some examples).

One problem is validating user interface patterns. Psychological experiments are of only limited use for pattern identification or validiation. Patterns cannot be reduced to simple numbers that can be verfied statistically. Nevertheless, the interaction techniques of an interface which are the result of applying some patterns can be assessed according to user acceptance. These user tests are the main focus of the research presented in this paper.

1.2 Empirical Studies on Multimodal Interaction – State of the Art

A Wizard of Oz simulation revealed that users like to combine speech and gestures for graphic editing [21]. A semi-automatic simulation of a multimodal service transaction system on a LCD-tablet revealed that 87% of all in-

ICMI'08, October 20–22, 2008, Chania, Crete, Greece.

Problem	Solution
Voice-based Interaction Shortcut	
The user has to select items from a large set. Consider selecting an action from a menu or selecting a list item from a drop-down chooser. [...] Which interaction style allows the user to quickly select the desired item?	Selecting objects or actions via speaking them can significantly speed up interaction. This is especially true for frequent users to whom the command and item names are well known.
Speech-enabled Form	
The user has to input structured data which can be mapped to some kind of input form consisting of a set of atomic fields. [...] How to simplify string input in form filling applications?	Whereever possible determine acceptable values for each form field. Support value selection via *Drop-down Choosers* and, alternatively, via voice commands.

Table 1: Some Patterns for Multimodal Interaction

put words are spoken by the user [22]. Digits were more frequently written than text, and proper names were more frequently written than other text. Multimodal interaction was – according to a survey – preferred throughout all tasks.

Studies on the mUltimo3D system [12] revealed that test subjects used speech input more frequently for data input than for command input. The users dictated dates less frequently than other concepts.

Oviatt et al. [23] found out that a more structured (form-based) presentation leads to shorter user utterances and less information per utterance. In contrast, free formulated input showed more linguistic complexity, that is the language model for free spoken input had a by three times higher perplexity than the one for structured input. Structured input helped to avoid ambiguities and unpredictable formulation variations. Similar results on form-based multimodal input were reported by Angeli et al. [24].

The effects on utterance length were confirmed by subsequent studies on a multimodal map-based interface [25]. Multimodal input helped to shorten user utterances and thus avoid disfluencies. Aditionally, for map-based tasks was multimodal input the fastest interaction style. For numeric and verbal tasks, speech input was the fastest one.

These results were confirmed in empirical studies with the map-based QuickSet system [26, 27, 28, 29]. Almost everybody preferred to interact multimodally, but only 20% of the time, users really interacted this way. Furthermore, individual differences were observed. Some users consistently made use of simultaneous speaking and pointing, whereas others consistently pointed first and spoke after a short pause. The spoken parts of user input can be simplified significantly, if users are able to provide spatial information via pointing. Furthermore, increased efficiency and error avoidance were reported as consequences of multimodal interaction styles.

Multimodal combination helps to reduce recognition errors, especially in the case of non-native speakers [30], suboptimal microphones [31, 32], noisy environments [33] or exhausted speakers [34].

Increased efficiency due to multimodality was reported in graphic design applications [35], mobile messaging systems [36, 37] and map-based systems [38].

Problem	Solution
Continuous Filter [6]	
"The user needs to find an item in an ordered set".	"Provide a filter component with which the user can in real time filter only the items in the data that are of his interest."
Autocompletion [5]	
"The user types something predictable, such as a URL, [...] or a filename [...]".	"With each additional character that the user types, the software quietly forms a list of the possible completions to that partially entered string [...]".
Composed Command [4]	
How can the artifact best present the actions that the user may take?	Provide a way for the user to directly enter the command, such as by speech or by typing it in.
Two-panel Selector [5]	
"You're presenting a list of objects, categories, or even actions. [...]"	"Put two side-by-side panels on the interface. In the first show a set of items that the user can select at will; in the other, show the content of the selected item." [...]

Table 2: Tradtional User Interface Patterns

2. EXPLORATIVE STUDIES FOR PATTERN VERIFICATION – GENERAL SETUP

User interfaces as well as user interface patterns are highly context dependent and cannot be simply verified via mathematical computations or standardised psychological settings as can be done with natural laws. Furthermore, patterns are based on proven concepts of design. That means, that patterns rarely cover innovative but well-known solutions.

Nevertheless, to get some feeling about the users' actual behaviour during multimodal interaction and to assess the plausibility and acceptance of the pattern-based interaction styles explorative studies were conducted.

Two studies are presented which compare multimodal PDA and desktop interaction with an e-mail organiser prototype. The first study is a Wizard of Oz simulation of the spoken parts of interaction whereas the prototypes for the second study make use of speech recognition.

After filling in a demographic survey, the test subjects were required to interact with both the desktop and Pocket PC. One part of the subjects interacted with the Pocket PC first whereas the other part started with the desktop PC. The interactive tasks comprised retrieving, forwarding and answering e-mails.

For some of the first tasks, the subjects were instructed to use either only pen (in the desktop-setting: mouse and keyboard) or only speech in order to become acquainted with the different interaction styles.

In order to avoid a one-sided bias for the subsequent tasks, the first tasks were permutated throughout the test subjects so that some of them began with pen-/mouse-/keyboard-only tasks and the other ones with speech-only tasks. For the remaining tasks the users were free to interact either traditionally (pen or keyboard and mouse), via speech input, or in a combined fashion.

After the test, the users were asked to fill in a questionary gathering their subjective judgments about the interaction.

3. SYSTEM SETUP

The functionality of the prototypes consisted of:

- Filtering e-mails according to sender or subject/text
- Answering, forwarding and creating e-mails

The first case study was a Wizard of Oz simulation of a multimodal e-mail organiser for both desktop and PDA systems. 35 test subjects participated in this study and interacted both with the desktop and PDA scenario. Patterns applied in this setting were

- Autocompletion (of sender / receiver names during filtering and creating e-mails)
- Two-panel Selector (in the case of the desktop setting: A message preview area was located beneath the message selection list)
- Composed Command (for creating e-mails and providing receivers in one step)
- Voice-based Interaction Shortcut (for forwarding / answering e-mails without the need to open them)

The Wizard of Oz setting consisted of two desktop computers – one was the wizard computer, the other one was the interaction computer for the desktop setting – and a COMPAQ PocketPC for the PDA-setting. The desktop-computers were connected via the university LAN. The PocketPC communicated with the wizard PC via bluetooth.

In the second study, a stand-alone prototype that was linked to a speech recognition engine was used. 25 test persons took part in these test runs. The design was refined and additional user interface patterns were applied such as

- Continuous Filter (to automatically update the message list according to letters input into the search field)
- Speech-enabled Form (for inputting search words, or receiver, subject and text of a newly created message)

In the second study, only one desktop PC was needed. That is, in the PDA-scenario, the Pocket PC was linked via bluetooth to the speech recognition server on the desktop PC. In the desktop-scenario, the speech recognition server and the user interface were run both on the same desktop PC.

4. RESULTS

4.1 Subjective Judgements

Most of the test persons stated that they would like (or rather like) than dislike to use the simulated PDA and desktop system as well as the stand-alone desktop prototype. The acceptance of the stand-alone PDA prototype was low in contrary. The usage was accordingly evaluated as mostly rather motivating for the simulated prototypes and the stand-alone desktop system and rather frustrating in the stand-alone PDA prototype. The desktop setting tended to be judged more usable than the PDA system for both the wizard setup and the stand-alone prototypes.

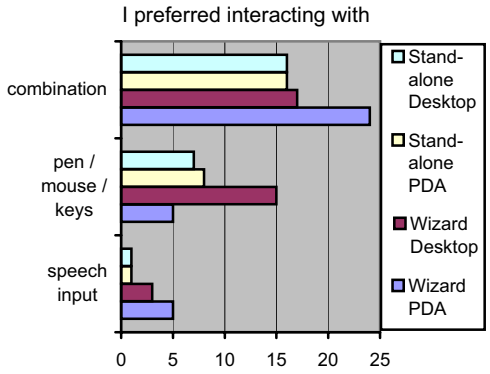

Figure 1: Preferred Interaction Modalities

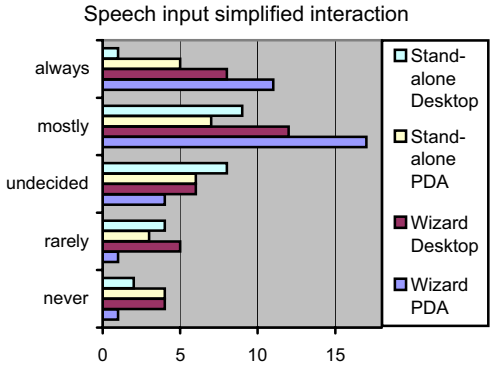

Figure 2: The Value of Speech Input

When asked about preferred interaction styles, far most test persons preferred the combination of pen and speech in the PDA wizard prototype and as well in the stand-alone PDA and desktop systems. In the desktop wizard prototype, the combination of mouse, keyboard and speech found the most proponents closely followed by traditional mouse-keyboard interaction (view figure 1).

According to most of the test persons, speech simplifies input, especially in the PDA wizard simulation but also, yet less clearly, in the stand-alone systems (view figure 2).

Speech input was the most difficult input style except in the wizard-simulated PDA setting, where pen-input was assessed worse.

4.2 Interaction Efficiency

The interaction time was compared for the warming-up task which required the test subjects to use either speech-only or traditional interaction. A naive examination of the interaction time in the PDA setting revealed a highly significant advantage of speech input.

A deeper analysis of the log-files and questionaries revealed, however, that these data may not be generalisable. Six test persons complained that they were not notified whether the e-mail they tried to send was really sent by the system. Fifteen test persons thus forwarded the mail under question several times.

This problem did not occur with speech interaction because in this case additional speech feedback was provided. There where several other delays caused by technical problems.

When these delays were substracted for the traditionally performed tasks, the mean effect was reduced to negligible 2 seconds.[1] When considering only those (fifteen) test persons that interacted with the desktop system first and therefore had already a notion of speech interaction when interacting with the PDA, nevertheless a significant result slightly in favour of speech interaction was observed (view table 3).

	pen	speech	⊘ effect	p(α) one-sided
average	85,5 s	75,8 s	9,7 s	2,5% (Wilcoxon)
std. ϵ	4,1	6,6		

Table 3: Interaction Speed (PDA Wizard Setting)

When analysing interaction times in the desktop scenario, a non-significant and negligible effect can be observed.

When comparing speech interaction times during the first and second test runs, a significant learning effect was observed. In almost every case, speech interaction was faster in the second test run than in the first run.

4.3 Recognition Accuracy

The recognition accuracy was 60.8% for command input (58.8% in the PDA-setting, 64.4% in the Desktop-setting) and 50.2% for dictation (PDA-setting 60%, Desktop-setting 34%).

The poor recognition performance for dictation in the Desktop-setting in contrast to the PDA-setting comes from the fact that most test persons typed the text after recognition errors instead of giving the recogniser one more chance in the desktop setting.

Despite this low recognition performance, the proportion of speech input was surprisingly high.

4.4 Modality Distribution

The modality distribution was examined in connection with two tasks of different complexity (simpler task and more complex task). Users had to reply to an e-mail which asked them to check some other messages before composing the answer text. The users were free to interact traditionally, via speech or in a combined fashion.

The independent variables were

- the setting (Wizard of Oz vs. stand-alone),
- the scenario (desktop vs. PDA),
- the task (simpler vs. more complex), and
- the subtask (search by name, search by text, navigation, create reply message, create text, send the message).

The dependent variable was the interaction style (speech only, traditional interaction, combined interaction). Cochran's Q and pairwise McNemar tests were used in an explorative way in order to determine whether differences in the proportion of purely traditional input vs. speech-enhanced input were significant on the 5% level across variations of the independent variables.

[1]Admittedly, this correction might have lead to a bias in favour of traditional interaction.

4.4.1 Overall Distribution

In the less complicated task (wizard setting), when interacting with the PDA, twelve testpersons used only speech interaction, six subjects used only the pen, and fifteen combined pen and speech across the subtasks. When interacting with the desktop PC, eight persons used speech-only, forteen keyboard and mouse, and thirteen combined interaction (view figure 3).

In the more complex task (wizard setting), eleven test persons used only speech, seven only pen, and fifteen combined input when interacting with the PDA. When interacting with the desktop PC, eight subjects exclusively used speech, thirteen traditional, and forteen mixed input (view figure 3).

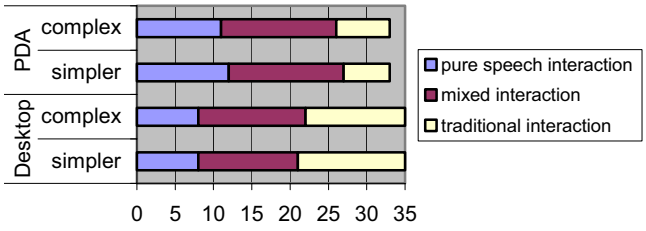

Figure 3: Modality Distribution in the Wizard Study

In the stand-alone setting, most test persons mixed speech interaction with traditional input. Only in few cases, the users gave up to use speech input (speech trial) due to low recognition accuracy or made use of traditional input from the beginning due to bad experience using speech during the test run (view figure 4).

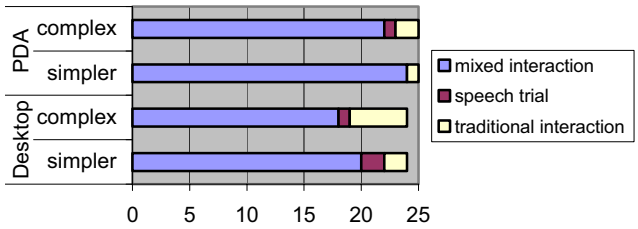

Figure 4: Modality Distribution in the Stand-alone Prototype Study

An application of Cochran's Q test revealed differences in the proportion of traditional vs. speech-based interaction styles among systems (PDA and desktop) and tasks (simple and complex) ($p(\alpha) = 0.018$ in the Wizard of Oz scenario, and $p(\alpha) = 0.024$ in the stand-alone prototype).

4.4.2 One-click vs. Text Input Actions

Cochran's Q tests on the proportion of purely traditional interaction for each subtask revealed that the proportion of purely traditionally interacting users differs significantly across the subtasks (name search, text search, reply, send, navigate):

- in the wizard setting within the simpler ($p(\alpha) = 0.001$) and more complex task ($p(\alpha) < 0.001$) of the PDA scenario and the simpler task of the desktop scenario ($p(\alpha) < 0.001$)

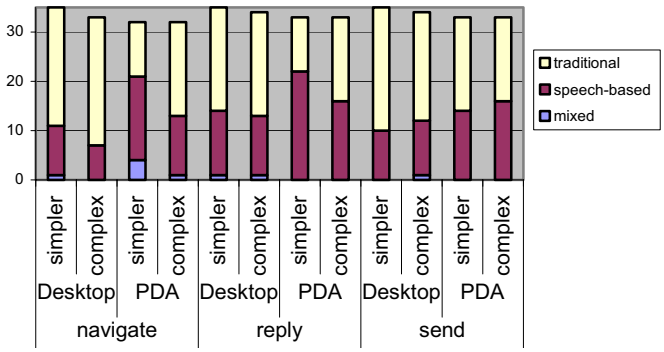

Figure 5: Modality Distribution for Simple Commands in the Wizard Study

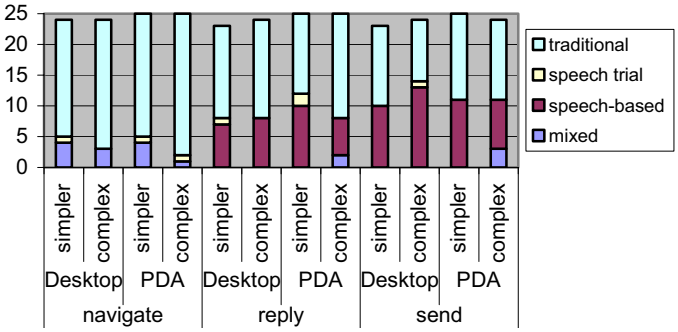

Figure 6: Modality Distribution for Simple Commands in the Stand-alone Prototype Study

- in the stand-alone setting within the simpler ($p(\alpha) = 0.002$) and more complex task ($p(\alpha) = 0.003$) of the PDA scenario and the more complex task of the desktop scenario ($p(\alpha) = 0.023$).

Explorative pairwise McNemar tests underpinned the observation that for simple (one-click) actions users interacted more frequently in a purely traditional way than for name or text searching.

Simple commands can be uttered by simply pressing the adequate button, selecting the necessary list item etc. whereas traditional text input requires typing the word letter by letter. In the case of name (sender/receiver) input, an autocompletion list helps to accelerate input. Nevertheless, most users preferred speech input in these cases. This holds for both the desktop and PDA settings.

4.4.3 Impact of Screen and Prompt Design

An application of Cochran's Q test revealed significant differences within each subtask across settings and tasks:

- in the wizard setting for the subtasks *navigation* ($p(\alpha) = 0.002$) and *answering* ($p(\alpha) = 0.029$)
- in the stand-alone setting for the subtasks *namesearch* ($p(\alpha) = 0.017$) and *textinput* ($p(\alpha) = 0.035$)

Pairwise McNemar tests underpinned the observations illustrated by figure 5: In the wizard setting the proportion of speech-related versus purely traditional interaction differed significantly:

- within the subtask *navigation*: In the simpler task the test persons used significantly more frequently speech interaction for navigating in the PDA scenario than for navigating in the desktop scenario ($p(\alpha) = 0.007$). This might be due to prompt design, which was more verbose in the PDA setting for navigation hints. Within the PDA setting, speech input for navigation was used significantly less frequently in the subsequent, more complex task than in the preceding simpler task ($p(\alpha) = 0.020$). This might be caused by some learning effect which counterbalanced the prompting-related bias towards speech input.
- wihtin the subtask *answering*: In the simpler task the test persons used significantly more frequently speech input for initializing a reply message in the PDA scenario than for doing so in the desktop-scenario ($p(\alpha) = 0.032$). This might be caused by screen design: The main window did not contain a reply button. Thus the users had to open the selected message to be able to answer them purely traditionally in the PDA setting. This extra-step could be circumvented easily using the associated speech command. Within the PDA setting, there was a significant drop in the frequency of using speech for initializing the reply message in the subsequent more complicated task in contrast to the preceding simpler one ($p(\alpha) = 0.016$).

In the simpler task of the PDA scenario (wizard setting) users interacted significantly more frequently via speech for navigating ($p(\alpha) = 0.020$) and initializing a reply message ($p(\alpha) = 0.004$) than for sending the message. No significant difference could be observed in the simple task of the desktop scenario and in the more complex task of the PDA scenario and desktop scenario.

In the stand-alone prototypes, where screen design was simplified for the PDA and spoken prompts were avoided totally, the ratio of speech vs. traditional input of simple commands seems not to differ in the desktop and PDA setting. Even in the desktop setting there is quite a high amount of test persons who appreciated to use short speech commands instead of moving hands between keyboard and mouse and moving around the mouse cursor (view figure 6).

In contrast to the wizard setting, where in the simple task of the PDA scenario speech input for navigation was used more frequently than for other simple commands, in the stand-alone setting, navigation was more frequently performed purely traditionally, which was significant throughout all tasks and scenarios:

- In the simpler task of the desktop-scenario: *navigation* vs. *send* ($p(\alpha) = 0.035$)
- In the more complex task of the desktop-scenario; *navigation* vs. *answering* ($p(\alpha) = 0.032$) and *navigation* vs. *send* ($p(\alpha) = 0.001$)
- In the simpler task of the PDA scenario: *navigation* vs. *send* ($p(\alpha) = 0.033$)
- In the more complex task of the PDA scenario: *navigation* vs. *answering* ($p(\alpha) = 0.016$) and *navigation* vs. *send* ($p(\alpha) = 0.002$)

4.4.4 Impact of Input Devices

In the stand-alone setting the proportion of speech-related versus purely traditional interaction differed significantly within the subtask *text input*: In the more complex task, far more users exclusively typed the text in the desktop scenario than in the PDA scenario ($p(\alpha) = 0.011$). Obviously, the availability of a comfortable keyboard lets the user more quickly give up speech input after bad experiences. In the desktop scenario, text editing was done significantly more frequently purely traditionally than searching by sender names in both the simpler task ($p(\alpha) = 0.032$) and the more complex task ($p(\alpha) = 0.008$).

At the same time, only in the PDA scenario, test persons used pure pen input for simple commands significantly more frequently than for editing the message text. This was significant in following cases:

- in the simpler task of the PDA scenario: *navigation* vs. *text input* ($p(\alpha) < 0.001$), *answering* vs. *text input* ($p(\alpha) = 0.011$) and *text input* vs. *send* ($p(\alpha) = 0.020$)
- and in the more complex task of the PDA scenario: *navigation* vs. *text input* ($p(\alpha) < 0.001$), *answering* vs. *text input* ($p(\alpha) = 0.003$) and *text input* vs. *send* ($p(\alpha) = 0.011$).

4.4.5 Input Forms and Free Text Input

In the wizard setting, most test persons used speech input to provide search criteria to filter the message list. This tendency seems to be strongest in the PDA system (view figure 7).

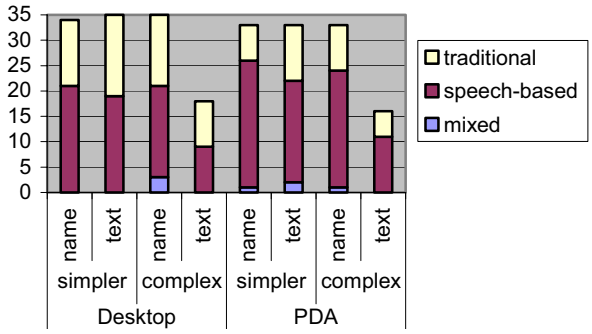

Figure 7: Modality Distribution for Searching in the Wizard Study

The stand-alone prototype allowed a more flexible combination of input styles. The observation seems to confirm the findings from the wizard setting that most people used (or tried to use) speech-based interaction styles when filtering the email list and that this ratio is even higher in the PDA setting (view figure 8). Inputting free text search criteria, on the contrary, was frequently abandonned due to poor recognition performance (and very strange substitution errors). This fact might change when personalising and training the speech recogniser.

Speech-related input for searching / filtering can be classified into *pure speech input* and *multimodal input* (selecting the form field and dictating the content). Pure speech input can be subclassified into *menu-like* (first selecting and then filling in) and *natural language* (one utterance for selecting and filling in the field) styles.

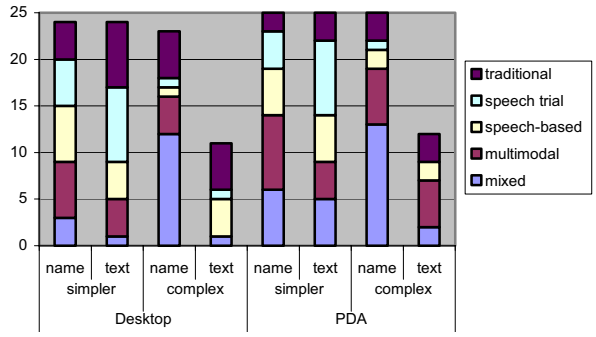

Figure 8: Modality Distribution for Searching in the Stand-alone Prototype Study

A deeper analysis shows that most of the speech-related search commands were uttered multimodally (view figure 9).

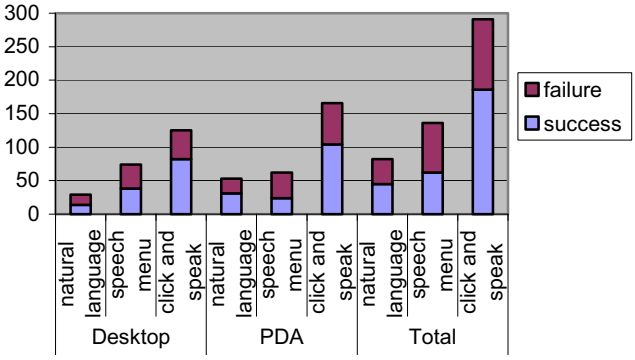

Figure 9: Interaction Styles during Searching (Stand-alone Prototype)

In the stand-alone setting the users were able to input text using speech recognition. In the PDA system only few test persons typed the text. Those who used speech-based text input mixed interaction styles, that is they dictated the content and corrected misrecognitions in a traditional or mixed way. In the more complex task, users made use of multimodal text completion (view figure 10). That is, users re-used the content from the original message and completed it with their own remarks. To do so, they pointed into the re-used text and dictated the additional comments.

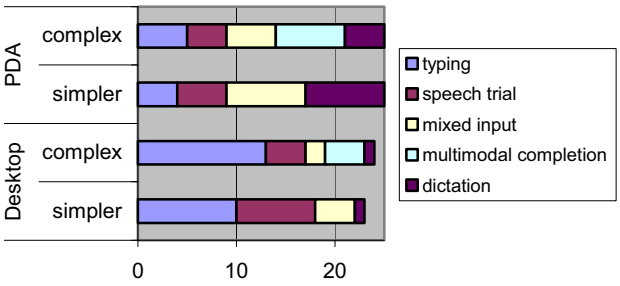

Figure 10: Free Text Input (Stand-alone Prototype)

4.4.6 Summary of General Tendencies

To summarise, following tendencies were observed:

- Users tend to use speech for text input more frequently than for selecting options that are displayed directly on screen.
- Users tend to use speech to circumvent navigation steps – to select "hidden" options (cf. *Voice-based Interaction Shortcut*).
- Novice users tend to use speech after being prompted to do so. With growing expertise, the influence of prompts is reduced.
- Users tend to deviate to traditional text input after recognition errors, especially when comfortable desktop keyboards (instead of tiny PDA on-screen keyboards) are available.
- To fill in forms, users tend to use pointing gestures for selecting the input field and speech for inputting data (cf. *Speech-enabled Form*).

4.5 Pattern-related Observations

The observations made above are interpreted in relation to the formerly identified user interface patterns.

4.5.1 Autocompletion

For name input almost every test subject made use of *Autocompletion* while interacting traditionally in the wizard setting. Only nine of 35 did not use it at first go, partially because of technical malfunctions.

4.5.2 Voice-based Interaction Shortcut

For inputting names in the wizard setting, most of the users preferred to use *Voice-based Interaction Shortcut* when having the choice between traditional and speech-based interaction styles – despite the presence of *Autocompletion* and the poor recognition performance in the stand-alone szenario. In the PDA setting, speech input seems to speed up interaction (view section 4.2).

During answering / forwarding emails with the PDA via speech input, most subjects uttered direcly the associated speech command without – which would have been necessary during pen interaction – explicitly opening the mail.

4.5.3 Composed (Spoken) Command

For forwarding an email and providing a receiver in one interaction step, only few test persons used a *Composed (Spoken) Command*. Most of the users took two interaction steps instead. The same is true for inputting search criteria (view figure 9, section 4.4).

Might be, with increasing expertise more users are willing to make use of spoken composed commands.

4.5.4 Speech-enabled Form

For filtering messages, the most frequently used interaction style in the stand-alone prototype was the multimodal one: The users selected the necessary form field with pen gestures and dictated the desired content (view figure 9, section 4.4).

5. CONCLUSIONS

This paper presents two empirical studies on multimodal interaction for an e-mail task. Two scenario prototypes (PDA, desktop) were evaluated in two settings (Wizard of Oz, stand-alone).

The results indicate that the availablility of speech input increases interaction speed and leads to higher user satisfaction, especially in mobile settings. At the same time, it is shown that modality preference is related strongly to the respective subtask (text input, simple action commands etc.).

The findings underlay the application of traditional user interface patterns as well as of user interface patterns for multimodal interaction identified in earlier research steps.

6. ACKNOWLEDGMENTS

I would like to thank my supervisors professor Rainer Hammwöhner and professor Christian Wolff for making possible this research and for their numerous comments and feedback. I would like to thank Sarah Will for assisting me during the user tests.

Furthermore, I would like to thank the DFG (German Research Foundation) for funding this research trip to ICMI.

7. REFERENCES

[1] Mayhew, D.J.: The Usability Engineering Lifecycle. Morgan Kaufmann, San Francisco (1999)

[2] Ratzka, A., Wolff, C.: A pattern-based methodology for multimiodal interaction design. In Sojka, P., Kopeček, I., Pala, K., eds.: Proc. of Text, Speech, and Dialogue, TSD'06. LNAI 4188, Berlin, Heidelberg, Springer (2006) 677–686

[3] Ratzka, A.: Design patterns in the context of multi-modal interaction. In: To appear in: Proceedings of the 6th Nordic Conference on Pattern Languages of Programs 2007 VikingPLoP 2007. (2008)

[4] Tidwell, J.: Common ground: A pattern language for human-computer interface design. (1999)

[5] Tidwell, J.: Designing Interfaces: Patterns for Effective Interaction Design. O'Reilly (2005)

[6] van Welie, M., Trætteberg, H.: Interaction patterns in user interfaces. In: Proceedings of the Seventh Pattern Languages of Programs Conference. (2000)

[7] Ratzka, A.: A wizard-of-oz setting for multimodal interaction. an approach to user-based elicitation of design patterns. In Osswald, A., Stempfhuber, M., Wolff, C., eds.: Open Innovation. Proc. 10th International Symposium for Information Science, Konstanz, Universitätsverlag Konstanz (2007) 159–170

[8] Niedermaier, F.B.: Entwicklung und Bewertung eines Rapid-Prototyping Ansatzes zur multimodalen Mensch-Maschine-Interaktion im Kraftfahrzeug. PhD thesis, Fakultät für Elektrotechnik und Informationstechnik der Technischen Universität München (2003)

[9] Dragičević, P.: Un modèle d'interaction en entrée pour des systèmes interactifs multi-dispositifs hautement configurables. PhD thesis, Université de Nantes, école doctorale sciences et technologies de l'information et des matérieaux (Mars 2004)

[10] Bernsen, N.O.: A toolbox of output modalities: Representing output information in multimodal interfaces. (1997)

[11] Bernsen, N.O.: Multimodality in language and speech systems – from theory to design support tool. In Granström, B., ed.: Multimodality in Language and Speech Systems, Dordrecht, Kluwer (2001)
[12] Seifert, K.: Evaluation multimodaler Computer-Systeme in frühen Entwicklungsphasen. PhD thesis, Fakultät V – Verkehrs- und Maschinensysteme, Technische Universität Berlin (2002)
[13] Alexander, C.: The Timeless Way of Building. Oxford University Press (1979)
[14] Gabriel, D.: A pattern definition. http://hillside.net/patterns/definition.html (checked on 2007-05-09)
[15] Alexander, C., Ishikawa, S., Silverstein, M., Jacobson, M., Fiksdahl-King, I., Angel, S.: A Pattern Language. Oxford University Press (1977)
[16] Gamma, E., Helm, R., Johnson, R., Vlissides, J.: Design Patterns: Elements of Reusable Object-Oriented Software. Addison-Wesley (1995)
[17] Buschmann, F., Meunier, R., Rohnert, H., Sommerlad, P., Stal, M.: Pattern-orientierte Softwarearchitektur. Addison-Wesley, Bonn (1998)
[18] Borchers, J.O.: A pattern approach to interaction design. AI & Society Journal of Human-Centered Systems and Machine Intelligence **15**(4) (2001) 359–376
[19] Duyne, D.K.V., Landay, J., Hong, J.I.: The Design of Sites: Patterns, Principles, and Processes for Crafting a Customer-Centered Web Experience. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA (2002)
[20] van Welie, M.: Gui design patterns. In: www.welie.com – ...patterns in Interaction Design. (2003)
[21] Hauptmann, A.G.: Speech and gestures for graphic image manipulation. In: CHI '89: Proceedings of the SIGCHI conference on Human factors in computing systems, New York, NY, USA, ACM (1989) 241–245
[22] Oviatt, S., Olsen, E.: Integration themes in multimodal human-computer interaction. In Shirai, K., Furui, S., Kakehi, K., eds.: Proceedings of the International Conference on Spoken Language Processing. Volume 2., Yokohama, Japan, Acoustical Society of Japan (1994)
[23] Oviatt, S., Cohen, P.R., Wang, M.Q.: Toward interface design for human language technology: modality and structure as determinants of linguistic complexity. Speech Commun. **15**(3-4) (1994) 283–300
[24] Angeli, A.D., Gerbino, W., Cassano, G., Petrelli, D.: Visual display, pointing, and natural language: the power of multimodal interaction. In: AVI '98: Proceedings of the working conference on Advanced visual interfaces, New York, NY, USA, ACM Press (1998) 164–173
[25] Oviatt, S.: Multimodal interfaces for dynamic interactive maps. In: CHI '96: Proceedings of the SIGCHI conference on Human factors in computing systems, New York, NY, USA, ACM (1996) 95–102
[26] Cohen, P., Johnston, M., McGee, D., Oviatt, S.L., Clow, J., Smith, I.: The efficiency of multimodal interaction: A case study. In: ICSLP'98. (1998)
[27] Cohen, P., McGee, D., Clow, J.: The efficiency of multimodal interaction for a map-based task. In: Proceedings of the sixth conference on Applied natural language processing, San Francisco, CA, USA, Morgan Kaufmann Publishers Inc. (2000) 331–338
[28] Oviatt, S., Kuhn, K.: Referential features and linguistic indirection in multimodal language. In: Proceedings of the International Conference on Spoken Language Processing. Volume 6., ASSTA (1998) 2339–2342
[29] Oviatt, S.: Ten myths of multimodal interaction. Commun. ACM **42**(11) (1999) 74–81
[30] Oviatt, S.: Mutual disambiguation of recognition errors in a multimodal architecture. In: CHI '99: Proceedings of the SIGCHI conference on Human factors in computing systems, New York, NY, USA, ACM (1999) 576–583
[31] Oviatt, S.: Multimodal system processing in mobile environments. In: UIST '00: Proceedings of the 13th annual ACM symposium on User interface software and technology, New York, NY, USA, ACM Press (2000) 21–30
[32] Oviatt, S.L.: Multimodal signal processing in naturalistic noisy environments. In Yuan, B., Huang, T., Tang, X., eds.: Proceedings of the 6th International Conference on SPoken Language Processing (ICSLP). Volume 2., Peking, Chinese Friendship Publishers (2000) 696–699
[33] Oviatt, S.: Taming recognition errors with a multimodal interface. Commun. ACM **43**(9) (2000) 45–51
[34] Kumar, S., Cohen, P.R., Coulston, R.: Multimodal interaction under exerted conditions in a natural field setting. In: ICMI '04: Proceedings of the 6th international conference on Multimodal interfaces, New York, NY, USA, ACM Press (2004) 227–234
[35] Nishimoto, T., Shida, N., Kobayashi, T., Shirai, K.: Improving human interface in drawing tool using speech. In: Proceedings of 4th IEEE International Workshop on Robot and Human Communication, ROMAN'95. (1995) 107–112
[36] Huang, X., Acero, A., Chelba, C., Deng, L., Duchene, D., Goodman, J., Hon, H., Jacoby, D., Jiang, L., Loynd, R., Mahajan, M., Mau, P., Meredith, S., Mughal, S., Neto, S., Plumpe, M., Wang, K., Wang, Y.: Mipad: A next generation pda prototype. In: ICSLP, Peking (2000)
[37] Lai, J.: Facilitating mobile communication with multimodal access to email messages on a cell phone. In: CHI '04: CHI '04 extended abstracts on Human factors in computing systems, New York, NY, USA, ACM Press (2004) 1259–1262
[38] Jöst, M., Häußler, J., Merdes, M., Malaka, R.: Multimodal interaction for pedestrians: an evaluation study. In: IUI '05: Proceedings of the 10th international conference on Intelligent user interfaces, New York, NY, USA, ACM Press (2005) 59–66