The Influence of Expertise and Efficiency on Modality Selection Strategies and Perceived Mental Effort

Ina Wechsung[1], *Stefan Schaffer*[2], *Robert Schleicher*[1], *Anja Naumann*[1], *Sebastian Möller*[1]

[1]Deutsche Telekom Laboratories, Quality & Usability Lab, TU Berlin, Germany
[2]Research training group prometei, TU Berlin, Germany

{ina.wechsung, stefan.schaffer, robert.schleicher, anja.naumann, sebastian.moeller}@telekom.de

Abstract

This paper describes a user study investigating the influence of expertise and efficiency on modality selection (speech vs. virtual keyboard) and perceived mental effort. Efficiency was varied in terms of interaction steps. The goal was to investigate if the number of necessary interaction steps determines the preference for a specific modality. It is shown that the threshold for changing the modality selection strategy is at 3 (experts) respectively 4 (novices) interaction steps.

Index Terms: modality selection, multimodal systems

1. Introduction

Presenting information and offering interaction via multiple modalities is assumed to increase a system's quality. Studies showed multimodal interfaces to be more robust, efficient and flexible than unimodal systems [1]. Most of these studies were focusing on spatial-verbal tasks investigating parallel speech and pen input or on the effects of multimodal and/or multimedia output [2], [3]. For these tasks and systems the benefits of multimodality and modality selection strategies are well documented. However, for different tasks and multimodal systems using input modality combinations other than speech and pen, or for systems offering serial (not parallel input) the findings are less clear [4].

Furthermore multimodality may have negative effects: The decision process involved in choosing the "appropriate" modality may increase cognitive load [4] and, moreover, the modalities may even interfere [6] with each other. Research showed that although multimodality may lead to higher perceived quality, users often do not use multimodal interaction strategies and are instead sticking to one input modality [7],[8]. Although some studies provide ambiguous results, a clear advantage of speech could rarely be observed. A preference for speech was only shown for situations were the visual channel was already occupied. For example Lemmelä et al. [9] showed that speech is preferred in an in-car scenario and [10] reported speech to be preferred by the users and to be more efficient than the other input modalities in situations with limited visual feedback. However, if none of the human perception channels is busy due to the situation, the majority of studies shows that the most frequently chosen and often also best rated modality is input via the touch screen , e.g. [7], [8], [11], [12], [13], [14], [15] and [16].

Interestingly, in most of these studies an increase of speech input usage was observed for specific tasks. For example speech was preferred for selecting an element out of a very long list [16], for entering a long telephone number [8], or for searching for specific titles [7], [8], [12], [15]. All these tasks have in common that speech input offered a short-cut in terms of a reduced number of necessary interaction steps compared to the other input modalities. Thus, the preference of one modality over another seems to be highly dependent on the efficiency of that very modality. This assumption is supported by [17] comparing text-input to speech input: Speech input was reported to be more efficient and more preferred. It has to be noted that efficiency was not systematically varied in [17]. The conclusion of speech being preferred if more efficient was based on the observation that numbers were less likely to be entered via speech than input requiring more keystrokes. These findings are opposed by [18] where speech input was the less efficient but most preferred modality compared to keyboard and a scrolling bar. However, efficiency was measured in task-completion time and not in interaction steps. Hence the explanation of the results presented in [18] might be that speech actually was more efficient in terms of interaction steps.

Thus, although research indicates that shortcuts have an influence on modality choice, a systematic investigation is still missing. It is currently not clear how many interactions steps a shortcut must offer to skip in order to lead to a change in modality selection strategies. The aim of this paper is to investigate this threshold. Furthermore, the influence of efficiency on perceived mental effort was examined.

Since differences in interacting strategies between expert and novice users are well documented (e.g. [19], [20], [21]), expertise was taken into account as an additional variable.

2. Method

2.1. Participants

Twenty-six German-speaking subjects voluntarily participated in our study. All participants were either PhD students or student workers in the area of human-computer-interaction. Thus all of them were assumed to be expert users regarding Information and Communication Technology. One of the participants was excluded from further analysis as he did not follow the experimenter's instructions. The remaining twenty-five participants were between twenty-two and thirty-nine years of age (M = 29 years, SD = 3, 7 female). Twelve of them were experts regarding speech dialogue systems, working or having previously worked in the area of speech recognition or voice user interfaces design and had frequently been using such systems. The other thirteen participants were novice users with a different research background (e.g. tactile interaction, brain-computer interfaces). These users have never or only very seldom been using such systems. All of them were familiar with virtual keyboards and regularly using them.

Copyright © 2010 ISCA

2.2. Material

The application, a mobile jukebox, was installed on an Android-based smart-phone (G1 HTC Dream). The available input modalities were touch using the virtual keyboard and speech (Nuance Vocon). Speech recognition was started via a push-to-talk button. An acoustic signal, a short beep, is indicated that the speech recognizer was activated. The end of the utterance was detected automatically by the recognizer. If more than one result was found in the grammar, an n-best list is presented. If only one artist was recognized, a direct forwarding to the available tracks and albums was presented. If no artist was found, a pop-up saying "I did not understand you" was presented.

For input via the virtual keyboard only two scenarios were possible: Either the direct forwarding took place or a pop-up saying "Nothing found" was shown, as substring search (e.g. just typing QUE instead of QUEEN) were not supported and the systems had zero error tolerance. All misspelled search request were rejected.

The participants' task was to search for ten different artists. The length of the artists' names increased for one character for each task. The shortest artist name had two characters, the longest had eleven.

To measure perceived mental effort, the SEA-scale [19], the German version of the rating scale mental effort (also known as subjective mental effort questionnaire) was employed.

Figure 1*: Tested device and application*

2.3. Procedure

The participants were first asked to answer demographic questions. Although all individuals were recruited based on their research background, we asked them to self-assess their expertise with spoken dialogue systems. Only subjects were objective criteria and self-assessed expertise matched, were entitled experts.

Next, the device, the virtual keyboard, the speech control and the tasks were explained. For each task the artists were presented in written and oral form. Every participant received the same order of tasks, thus the test always started with the shortest name (2 characters) and ended with the longest (11 characters). The task was accomplished if the artist was in the results list or if the page displaying the artist's albums and tracks was shown. If the artist was not in the list the participants had to search again. This was repeated until the name was in the list, or respectively, until the artist's page was shown. After each accomplished task the SEA-scale had to be filled in. The participants could choose and switch the input modality at any time. The modality switches and the number of trials were logged.

3. Results

3.1. Perceived mental effort

Regarding the perceived overall mental effort (averaged over all tasks), differences between expert and novice users were not observed. When analyzing the single tasks differences between the user groups were found for task 7 (8 characters, t (23) = 1.92, p =.034) and task 10 (11 characters, t (17.57) = 2.53, p =.011). Novice users reported higher mental effort than expert users (cf. Figure 2).

Differences were also found between the tasks: Task 6 (7 characters) was perceived as most demanding (F (4.25, 97.83) =4.28, p=.002, cf. Figure 2).

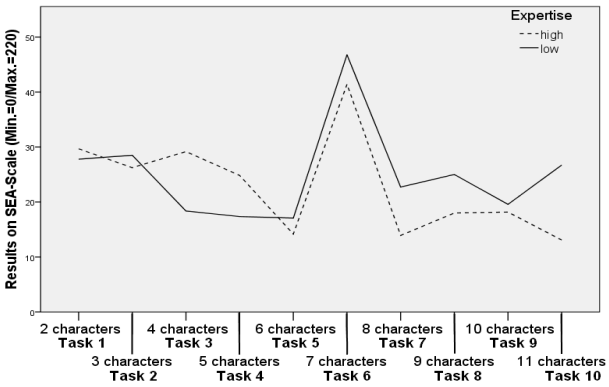

Figure 2*. Perceived mental effort by expertise and task*

3.2. Number of Attempts

Over all tasks, expertise had no influence. Besides the minimal number of ten attempts, one for each task, both experts and novices needed three additional attempts (M_{expert}=2.83, SD_{expert}=1.70, M_{novice} = 2.69, SD_{novoce}=2.14). An analysis for each single task revealed differences for task 3 (t(13.55) = 1.85, p=.043) and task 9 (t(11) = 1.92, p=.041): Surprisingly the experts needed more attempts than the novice users (cf. Figure 3).

As for mental effort, differences between the tasks were found for task 6, requiring the most attempts and thus being most error prone (F (5.40, 124.28) =3.06, p=.05, cf. Figure 3).

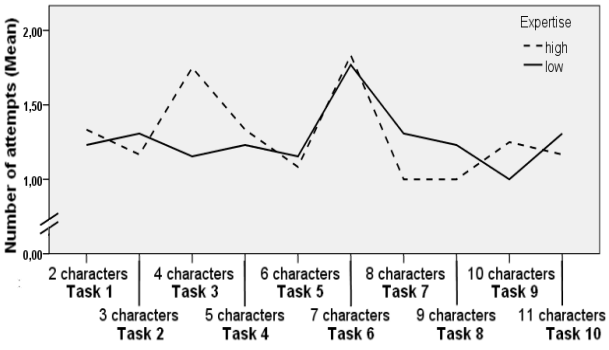

Figure 3. *Number of attempts by expertise and task*

3.3. Correlations between Mental Effort, Number of Attempts and Number of Characters

The number of attempts and the perceived mental effort were significantly correlated (Pearson's $r_{overall}$ (248) = .550, p<0.00, Pearson's r_{novice} (118) = .525, p<.000, Pearson's $r_{experts}$(128) = .582, p<.000). Tasks requiring more trials were rated as more demanding. No correlation was shown for the number of characters.

3.4. Interaction Steps and Modality Selection

The modality selection strategy for the first attempt did not differ between the first three tasks. For task 4 (5 charaters) speech usage increased and stayed similar for the last six tasks (cf. Figure 4). Thus, if the speech short-cut offered to skip at least four interaction steps the modality selection strategy changed resulting in an increase of speech usage (McNemar Test: χ^2 (1, N=25) = 2.77, p=.002).

A separate analysis of expert and novice users revealed that for experts the threshold is one interaction step lower: The shortcut needs to offer only three interaction steps that can be skipped to significantly change the modality preference (cf. Table 1). For novice users the results are ambiguous. Although the modality selection strategy switched after task 3 (4 characters), another switch was observed before and after task 6 (7 characters).

Task	Expert			Novice		
	% Speech (N)	% Touch (N)	χ^2 (p)	% Speech (N)	% Touch (N)	χ^2 (p)
1 (2 char.)	58.3 (7)	41.7 (5)	.333 (.387)	53.8 (7)	46.2 (6)	.077 (.500)
2 (3 char.)	66.7 (8)	33.3 (4)	1.333 (.194)	69.2 (9)	30.8 (4)	1.923 (.134)
3 (4 char.)	83.3 (10)	16.7 (2)	5.333 (.015)*	61.5 (8)	38.5 (5)	.692 (.291)
4 (5 char.)	100 (12)	0 (0)	-	84.6 (11)	15.4 (2)	6.231 (.011)*
5 (6 char.)	100 (12)	0 (0)	-	84.6 (11)	15.4 (2)	6.231 (.011)*
6 (7 char.)	100 (12)	0 (0)	-	69.2 (9)	30.8 (4)	1.923 (.134)
7 (8 char.)	100 (12)	0 (0)	-	84.6 (11)	15.4 (2)	6.231 (.011)*
8 (9 char.)	91.7 (11)	8.3 (1)	8.333 (.003)*	84.6 (11)	15.4 (2)	6.31 (.011)*
9 (10 char.)	100 (12)	0 (0)	-	76.9 (10)	23.1 (3)	3.769 (.046)*
10 (11 char.)	100 (12)	0 (0)	-	76.9 (10)	23.1 (3)	3.769 (.046)*

Table 1. *Results for modality selection strategies by expertise and task, * p<.05.*

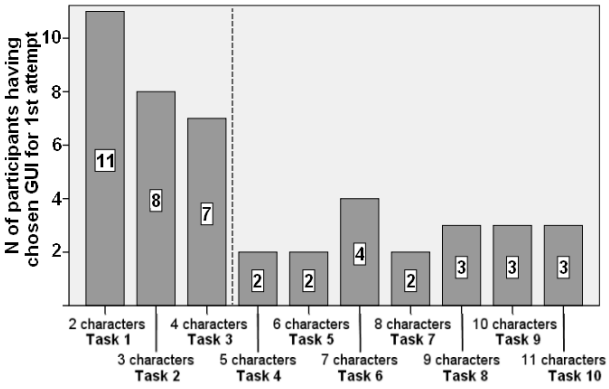

Figure 4. *Number of participants choosing GUI (touch) for 1st attempt by task. Dotted line displays tasks differing in modality selection strategy.*

3.5. Number of Trials and Modality Switches

The majority of participants did not change the input modality after the first unsuccessful attempt. Not until the second unsuccessful attempt did a majority of participants switch the input modality. If a third attempt was necessary, nearly all participants switched the modality. This pattern was the same for expert and novice users (cf. Table 2). Interestingly none of the novice users needed a fourth attempt.

Table 2. *Modality switches for additional attempts by expertise.*

% of Modality Switches	Expert		Novice	
	M	SD	M	SD
1st to 2nd attempt	27.27	38.92	40.15	42.30
2nd to 3rd attempt	75.00	41.83	64.29	47.56
3rd to 4th attempt	100	0	-	-

4. Discussion and Conclusion

The results show that perceived mental effort is hardly affected by expertise but strongly related to task success.

Also for modality selection, the number of necessary interaction steps is more relevant than expertise. In more detail, the results indicate that if at least three to four interaction steps can be skipped, speech is strongly preferred over touch. An influence of effectiveness and robustness on modality selection could be shown but only if a modality failed more than twice. Thus in the current study the most important factor is the number of interactions steps. Based on these results offering speech as an input modality is especially useful when providing shortcuts skipping more than three to four interaction steps. Or the other way around, if speech does not offer shortcuts users will probably not interact multimodally. However, these results are so far limited to the tested system. The used speech recognizer had fairly good recognition accuracy (~77%) and the virtual keyboard was relatively small. Thus, with less reliable speech recognition error avoidance would have become more important to the participants and the influence of effectiveness and robustness on modality selection would have been higher. As physical keyboards [23] are more usable than virtual ones the threshold might be different if speech would have been compared to the physical keyboard.

Moreover, only text and not numbers had to be entered. Since research reports that numbers are less likely to be entered [17] via speech the results may have been different for telephone numbers instead of artists.

After switching the modality once (from keyboard to speech), participants tend to stick to that modality, thus aiming for some kind of internal interaction consistency.

A critical point here is that participants could anticipate the increase of necessary interactions steps as the order of tasks was not randomized and the length of the artists' names increased by one character for each task. If this increase would have been less apparent the threshold probably would not have been this clear. However, the preference for speech for longer input is quite obvious. For tasks that are likely to require longer input like track or album titles, a subtle reminder of the speech option (e.g. a microphon icon) might encourage untrained users to switch to this modality.

5. Future Work

In a follow-up study the physical keyboard will also be offered as an input modality and the number of necessary interaction steps will not constantly increase by one but be randomized in order to see if the results of the present study can be verified. Also the number of errors will systematically be varied to investigate the relationship between efficiency and robustness regarding the influence on modality selection.

6. References

[1] Oviatt, S. Multimodal interfaces. In the Human-Computer interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications, In J. A. Jacko and A. Sears, Eds. Human Factors And Ergonomics. L. Erlbaum Associates, Hillsdale, NJ, 286-304, 2002.

[2] Elting, C., Zwickel, J., and Malaka, R. Device-dependant modality selection for user-interfaces: an empirical study. In Proceedings of the 7th international Conference on intelligent User interfaces. ACM, New York, NY, 55-62, 2002.

[3] Oviatt, S., Cohen, P., Wu, L., Vergo, J., Duncan, L., Suhm, B., Bers, J., Holzman, T., Winograd, T., Landay, J., Larson, J., and Ferro, D. Designing the user interface for multimodal speech and pen-based gesture applications: state-of-the-art systems and future research directions. Human-Computer Interaction 15 (4) 263-322, 2000.

[4] Sarter, N. Multiple-resource theory as a basis for multimodal interface design: Success stories, qualifications, and research needs. In Kramer, A., Wiegmann, D., and Kirlik, A. (Eds.), Attention: From Theory to Practice. Oxford: Oxford University Press, 187-195, 2006.

[5] Schomaker, L., Nijtmans, J., Camurri, A., Lavagetto, F., Morasso, P., Benoit, C., Guiard-Marigny, T., LeGoff, B., Robert-Ribes, J., Adjoudani, A., Defee, I., Munch, S., Hartung, K. and Blauert, J. A Taxonomy of Multimodal Interaction in the Human Information Processing System. A Report of the ESPRIT Project 8579 MIAMI. NICI, Nijmegen, 1995.

[6] Schnotz, W. Bannert, M. and Seufert, T. Towards an integrative view of text and picture comprehension: Visualization effects on the construction of mental models, in: J. Otero, A. Graesser & J. A. Leon (Eds.), The Psychology of Science Text Comprehension. Erlbaum, Mahwah, 385-416, 2002.

[7] Wechsung, I., Anja B. Naumann and Hurtienne, J. Multimodale Interaktion: Intuitiv, robust, bevorzugt und altersgerecht? [Multimodal Interaction: Intuitive, robust, preferred and senior-friendly?]. In Proceedings of Mensch und Computer 2009. 9. fachübergreifende Konferenz für interaktive und koooperative Medien - Grenzenlos frei 2009. Oldenbourg, 213-222, 2009.

[8] Naumann, A., Wechsung, I. and Möller, S. Factors Influencing Modality Choice in Multimodal Applications. Perception in Multimodal Dialogue Systems, 4th IEEE Tutorial and Research Workshop on Perception and Interactive Technologies for Speech-Based Systems. Berlin: Springer, 37–43, 2008.

[9] Lemmelä, S., Vetek, A., Mäkelä, K., and Trendafilov, D. Designing and evaluating multimodal interaction for mobile contexts. In Proceedings of the 10th international Conference on Multimodal interfaces. ACM, New York, 265-272, 2008.

[10] Cox, A.L., Cairns, P.A., Walton, A., Lee, S. Tlk or txt? Using voice input for SMS composition. Personal Ubiquitous Computing 12 (8), 567-588, 2008.

[11] Seebode, J., Schaffer, S., Wechsung, I. and Metze, F. (2009). Influence of Training on Direct and Indirect Measures for the Evaluation of Multimodal Systems. Proccedings of the 10th Annual Conference of the International Speech Communication Association, xxx-xxx, 2009.

[12] Metze, F., Wechsung, I., Schaffer, S., Seebode, J. and Möller, S. Reliable Evaluation of Multimodal Dialog Systems. Proceedings of the 13th International Conference on Human-Computer Interaction. Berlin: Springer, 75–83, 2009.

[13] Raisamo, R. Evaluating different touch-based interaction techniques in a public information kiosk. Report A-1999-11, Department of Computer Science, University of Tampere, 1999.

[14] Janienke Sturm, Bert Cranen, Jacques Terken and Ilse Bakx. Effects of Prolonged Use on the Usability of a Multimodal Form-Filling Interface. In: W. Minker, D. Bühler, and L. Dybkjær (Eds.). Spoken Multimodal Human-Computer Dialogue in Mobile Environments. Dordrecht, Kluwer Academic Publishers, 329-348, 2004.

[15] Wechsung, I., and Naumann, A.B. Bewertung eines multimodalen Systems in verschiedenen Testsituationen [Evaluation of a multimodal system in different test situations] Fortschritte der Akustik – DAGA 2010: Plenarvortr. u. Fachbeitr. d. 36. Dtsch. Jahrestg. f. Akustik, 2010.

[16] Lamel, L., Bennacef, S., Gauvain, J. L., Dartigues, H., and Temem, J. N. User evaluation of the MASK kiosk. Speech Commuication. 38 (1) 131-139, 2002.

[17] Bilici, V. Krahmer E., Riele, S., and Veldhuis, R. Preferred Modalities in Dialogue Systems. Sixth International Conference on Spoken Language Processing, 727--730, 2000.

[18] Rudnicky, A. I. Mode preference in a simple data-retrieval task. Proceedings of the Workshop on Human Language Technology, Human Language Technology Conference. Association for Computational Linguistics, 364-369, 1993.

[19] Kamm, C., Litman, D., and Walker, M. From novice to expert: the effect of tutorials on user expertise with spoken dialogue systems. Proccedings of the 9th Annual Conference of the International Speech Communication Association, 2008.

[20] Lazonder, A. W., Harm, J. A., and Woperies, I. J. Differences between novice and expert users in searching information on the world wide web. Journal of the American Society for Information Science, 51 (6), 576–581, 2000.

[21] Petrelli, D., De Angeli, A., Gerbino, W., and Cassano, G. Referring in multimodal systems: the importance of user expertise and system features. In Referring Phenomena in A Multimedia Context and their Computational Treatment, ACL, Morristown, 14-19, 1997.

[22] Eilers, K., Nachreiner, F., and Hänecke, K. Entwicklung und Überprüfung einer Skala zur Erfassung subjektiv erlebter Antrengung [Development and evaluation of a scale to assess subjectively perceived effort]. Zeitschrift für Arbeitswissenschaft, 40, 215–224, 1986.

[23] Hoggan, E., Brewster, S. A., and Johnston, J. Investigating the effectiveness of tactile feedback for mobile touchscreens. In Proceeding of the Twenty-Sixth Annual SIGCHI Conference on Human Factors in Computing Systems (CHI '08), ACM, New York, 1573-1582, 2008.