Modality Preferences of Different User Groups

Benjamin Weiss, Sebastian Möller and Matthias Schulz
Quality and Usability Lab
Dt. Telekom Laboratories, Technische Universität
Berlin, Germany
[*BWeiss* | *Sebastian.Moeller* | *Matthias-Schulz*] *@tu-berlin.de*

***Abstract*—In order to examine user group differences in modality preferences, participants of either gender and two age groups have been asked to rate their experience after interacting with a smart-home system offering unimodal and multimodal input possibilities (voice, free-hand gesture, smartphone touch screen). Effects for gender, but not for age (younger and older adults) have been found for modality preferences. Women prefer touch and voice over gesture for many scales assessed, whereas men do not show this pattern consistently. Instead, they prefer gesture over voice for hedonic quality scales. Comparable results are obtained for technological expertise assessed individually. This interrelation of gender and expertise could not be solved and is discussed along with consequences of the results obtained.**

Keywords-multimodal dialog system; evaluation; user factors.

I. INTRODUCTION

Current evaluations of multimodal interfaces already take into account user groups: Differences in users' interactive and rating behavior is analyzed regarding e.g., gender, age, user's experience with a system, etc. Unfortunately, the attitudes and expectations people have towards such systems are not well known yet [1]. Even more, expectations concerning novel multimodal application seem not to be that relevant for the actual user experience [1][2][3]. Modality preference and selection are dependent on task and efficiency [4], but general user expertise [5] also has to be taken into account.

Age and gender effects, for instance, are rarely examined together, with [3][6] as notable exceptions. While in most studies gender is balanced but not looked into further, studies on modality preferences are often limited to younger adults (e.g., [1][7]). Studies including older adults are mostly focusing on assisting technologies to support independent living [8][9][10], but age does not necessarily limit the number of products used [11]. For example, home entertainment and control is one of the major application domains for HCI and also is in the focus of this paper.

Exploring strategies for including older users, multimodality and touch were found to be more suitable than speech and motion control [6]. Furthermore, "older participants used the flexibility offered by the multiple input modalities to a lesser extent than younger users did" [6].

Comparing pointing times on a graphical user interface (GUI) using a mouse or touch-panel no significant difference between younger, middle-age and older adults was found for touch in contrast to mouse control [12]. The authors conclude that touch interfaces should be pursued to make information technology accessible to older adults.

Experience, although an established feature [13], is typically not a dimension to separate user groups in the field of multimodal systems. Multimodal interfaces are typically innovative and therefore performing evaluation experiments to compare trained versus novice users does not seem to be mandatory. Instead, general technological affinity is assessed in order to analyze this factor.

The aim of this paper is to have a closer look at the interaction between age and gender, as especially for age effects on rating behavior can be expected on the basis of the literature referred to, i.e. overall positive results for older adults [3][6] and gesture preference for younger adults [3]. But, we also want to look into other user differences and their interaction with age or gender. For this purpose, a small battery of assessments has been conducted in order to assess various aspects of technological affinity. For the domain of home entertainment and control user modality preferences (speech, 3d gesture, touch) are analyzed in order to find relevant user attributes to correlate and explain user modality preferences. After presenting the system used, we describe the experimental design, and results of the assessments, as well as the user ratings of the multimodal interaction session and the comparison of the ratings of the unimodal interaction sessions.

II. MULTIMODAL SYSTEM

For the experimental study, a smart-home system was used offering sequential use of voice, smartphone-based input (touch) and three-dimensional free-hand gesture control (gesture). This system is set up inside a fully functional living room. Possible interactions include the control of lamps and blinds, the TV, an IP-radio, an electronic program guide (EPG), video recorder, and a hi-fi system. Furthermore, the system offers an archive for music and supports the generation of playlists. The TV program, available radio stations, lists of recorded movies, an overview of the users' music (sorted by album, artist, etc.) or the playlist are displayed on the TV screen (cf. Figure 1). Those lists are also displayed on the smartphone to allow touch input for

 ISBN: 978-1-61208-177-9

Figure 1. Screen shot of information displayed on the TV screen.

(a) Main Screen (b) TV control (c) EPG screen

Figure 2. Screen shots of the smartphone display.

the selection of list entries and the execution of subsequent actions, such as recording a movie or deleting a song from the playlist (cf. Figure 2).

A German male voice was chosen for the TTS system. Thus, three different output channels are employed (TTS, GUI on the touch screen, and lists on the TV screen), in some cases offering complementary or redundant information in parallel. In order to keep input accuracy comparably high for all modalities, speech recognition was replaced by a transcribing wizard. Participants were told that there was a speech recognizer in place, and a lapel microphone was used to further strengthen this impression.

A simple graphical user interface was developed and implemented on an Apple iPhone 3GS, which communicated via wireless LAN with the smart-home system. To control the lamps, blinds, radio and TV the corresponding button on the main screen had to be pressed with a fingertip. This opens a list of options available for the respective device. Further buttons open a music archive, a music playlist, or an overview of recorded movies. List navigation was possible via scrolling (slide finger across screen) and selecting (touching an entry).

A camera-based gesture recognition for simple and often used gestures (TV, radio, lamps, blinds) was simulated by placing two cameras in front of the participants at a distance of approximately two meters, and below the TV screen. The actual recognition was done by the wizard who could monitor the participant via the cameras and enter the recognition result as attribute-value pairs (e. g. [device:blinds; action:down]) into the system. Each fourth of the participants was presented a system with either perfect recognition rate (due to the wizard), reduced speech recognition (10% error rate), reduced gesture recognition (10%) or both (gender and age group balanced).

A set of five three-dimensional gestures was used in this experiment (see Table I). By pointing towards a device with the hand this device is selected. The same gesture could thus be used to initiate the same effect for different devices. Reusing the same concept for different system controls reduces the gesture set considerably. For more detailed information please refer to [14].

Table I
GESTURE-COMMAND MAPPING.

Gesture	Command	Device
Swing up	Volume up	TV, Radio
	Brighter	Lamps
	Open	Blinds
Swing down	Volume down	TV, Radio
	Dim	Lamps
	Close	Blinds
Point forward	Turn on/off	TV, Radio
Swing to the right	Next channel	TV, Radio
	Stop	Blinds
Swing to the left	Previous channel	TV, Radio

III. EXPERIMENT

A. Participants

17 young and 17 older adults were asked to participate in the study. For the analysis, data from two subjects (one older male and one younger female adult) has been excluded, as one (younger adult) immediately recognized the Wizard-of-Oz scenario, whereas another (older adult) experienced an unstable system. This results in a group of 16 younger participants (20–29 years, M=24, SD=2.7, 8 female), who have been recruited at the university campus. The 16 older participants (51-67 years, M=59, SD=4.6, 9 female) were selected to also represent the target group of the home entertainment and device control system and thus did not exhibit physical or cognitive disabilities, which would result in special technical requirements. Therefore, they were recruited via notices placed, e.g., in supermarkets. All subjects were paid for their participation. None of the participants was familiar with the system used in this study.

B. Procedure

The experiment was split into four parts:

A: Judgment of the system output (passive scenario)
B: Judgment of the unimodal input (3 interactive scenarios)
C: Judgment of the multimodal input (interactive scenario)
D: Battery of user related assessments

 ISBN: 978-1-61208-177-9

In the first part (Part A), participants were asked to rate each of the three different output channels (TTS, touch screen and TV screen) after the presentation of a series of three to seven examples of one output channel. According to [15], it is sufficient to show a web page for less than one second to judge its aesthetics. Thus, each interface was presented only very shortly to the participants.

In the second part (Part B), the participants were guided through three identical task-based interactions, each time using a different input (touch, voice and gesture). The tasks were short, simple, and closely defined, such as "Lower the blinds and stop them midway" or "Turn on the radio and switch to the next station". This part was used to collect judgments for each input modality and to train the participants in the use of the modalities and the system. The sequence of output and input in Part A and B followed a full Latin square design to counterbalance order effects.

In the Part C, the user was guided by four tasks displayed one at a time on the screen in front of them. This time participants could choose freely which modality they wanted to use and change the modality whenever they felt like it. The first task consisted of all the interactions that had been conducted in Part B, but in this part the subtasks were less precisely defined (e.g., "Find a radio station you like"). The second and third task asked the participants to do something they had not done before, such as programming a movie or adding songs to their playlist. These tasks could not be solved via gestural interaction. As participants were not explicit informed about this, some tried nevertheless. The fourth task was open; users were asked to "play" with the system, again try something they had not done yet or use a modality they had not used often.

In the final part (D), each participant had to perform the Digit-Span test [16] to assess memory capacity as control variable and fill out questionnaires assessing technological affinity [17], ICT experience/attitude in order to assess user features apart from age and gender that are potentially related to modality preferences.

C. Assessments

All participants were asked for their ratings of the three output channels (Part A), the three unimodal input channels (Part B) and the multimodal interface (Part C) on the AttrakDiff questionnaire [18], resulting in seven questionnaires filled in per participant (3,3,1). The AttrakDiff questionnaire contains antonym pairs rated on a 7-point scale ([-3,+3]), yielding the subscales *Attractiveness* (ATT), *Pragmatic Qualities* (PQ), *Hedonic Quality – Stimulation* (HQS) and *Hedonic Quality – Identity* (HQI).

According to [19], overall *Attractiveness* (i. e., valence, beauty) is the result of a simple linear combination of PQ (i.e., simple and functional), HQS and HQI. Of the hedonic qualities, *Identity* describes how well a user identifies with the product. *Stimulation* indicates the extent to which a product supports the needs to develop and move forward by offering novel, interesting and stimulating functions, contents, interactions and styles of presentation.

Table II
SIGNIFICANT RESULTS FOR ASSESSMENTS OF USER CHARACTERISTICS. F-VALUES ($F_{(1,28)}$) AND SIGNIFICANCE LEVEL (ASTERISK).

Data	gender	age	gender:age
Digit-Span value	—	—	—
Technical Expertise (TA)	F=7.29*	F=9.88**	—
Positive Technological Consequences (TA)	—	—	—
Negative Technological Consequences (TA)	—	—	—
Anxiety (ICT)	—	—	F=13.54***
Gadget Loving (ICT)	—	—	—
Training Need (ICT)	—	F=5.62*	—

IV. RESULTS

The reduced recognition rates for some participants did not result in significant rating differences on any ratings scale for any modality condition ($\alpha = .05$) and can thus be neglected in the following.

User variables assessed in Block D are checked for cross-correlations: The following subscales from both questionnaires *technical affinity* (TA) and *ICT attitude/experience* (ICT) have been excluded from analysis, as they seem to assess related constructs due to significant product-moment correlations ($\alpha = .05$, $p > .35$) with other subscales: *Fascination* (TA) (correlates with *Expertise* (TA) and *Gadget Loving* (ICT)); *Exploratory Behavior* (ICT) (correlates with *Anxiety* (ICT)); *Design Oriented* (ICT) (correlates with *Need For Training* (ICT)); *Riskiness* (ICT) (correlates with *Assumed Negative Consequences* (TA)).

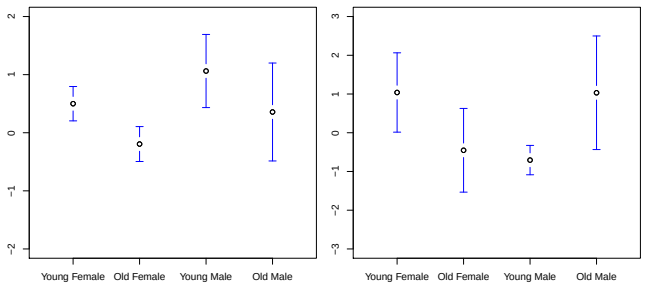
Then, the participants were divided into two groups of age, and gender, respectively. The resulting four groups are tested for differences in the remaining user specific assessments (TA, ICT, Digit-Span test). Age and gender give significant results for some scales assessed (see Table II).

Both age groups do not show any difference in their memorizing abilities. The older adults recruited can be considered as belonging to a possible target group of our multi-modal test system, as they do not exhibit discrepancies in their cognitive abilities and obviously are not physically disabled and thus are not be in need of assistive technology.

Self-reported *Expertise* is lower for both, the older and the female group compared to the younger and male groups (see Figure 3a), which is in line with expectations based on [20][21]. Interestingly, older men and younger women report a higher technological anxiety (Figure 3b) whereas older subjects report of being in need of more professional training with ICT (Figure 4).

These significant differences give information about attitude towards technology in general and may also be used to

 ISBN: 978-1-61208-177-9



(a) Technical Expertise (b) ICT Anxiety

Figure 3. Self-reported Techn. Expertise (a) and ICT Riskiness (b).

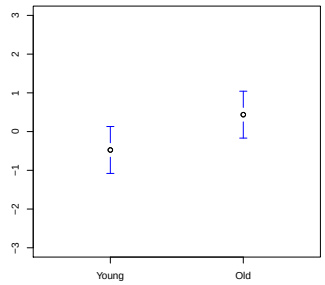


Figure 4. Self-reported Need for Training.

Table III
SIGNIFICANT RESULTS FOR ASSESSMENTS OF THE MULTIMODAL SYSTEM. PEARSON'S R AND SIGNIFICANCE LEVEL (ASTERISK).

Data	ATT	PQ	HQI	HQS
Positive Technological Consequences (TA)	—	—	$r = .43$*	—
Negative Technological Consequences (TA)	—	$r = -.44$*	—	—

explain user group differences not easily explained by age or gender concerning interaction with the system, as well as differences in rating the system, which is analyzed in the following section (IV-A).

A. User group dependent rating of the multimodal system

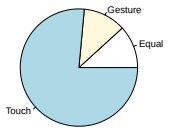
The rating of the whole system was done after the last and most flexible interaction with the multimodal system. The four subscales of the AttrakDiff were used to assess the participants' evaluation of the whole system at that instance. The ratings of the subscales differs neither for age nor gender ($\alpha = .05$).

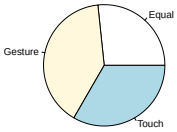
Only when relying on the user group information, there are significant effects. Table III depicts the results of linear correlation analyses with the AttrakDiff subscales and the Part D user assessments as metrical variables:

- The Pragmatic Quality increases with lower expectations of Negative Consequences of technology.
- Hedonic Quality – Identity increases with assumed Positive technological Consequences.

Table IV
SIGNIFICANT RESULTS FOR MODALITY PREFERENCES FOR AGE AND GENDER. F-VALUES ($F_{(1,28)}$) AND SIGNIFICANCE LEVEL (ASTERISK).

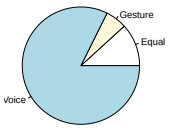
Data	ATT	PQ	HQI	HQS	
Touch-Gesture	4.71*	—	—	—	
Voice-Gesture	12.27**	9.57**	6.84*	4.38*	gender
Touch-Voice	—	—	—	—	
Touch-Gesture	—	—	—	—	
Voice-Gesture	—	—	—	—	age
Touch-Voice	—	—	—	—	

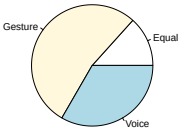




(a) Females (b) Males

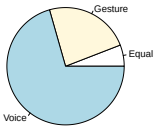
Figure 5. Modality preferences of ATT (Touch or Gesture).

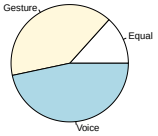




(a) Females (b) Males

Figure 6. Modality preferences of ATT (Voice or Gesture).





(a) Females (b) Males

Figure 7. Modality preferences of PQ (Voice or Gesture).

B. User group dependent modality preferences

But what about modality preferences? Do the ratings of the single modality interactions (Part B) differ for the groups and user variables? This was tested for modality preferences as individual difference between the AttrakDiff subscales of all three modality pairs (Touch-Voice, Touch-Gesture, Voice-Gesture). Table IV summarizes the significant results of the ANOVAs; i.e., that gesture is preferred differently for gender, not for age. We decided to visualize the significant results categorically in Figures 5–9. It can be seen that the overall preference of using touch or voice over gestures concerning *Attractiveness* is dominant for female participants. A similar pattern is not as strong for the other three subscales. In contrast, male participants are divided concerning *Attractiveness*. Additionally, they prefer gestures over voice concerning HQS.

Can we get more insight into these results by analyzing

(a) Females (b) Males

Figure 8. Modality preferences of HQI (Voice or Gesture).

(a) Females (b) Males

Figure 9. Modality preferences of HQS (Voice or Gesture).

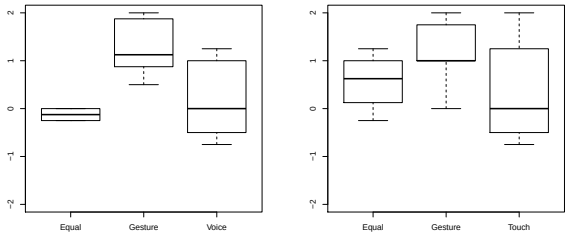
Table V
SIGNIFICANT CORRELATIONS FOR MODALITY PREFERENCES WITH USER ASSESSMENTS. PEARSON'S R AND SIGNIFICANCE LEVEL (ASTERISK).

Data	Touch-Voice Technical Expertise	Voice-Gesture Technical Expertise
ATT	$r = .43$*	$r = -.46$**
PQ	$r = .51$**	$r = -.56$***
HQI	—	$r = -.38$*
HQS	—	—

modality preferences with user variables assessed (see Table V)? The significant negative correlations between Technical Expertise (TA) and the preference of voice and touch over gesture (right column) are similar to the result of female participants preferring voice over gestures in general. Additionally, there is a significant correlation between Technical Expertise and preference of touch over voice (left column) not given for age nor gender. For touch and gesture there is no significant result. Also, neither the females' preference of voice over gesture on HQS and HQI with the opposite for males (Figures 8-9), nor the females' preference of touch over gesture on ATT (5) can be replicated with any of the user characteristics assessed by questionnaires. Thus, effects for Technical Expertise do not help to explain or further describe modality preference effects of gender. For example, the subscale Technical Expertise cannot significantly explain the inconclusive preferences between voice or touch and gesture for male participants (5, 6), although there is a visible tendency to prefer gesture with higher self-reported Expertise (see Figure 10).

V. SUMMARY AND DISCUSSION

When analyzing rating results of participants interacting with a smart-home system, the multimodal system was not judged differently for groups of age or gender. However, using questionnaire-based user characteristics, pragmatic quality increases with participants' decreasing assumed Negative



(a) Voice or Gesture (b) Touch or Gesture

Figure 10. Boxplots of Technical Expertise and preferred modality (men).

Consequences of technology, and Hedonic Quality *Identity* correlates positively with assumed Positive Consequences of technology. Both user characteristics are not affected by age or gender, but give additional information on why individuals considered the interactive system as usable or to identify with it. The latter aspect may be considered even more relevant from a business point of view.

Regarding modality preferences based on rating differences of the unimodal interactions, there are significant results for gender and self-reported Technical Expertise: Whereas in general touch and voice are preferred over gestures, this result is valid only for female participants. Male subjects do not show clear preferences for touch or gesture (on the subscale ATT), and voice or gesture (ATT, PQ, HQS). On the two hedonic subscales, males even mostly prefer gesture over voice in opposition to females. Roughly comparable results are obtained with the self-reported Expertise information instead of gender.

However, as Technical Expertise is interrelated with age and gender, a final conclusion, which factor is causing the effect described – gender or Technical Expertise –, cannot be drawn. As the tendency of males' preference being dependent on Technical Expertise is not significant, a relevant effect size is not expected even with more subjects.

Not being directly related with the user group dependent ratings of the system and single modalities, results form the user assessments give rise to the question, why young female and old male participants seem to avoid ICT (Anxiety, Figure 3b) significantly more than old female and young male subjects, although this result is not in concordance with, e.g., Technical Expertise (Figure 3a).

VI. CONCLUSION AND FUTURE WORK

Even quite strong differences in age of adults do not result in different rating behavior of the multimodal interactive system or the preferences of single input modalities. Instead and surprisingly, gender seems to be a strong factor affecting modality preferences of unimodal interaction. These findings are opposed to results found in [6], which found limited influence of gender, especially considering age, however for task efficiency. In [3], there was no age effect for ratings after interaction, but a positive effect for females concerning

 ISBN: 978-1-61208-177-9

functional and usability aspects of the multimodal system used. Although limited in number of participants and limited to one experimental system (like [3]), the conclusion is to take gender into account much more for interactive systems than is done currently (e.g., [1][7][22]), especially when deciding on the investment into voice and/or gesture control. In this light, results of [22] that 55% of their subjects preferred controlling the home entertainment system via a GUI, but that users also stated that speech input would be their first choice if the speech recognizer had a lower error rate, would be interesting to reanalyze taking into account gender as well. Still, the nature of the interrelation between age, gender and technological expertise is still to be identified with, e.g., special recruited participants. From a pragmatic point of view, grouping users according to gender is much easier than assessing technical expertise.

Furthermore, using other assessment methods will be necessary for answering the questions raised here. For example, addressing the impact of degrees of cognitive abilities was not possible with our recruited participants, although beneficial to the purpose of this paper. Also, we observed single older adults having trouble using the touch screen efficiently. This did not affect the results, but for additional experiments with older adults, assessing dexterity, e.g., with the Grooved Pegboard seems to be advisable.

REFERENCES

[1] M. Turunen, A. Melto, J. Hella, T. Heimonen, J. Hakulinen, E. Mäkinen, T. Laivo, and H. Soronen, "User expectations and user experience with different modalities in a mobile phone controlled home entertainment system," in *MobileHCI*, 2009.

[2] M. Turunen and T. Hakulinen, J. Heimonen, "Assessment of spoken and multimodal applications: Lessons learned from laboratory and field studies," in *Interspeech*, 2010, pp. 1333–1336.

[3] T. Jokinen, K. und Hurtig, "User expectations and real experience on a multimodal interactive," in *Interspeech*, 2006, pp. 1049–1052.

[4] A. Naumann, I. Wechsung, and S. Möller, "Factors influencing modality choice in multimodal applications," in *Perception in Multimodal Dialogue Systems*, 2008, pp. 37–43.

[5] D. Petrelli, A. De Angeli, W. Gerbino, and G. Cassano, "Referring in multimodal systems: the importance of user expertise and system features," in *ACL Workshop on Referring Phenomena*, 1997, pp. 14–19.

[6] A. Naumann, I. Wechsung, and J. Hurtienne, "Multimodal interaction: A suitable strategy for including older users?" *Interacting with Computers*, vol. 22, no. 6, pp. 465–474, 2010.

[7] K. Jeong, R. Proctor, and G. Salvendy, "A survey of smart home interface preferences for U.S. and Korean users," in *Human Factors and Ergonomics Society Annual Meeting*, vol. 53, 2009, pp. 541–545.

[8] T. Kleinberger, M. Becker, E. Ras, A. Holzinger, and P. Müller, "Ambient intelligence in assisted living: Enable elderly people to handle future interfaces," in *Universal Access in HCI, HCI International*, 2007, pp. 103–112.

[9] M. Perry, A. Dowdall, L. Lines, and K. Hone, "Multimodal and ubiquitous computing systems: Supporting independent-living older users information technology in biomedicine," *IEEE Transactions on Information Technology in Biomedicine*, vol. 8, pp. 258–270, 2004.

[10] M. Zajicek and W. Morrissey, "Multimodality and interactional differences in older adults," *Universal Access in the Information Society*, vol. 2, pp. 125–133, 2003.

[11] W. Rogers and A. Fisk, "Toward a psychological science of advanced technology design for older adult," *Journals of Gerontology*, vol. 65B, no. 6, pp. 645–653, 2010.

[12] A. Murata and H. Iwase, "Usability of touch-panel interfaces for older adults," *Journal of the Human Factors and Ergonomics Society*, vol. 47, no. 4, pp. 767–776, 2005.

[13] J. Nielsen, *Usability Engineering*. San Francisco: Morgan Kaufmann, 1993.

[14] C. Kühnel, B. Weiss, and S. Möller, "Evaluating multimodal systems – A comparison of established questionnaires and interaction parameters," in *NordiCHI*, 2010, pp. 286–293.

[15] N. Tractinsky, A. Cokhavi, M. Kirschenbaum, and T. Sharfi, "Evaluating the consistency of immediate aesthetic perceptions of web pages," *International Journal on Human-Computer Studies*, vol. 64, pp. 1071–1083, 2006.

[16] D. Wechsler, *Manual for Wechsler Memory Scaled – Revised*. New York: Psychological Corporation, 1981.

[17] K. Karrer, C. Glaser, C. Clemens, and C. Bruder, "Technikaffinität erfassen – der Fragebogen TA-EG," in *Der Mensch im Mittelpunkt technischer Systeme. 8. Berliner Werkstatt Mensch-Maschine-Systeme*, ser. ZMMS Spektrum, vol. 22, no. 29. Düsseldorf: VDI Verlag, 2009, pp. 196–201.

[18] M. Hassenzahl and A. Monk, "The inference of perceived usability from beauty," *Human-Computer Interaction*, vol. 25, no. 3, pp. 235–260, 2010.

[19] M. Hassenzahl, "The interplay of beauty, goodness, and usability in interactive products," *Human-Computer Interaction*, vol. 19, pp. 319–349, 2008.

[20] J. Cooper and M. Kugler, "The digital divide: The role of gender in human-computer interaction," in *The human-computer interaction handbook: Fundamentals, evolving technologies and emerging applications*, 2nd ed. New York: Lawrence Erlbaum, 2008, pp. 763–775.

[21] S. Czaja and C. Lee, "Information technology and older adults," in *The human-computer interaction handbook: Fundamentals, evolving technologies and emerging applications*, 2nd ed. New York: Lawrence Erlbaum, 2008, pp. 777–792.

[22] M. Johnston, L. D Haro, M. Levine, and B. Renger, "A multimodal interface for access to content in the home," in *Annual Meeting of the Association for Computational Linguistics*, 2007, pp. 376–383.

 ISBN: 978-1-61208-177-9