INTERSPEECH 2019
September 15–19, 2019, Graz, Austria

The Dependability of Voice on Elders' Acceptance of Humanoid Agents

Anna Esposito[1], *Terry Amorese*[1], *Marialucia Cuciniello*[1], *Maria Teresa Riviello*[1], *Antonietta M. Esposito*[2], *Alda Troncone*[1], *and Gennaro Cordasco*[1]

[1]Università degli Studi della Campania "Luigi Vanvitelli", Department of Psychology and International Institute for Advanced Scientific Studies (IIASS), Italy.
[2]Istituto Nazionale di Geofisica e Vulcanologia, Sez. di Napoli Osservatorio Vesuviano, Italy.
anna.esposito@unicampania.it

Abstract

The research on ambient assistive technology is concerned with features humanoid agents should show in order to gain user acceptance. However, differently aged groups may have different requirements. This paper is particularly focused on agent's voice preferences among elders, young adults, and adolescents.

To this aim 316 users organized in groups of 45/46 subjects of which 3 groups of elders (65+ years old), 2 of young adults (aged between 22-35 years), and 2 of adolescents (aged between 14-16 years) were recruited and administered the Virtual Agent Acceptance Questionnaire (VAAQ), after watching video-clips of mute and speaking agents, in order to test their preferences in terms of willingness to interact, pragmatic and hedonic qualities, and attractiveness, of proposed speaking and mute agents. In addition, the elders were also tested on listening only the agent's. The results suggest that voice is primary for getting elder's acceptance of virtual humanoid agents in contrast to young adults and adolescents which accept equally well either mute or speaking agents.

Index Terms: Humanoid agents, Voice role, Users'acceptability, Differently aged users.

1. Introduction

Advances in research and technology have brought about the possibility of humanoid virtual agents becoming part of everyday living, founding applications in several areas of human activity. They are being used as trainer [1, 2]; automated coach [3], museum guide [4], touristic assistant [5], companion or assistant for seniors [6, 7], customer assistant in e-commerce [8], in videogame [9], and in many different ICT (Information Communication Technologies) applications devoted to social services.

This ICT wide range of current and potential applications involves different categories of end users. Thus, the developed agents are required to be easy to use, and able to arouse feelings of trustworthiness and pleasure in order to match the average citizen needs and expectations and favor their use, especially by vulnerable people, as in the case of seniors. They must be attractive and engaging for users of different gender, education, and social status. Indeed, the very critical challenge in developing intelligent virtual agents is on how to elicit acceptance, positive emotions and establish long-term relationships between users and agents.

Most of the researches on this line investigate features, such as agents' human-like appearance, which may affect agent-user's interaction experience, focusing on specific target-groups. The rationale is that users' acceptance and efficient use of such technological supports rely on specific utilitarian (such as accessibility and usability) and hedonic (such as the pleasure associated with its use, and the ability to perform social behavior, e.g. being caring, empathic, intelligent, with human-like communication capabilities) factors, but also characteristics (such as age, gender, general interest in technology, attitude, and personal traits) that are peculiar to each category of users [10, 11].

Few studies however, have been conducted on virtual agent-user's interaction's experience comparing concurrently different categories of users, such as young and elderly persons. Among these, Yaghoubzadeh [12] explored how elderly people and young adults with cognitive impairments assess acceptability and feasibility of virtual agents assisting them in managing their daily life schedule. The authors found that elders were more reluctant than younger to use such systems. Beer [13] showed age-related differences among elders and young adults in recognizing facial emotional expressions of virtual agents, revealing a difficulty for elders in attributing correctly to the facial expressions of such agents emotions of anger, disgust, fear, happiness, sadness, and neutral. Recently, Hosseinpanah [14] investigated the role of age in the perception of emotional non-verbal behaviors of a virtual assistant and observed that compared to young people seniors rely more on nonverbal behaviors in order to perceive the assistant as empathic.

Following this research approach, the present study aims at investigating the role of agents' voice when virtual agents interact with users of different age and what change for elders when instead of a humanoid characters, only a voice user interface is exploited. To our knowledge, researches assessing the influence of agent's voice (on differently aged groups of users), and the weight of voice alone (on elders) are not yet available in literature, being current investigations rather focused on visual agent's features. In particular, this study explores preferences of different groups of adolescents, young adults and elders to interact with humanoid male and female speaking and mute virtual agents.

Our study intends to assess: (i) whether the agents' ability to use voice will encourage users to initiate and maintain an interaction with them; (ii) users' age related differences in the preference of speaking or mute agents; (iii) the effect of voice in determining users' preference of agent's gender.

The final purpose is to disclose differences and similarities in the attitude of differently aged users towards crucial features and attributes humanoid agents should be endorsed in order for them to gain user's acceptance and trust. This will help to improve the definition of such technologies for transversal applications, as well as identify agents' features that are considered critical for the agent to be accepted by specific target groups of users.

http://dx.doi.org/10.21437/Interspeech.2019-1734

2. Material and Method

In order to accomplish the above-mentioned goals, experiments were conducted involving seven differently aged groups of participants (three elders, two young adults, and two adolescents. On the collected data two investigations were conducted. The first was devoted to assess the degree of acceptance of virtual agents and the influence of agents' voice on participants' satisfaction. The latter was devoted to analyse the scores obtained from the elders in order to deepen the role of agents' voice on their preferences.

2.1. Stimuli

Four virtual agents selected from the website BOTLIBRE (www.botlibre.com) by three experts were defined and exploited in the experiments. For a detailed description of the selection procedure please refer to [7].

The agents, two males (Michele and Edoardo) and two females (Giulia and Clara) were showed half torso, with definite clothes, as presented in figure 1.

Figure 1: *The four selected agents.*

Each agent was provided with a different synthetic voice, created through the website Natural Reader (`www.naturalreaders.com`) and producing the Italian sentence "Ciao sono Michele / Edoardo / Clara / Giulia, se vuoi posso aiutarti nelle tue attività quotidiane" (Hi, my name is Michele / Edoardo / Clara / Giulia. If you want, I would like to assist in your daily activities). The voices (recorded using the free software Audacity) were set into each agent's video-clips, which had an average duration of about 6 seconds.

The video-clips were randomly presented to a group of elders, young adults and adolescents with speaking agents, a group of elders, young adults, and adolescents with mute agents, and to an additional group of elders it was required to listen only the agent's voice.

2.2. Participants

The experiments concerned 316 participants split in 7 groups.

Group 1 (22 males and 24 females, mean age=71.59, SD=$\pm$ 6.32) and Group 2 (20 males and 25 females, mean age=71.22, SD=$\pm$ 6.66) consisted of seniors who were asked to assess speaking and mute agents respectively.

Group 3 (21 males and 24 females, mean age=26.07, SD=$\pm$ 3.09) and Group 4 (24 males and 21 females, mean age=25.38, SD=$\pm$ 3.75) consisted of young adults assessing speaking and mute agents respectively.

Group 5 (22 males and 23 females, mean age=14.44, SD=$\pm$ .50) and Group 6 (20 males and 25 females, mean age=14.69, SD=$\pm$ 0.60) were composed of adolescents who judged speaking and mute agents respectively.

Group 7 (21 males and 24 females, mean age=72.73, SD=$\pm$6.20) consisted of seniors who only listen to agents' voice. Results of the first experiment involves groups from 1 to 6 while the second experiment considers only elders (i.e., groups 1, 2 and 7).

Participants were recruited in Campania, in the south of Italy and accepted to participate at the experiment on a voluntary basis. All the participants declared a good degree of experience with technology. They signed an informed consent based on the privacy and data protection procedures established by the current Italian and European laws. The ethical committee of the Department of Psychology at the Università degli Studi della Campania, "Luigi Vanvitelli", authorized this research with the protocol number 25/2017.

2.3. Tools and Procedures

An ad hoc questionnaire VAAQ (Virtual Agent Acceptance Questionnaire) described in [7] was developed to explore participants' satisfaction in interacting with virtual agents. The questionnaire aims to provide hints on agent's perception and technology acceptance degrees among users relatively to diverse agents, with special regard to agent's gender and voice.

The first section of the questionnaire collects information about participants' socio-demographic status and their experience with technology (their ability to use smartphone, tablet and the laptop). The second section, composed by 1 item, is devoted to assess participants' willingness to be involved in interaction with the proposed agents. The third section investigates participants' perception of agent's features. This part of the questionnaire, inspired by Hassenzahl's theoretical model [15, 16, 17] consists in four sub-sections, each composed by 10 items, and assess which features, among those listed below, the proposed interactive agents should possess in order to be highly accepted by their users:

1. Pragmatic Qualities (PQ): how useful, effective, practical, clear and controllable the agents are.
2. Hedonic Qualities- Identity (HQI): how original, creative, captivating as well as presentable, professional and of good taste, the agents appear.
3. Hedonic Qualities- Feeling (HQF): how innovative, exciting and engaging the agents are perceived.
4. Attractiveness (ATT): how attractive the agents are considered, encouraging increased use and positive emotions.

For each VAAQ item, participants' answers were given on a 5-point Likert scale from 1=strongly agree, 2=agree, 3=I don't know, 4=disagree, to 5=strongly disagree.

Since sections 2 and 3 of the questionnaire contain positive and negative items evaluated on a 5-point Likert scale, scores from negative items were corrected in a reverse way, thus low scores summon to positive evaluations, whereas high scores to negative ones. Participants from each group were first asked to provide answers to items of section 1, then they were asked to watch and/or listen to each agents and immediately after completing the items from sections 2 and 3 of VAAQ.

3. Results

In this section, results of the proposed experiments are reported. Several repeated measures ANOVA were carried out [18]. Participants' gender and their belonging to one of the three experimental groups (elders, young adults, and adolescents) were considered as between factors. The involved groups for the first analysis were elders (group 1 and 2), young adults (group 3

and 4), and adolescents (group 5 and 6). Each age category assessed speaking and mute agents respectively. For the second investigation, comparisons were made among elders belonging to groups 1, 2, and 7, the latter involving elders listening only the agent's voice.

The scores obtained at each VAAQ section (Willingness to interact, Pragmatic Qualities, Hedonic Qualities, Identity and Feeling, and Attractiveness) by each agent were considered as within factors in both the analyses. The significance was set at $\alpha < .05$ and differences among means were assessed through Bonferroni's post hoc tests. It is worth recalling that due to the reverse correction of negative items, low scores summon to positive agents' assessments whereas high scores to negative ones.

3.1. Results - First Analysis

This section reports on the first analysis involving scores assigned by elders, young adults, and adolescents to speaking and mute agents.

Willingness to interact

Significant differences (F (5,259) =5.651, p<<.01) emerged among groups (1, . . ., 6). Bonferroni post hoc tests showed that these differences were due to the scores seniors attributed to mute agents (mean = 5.75) which were worse than those they attributed to speaking ones (mean=4.13, p<<.01) and worse than those adolescents attributed to speaking (mean =4.38, p<<.01) and mute (mean=4.22, p <<.01) agents and young adults attributed to mute agents (mean=4.30, p<<.01). These results suggest that seniors are particularly keen to interact with speaking agents more than adolescents and young adults.

Significant differences emerged between male (mean=4.73) and female (mean = 4.44) agents in terms of willingness to interact (F (1,259) =10.832, p<<.01) with a preference for female agents. A significant interaction was found between groups and agent's gender (F (5,259) =11.202, p<<.01). Bonferroni post hoc tests revealed that elders favored female (mean = 3.27) rather than male speaking agents (mean = 4.99) in terms of willingness to interact (p<<.01). This was not the case for mute male and female agents for elders and for both mute and speaking male and female agents young adults and adolescents.

Pragmatic qualities (PQ)

No significant differences (F(5,259) =1.740, p=.126) emerged among groups (1, . . ., 6). Significant differences (F (1,259) =45.330, p<<.01) emerged between male (mean=58.11) and female (mean = 52.93) agents in terms of pragmatic qualities, with a preference for female agents. A significant interaction was found between groups and agent's gender (F (5,259) =26.359, p<<.01). Bonferroni post hoc tests revealed that elders favored female (mean = 41.24) rather than male speaking agents (mean = 65.77, p<<.01). This was not the case for young adults and adolescents, both for speaking and not speaking agents, and for elders interacting with mute agents.

Hedonic qualities- identity (HQI)

Significant differences (F (5,259) =2.645 p=.02) emerged among groups (1,. . ., 6). However, since Bonferroni post hoc tests perform adjustments as a function of the number of comparisons, no significant differences emerged among the six groups. A significant difference (F (1,259) =51.860, <<.01) emerged between male (mean=56.91) and female (mean = 51) agents in terms of hedonic (identity) qualities. A significant interaction was found between groups and agent's gender (F (5,259) =19.775, p<<.01). Bonferroni post hoc tests revealed that elders favored female (mean = 44.50) rather than male speaking agents (mean = 68.06, p<<.01). This was not the case for adolescents, both for speaking and not speaking agents, and for elders interacting with mute agents. Instead, there was a slightly significant preference of young adults toward female not speaking agents (mean =51.613) rather than male not speaking agents (mean=55.970, p=.03).

Hedonic qualities- feeling (HQF)

Significant differences (F (5,259) =2.613, p=.02) emerged among groups (1,. . ., 6). Bonferroni post hoc tests showed slightly significant differences between the scores seniors (mean = 55.18) and adolescents (mean =48.73, p=.04) attributed to mute agents. Significant differences (F (1,259) =33.852, p<<.01) emerged between male (mean=55.80) and female (mean = 51.10) agents in terms of hedonic (feeling) qualities. A significant interaction was found between groups and agent's gender (F (5,259) =18.738, p<<.01). Bonferroni post hoc tests revealed that elders favored female (mean = 43.23) rather than male speaking agents (mean = 65.22, p<<.01). This was not the case for young adults and adolescents, both for speaking and not speaking agents, and for elders interacting with mute agents.

Attractiveness (ATT)

Significant differences (F (5,259) =3.047, p=.01) emerged among groups (1,. . ., 6). Bonferroni post hoc tests showed that these differences were due to scores seniors attributed to mute agents (mean = 55.99) which were worse than those attributed to them by adolescents (mean =48.99, p=.01). A significant difference (F (1,259) =34.215, p<<.01) emerged between male (mean=55.74) and female (mean = 51.26) agents in terms of attractiveness. A significant interaction was found between groups and agent's gender (F (5,259) =27.666, p<<.01). Bonferroni post hoc tests revealed that elders favored female (mean = 41.88) rather than male speaking agents (mean = 66.31, p<<.01). This was not the case for young adults and adolescents, both for speaking and not speaking agents, and for elders interacting with mute agents.

In summary these results show strong preferences of seniors toward female agents only when they were also speaking. To this extent, voice seems to play a strong role in influencing seniors choice of the agent gender. Young adults and adolescents are not affected by agent's voice. For sake of clarity, these results are also depicted in figure 2, where on the top are displayed the scores obtained by the six differently aged groups for the willingness to interact either with a speaking or mute agent. Low scores indicate more willingness to interact than high scores. The graph clearly show that elders are significantly more in favor to interact with speaking rather than mute agents. The bottom of figure 2 displays the scores attributed by the six differently aged groups to the pragmatic, hedonic (identification and feelings), and attractive features of the speaking and mute agents. In both the figure the significant differences are highlighted with *.

3.2. Results - Second Experiment

This section reports on the analysis experiment performed on the scores obtained by the three groups of elders who evaluated speaking and mute agents, as well as, only the agent's voice.

Willingness to interact

Significant differences (F (2,130) =9.442, p<<.01) emerged among groups (1, 2 and 7). Bonferroni post hoc tests showed that seniors preferred to interact more both with speaking agents (mean =4.13, p<.001) and agents' voice only (mean =4.50, p<.01) rather than with mute agents (mean =5.7).

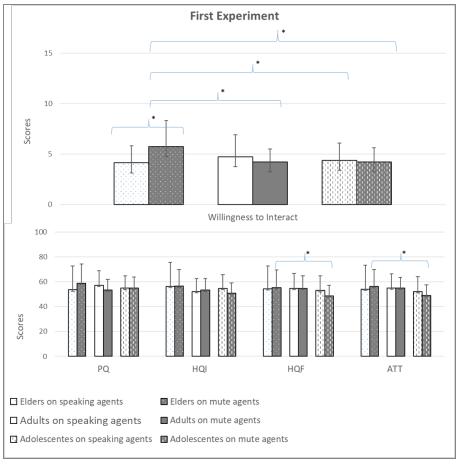

Figure 2: *On the top, the scores obtained by the speaking and mute agents in terms of willingness to interact. On the bottom, the scores obtained on the four considered acceptance dimensions.*

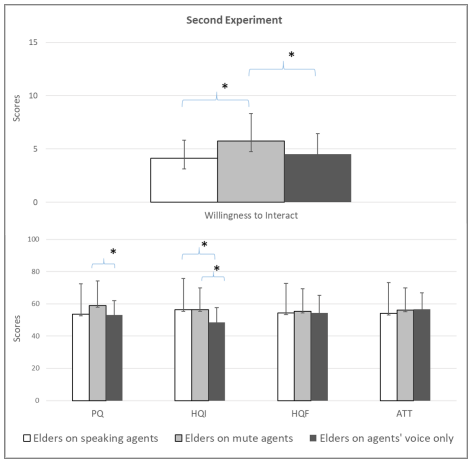

Figure 3: *The scores obtained by the elders to the speaking, and mute agents and to the only voice agents.*

Pragmatic qualities (PQ)
Significant differences (F (2, 130) =3.952, p=.02) emerged among groups (1, 2 and 7). Bonferroni post hoc tests revealed that seniors preferred to interact more only with agent's voice (mean=52.96, p<.05) rather than mute agents, suggesting that from a pragmatic point of view, voice was preferred to the mute visual semblance of the agent.

Hedonic qualities- identity (HQI)
Significant differences (F (2, 130) =9.682, p<<.01) emerged among groups (1, 2 and 7). Bonferroni post hoc tests showed that these differences were due to scores seniors attributed to agents' voices (mean = 48.61) which were better than those attributed to speaking agents (mean =56.28, p<.01) and to mute agents (mean =56.37, p<.01). Therefore, for the hedonic identity, seniors preferred voice to both speaking and mute agent's visual semblances.

Hedonic qualities- feeling (HQF) and Attractiveness
No significant differences emerged among groups (1, 2 and 7) concerning the assessment of agents' Hedonic (feeling) qualities (F (2,130) =.134, p=.874) and Attractiveness (F (2,130) =.773, p=.464).

The results of the second analysis are depicted on figure 3. The graph on the top displays the scores attributed by elders to speaking and mute agents, and only to the agent's voice in terms of willingness to interact, clearly showing legitimate elders' preferences toward voice only or speaking agents. The graph on the bottom of figure 3 clearly illustrates elders' preferences, in terms of pragmatic and hedonic identity dimensions, toward only voice or speaking agents.

4. Conclusion

The present study reports on experiments investigating agents' features affecting differently aged user's satisfaction to interact with diverse shapes of ambient assistive technologies.

In particular, it explored the role played by the agents' voice, in engaging different groups of adolescents, young adults and elders to interact with humanoid male and female speaking and mute virtual agents. In addition, elders were also administered only vocal stimuli.

The data obtained are extremely interesting for guiding the development of interactive technologies serving as assistants or supporting elders in their daily life. This is because they reveal the dependability of these technologies on voice, particularly for the current aged population. In fact, it clearly appears that elders prefer to interact with speaking agents or even with agent's voices only rather than with mute agents. In addition, voice dominates elder's positive assessment of agent's hedonic and pragmatic qualities, and their gender preference, giving to speech a fundamental role for the acceptance of ambient assistive living agents. In addition, elders show a clear gender preference toward female rather than male speaking agents. This preference disappears when both male and female agents are mute.

Interestingly this is not the case for young adults and adolescents. For them, it does not matter whether the agent they are interacting with, is speaking or mute, female or male.

It comes natural to ask ourselves why is that. What bring elders to be so affected by voice in their preference for a virtual assistant. One reason that can be proposed is the experience with technologies. Elders are surely less familiar with mute technologies and therefore the speaking ones appear to them more appealing and easy to use. In addition, voice does not require cognitive labors, acclimatization, memory resources, attention, is more flexible with respect to traditional graphical menus and visual interfaces, and is a long term exploited modality (for elders more than young people) for daily interactional exchanges. Finally, we would like to suggest that, in assistive technologies, elders are looking for more human-like features than young people, and may not appreciate the agent lacking of them in relation to the seriousness of the task (being assistant) they have to perform, and in the context of assistance, speech is surely one of the ultimate human characteristic.

What remains to be investigated, is the degree to which voice only may affect the acceptance of interactive technologies by adolescents and young adults.

5. Acknowledgements

The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 769872 (EMPATHIC) and No 823907 (MENHIR).

6. References

[1] M. Macedonia, I. Groher and F. Roithmayr, F. "Intelligent virtual agents as language trainers facilitate multilingualism," *Frontiers in psychology*, vol 5, 295, doi:10.3389/fpsyg.2014.00295, 2014.

[2] S. Provoost, H.M. Lau, J. Ruwaard and H. Riper, "Embodied Conversational Agents in Clinical Psychology: A Scoping Review," *J. Med. Internet Res.*, vol 19(5):e151 doi:10.2196/jmir.6553, 2017.

[3] A. van Wissen, C. Vinkers and A. van Halteren, "Developing a Virtual Coach for Chronic Patients: A User study on the Impact of Similarity, Familiarity and Realism". In Proceedings of *Persuasive Technology - 11th International Conference* (PERSUASIVE 2016), vol. 9638, pp. 263-275, Springer/Verlag. `https://doi.org/10.1007/978-3-319-31510-2_23`, 2016.

[4] S. Kopp, L. Gesellensetter, N.C. Krämer and I. Wachsmuth, "A Conversational Agent as Museum Guide – Design and Evaluation of a Real-World Application". In Proceedings of *Intelligent Virtual Agent* (IVA 2005). Lecture Notes in Computer Science, vol 3661, Springer, Berlin, Heidelberg, 2005.

[5] A.I. Niculescu, H.Y. Kheng, L.F. D'Haro, S. Kim, R. Jiang and R. E. Banchs, "Design and evaluation of a conversational agent for the touristic domain". In Proceedings of *Signal and Information Processing Association Annual Summit and Conference* (APSIPA'14), doi:10.1109/APSIPA.2014.7041744, 2014.

[6] N.A. Shaked, "Avatars and virtual agents - relationship interfaces for the elderly". In Healthcare technology letters, vol. 4(3), pp. 83-87, doi:10.1049/htl.2017.0009, 2017.

[7] A. Esposito, T. Amorese, M. Cuciniello, A.M. Esposito, A. Troncone, M.I. Torres, S. Schlgl and G. Cordasco, "Seniors' Acceptance of Virtual Humanoid Agents". In Proceedings of *Forum Italiano Ambient Assisted Living* (ForItAAL 2018), Lecture Notes in Electrical Engineering, vol. 544. Springer, Cham, 2019.

[8] T. Sandbank, M. Shmueli-Scheuer, D. Konopnicki, J. Herzig, J. Richards and D. Piorkowski, "Detecting Egregious Conversations between Customers and Virtual Agents". CoRR. http://arxiv.org/abs/1711.05780, 2017.

[9] J. Fraser, I. Papaioannou and O. Lemon, "Spoken conversational AI in video games – Emotional dialogue management increases user engagement," In Proceedings of *the 18th International Conference on Intelligent Virtual Agents* (IVA 2018), pp. 179-184, ACM, DOI: 10.1145/3267851.3267896, 2018.

[10] J. Li, M.X. Zhou, H. Yang and G. Mark, "Confiding in and Listening to Virtual Agents: The Effect of Personality", In Proceedings of *the 22nd International Conference on Intelligent User Interfaces* (IUI 2017), pp. 275-286, 2017.

[11] D.M. Cereghetti, S. Kleanthous, C. Christophorou, C. Tsiourti, C. Wings and E. Christodoulou, "Virtual partners for seniors: analysis of the users' preferences and expectations on personality and appearance". In *European conference on ambient intelligence* (AmI'15), Athens, Greece, 2015.

[12] R. Yaghoubzadeh, M. Kramer, K. Pitsch and S. Kopp, "Virtual Agents as Daily Assistants for Elderly or Cognitively Impaired People - Studies on Acceptance and Interaction Feasibility," In Proceedings of *International Conference on Intelligent Virtual Agents* (IVA 2013), 2013.

[13] J. M. Beer, C. A Smarr. A. D., Fisk and W.A. Rogers, "Younger and Older Users' Recognition of Virtual Agent Facial Expressions". *International journal of human-computer studies*, vol. 75, pp. 1-20, 2015.

[14] A. Hosseinpanah, N.C. Krämer, C. Straßmann, "Empathy for Everyone?: The Effect of Age When Evaluating a Virtual Agent," In Proceedings of *the 6th International Conference on Human-Agent Interaction* (HAI '18), Southampton, United Kingdom, 2018.

[15] M. Hassenzahl, "AttrakDiff(tm)". *Internet Resource* http://www.attrakdiff.de, 2019.

[16] M. Hassenzahl, "The interplay of beauty, goodness, and usability in interactive products". *Human-Computer Interaction*, vol. 19, pp. 319-349, 2004.

[17] M. Hassenzahl, "The thing and I: Understanding the relationship between user and product," *Funology. Human-Computer Interaction Series*, vol. 3, pp. 31–42, Springer, Dordrecht, 2003.

[18] C. S. Davis, "Statistical Methods for the Analysis of Repeated Measurements," *Springer-Verlag New York*, pp. XXIV, 416, ISSN:1431-875X, DOI:10.1007/b97287, 2002.