USER ACCEPTANCE OF EMOTION- AWARE MOOD-IMPROVING VOICE ASSISTANTS

R. Bernhaupt, C. Murko, G. Pottier, A. Battut

ruwido, Austria

ABSTRACT

Speech as an interaction mechanism for television control is perceived as fast and easy. While speech and speech to text are standard mechanisms and commonly used, the usage of voice and voice-based information like emotional state recognition of the user is underexplored. To understand the potential of voice aware and mood improving services a mixed method approach including a web-based study with 130 participants from the US and a user experience study in Austria and France (n=20) was conducted. For acceptance, results indicate that technology interest is a key value to predict acceptance of such systems, with acceptance rates of up to 80% for such systems. Main areas for innovation are services for people that feel stressed, users that would like to have a fun evening, or situations where people have difficulties in falling asleep.

INTRODUCTION

Speech interaction is becoming increasingly available and popular in smart living rooms due to the rise of digital voice assistants such as Alexa or Google Assistant [9]. A speech request carries verbal information, the command and based on the users' individual voice, the tone and frequency allows us to gather different emotional states. [11]. There are different fields where emotion-recognition can be applied, see already existing applications for example in call centers for the training of customer support [4] or emotion-aware infotainment systems in cars [10].

Today, smart speakers make use of what the user is saying, but not (yet) how the user is talking to the machine (the users voice print). This ability could be implemented soon in any standard smart speaker, thanks to the availability of data, efficient machine learning techniques and the recent progress in emotion recognition technology [e. g. 6].

The main goal of this research was to investigate user acceptance of emotionally aware systems and services that enhance mood. Particularly in terms of a reported gulf between users' expectations and experiences regarding those "artificial roommates" [9], insights into users' expectations as well as users' experiences with existing products are necessary contributions for the development of products. As part of such contributions, the environment (or context), personality (neuroticism) and gender have been highlighted for evaluations of user experience regarding emotion-aware systems [3]. Additionally the general dimensions of user experience for interactive television [2] are important, particularly the dimension

emotion [2] might benefit from three factors that reflect rewarding feelings concerning user's entertainment experience: 1) fun, 2) thrill, and 3) empathic sadness [1]. All of those factors were taken into account for the current report and study design.

Central questions for this report were what to do with the recognized emotion not only in the context of watching TV, but beyond taking into account (future) smart environments in the home? What will be the potential benefits for the user if the virtual assistant knows the emotional context of a request? Do users value emotionally aware services in a different way when they are in different mood, e.g. sad or happy [7]?

In the short term, the virtual assistant may provide suggestions for additional actions to realize (e.g. movie or music recommendations). In the longer term, the virtual assistant could be able to assess the user's state of mind at a given time, her mood in the long run, and even change available functionalities in the smart home to make for example the user's life more convenient or more relaxing (e.g. mood-related adaption of lights and music playlists).

The remainder of this article presents the research goal and the selected methods, with descriptions of the study set-up and prototypical system used, results and a discussion.

RESEARCH GOAL AND METHOD

To understand what is needed to improve user acceptance of voice assistants that can recognize user's emotions and provide emotion-based services to enhance user's mood, three research questions are central:

- Do users accept smart speakers in the living room in general and are emotion-enabled smart speakers acceptable specifically?
- What are the potential benefits for the user if the virtual assistant knows the emotional context of a voice request and are there differences in experience between users who were influenced to be in one of two mood conditions, namely sad or happy mood?
- What requirements can be identified for the technical implementation of emotionally aware voice assistants, if they have to be widely accepted by users?

A mixed-method approach combining a web survey and a lab study was performed. The web survey running in April 2019 focused on the ownership and usage of smart speakers in the United States of America. The recruitment of study participants was conducted via SurveyMonkey Audience. To complement these findings a lab study was performed in May 2019 with a focus on user experience to understand the impact of emotion-aware voice interaction service integrated in an IPTV user interface.

STUDY PART I:WEB-BASED SURVEY

The general findings about user's trust in current smart speakers such as Alexa devices, Google Home, etc. are based on data from a random sample of 130 participants in US households. The web-based survey took on average 10 minutes to finish. It was presented in English and was addressed to smart speaker users (n = 74; 57%) and non-smart speaker users (n = 56; 43%). The survey included questions in terms of demographic data (age, gender), household size, interest in technology, privacy concerns regarding voice assistants and usage of emotionally aware smart speakers.

The distribution of households was the following: 40.8% of the sample live with family (40.8%), in marriage (33.1%) or alone (16.2%). 4.6% live in partnership, 0.7% in flat sharing and 4.6% in other relationships. The age of people participating in the survey ranged from 18 to 87 years with a mean age of 50 years old (M = 49.77; SD = 16.24). 2 participants preferred not to answer about the age. Overall the study included 54 male (41.5%), 75 female (57.7%) participants and one participant self-described as gay (0.8%).

48 people of smart speaker users are owners of Alexa devices (64.9%), 23 people use Google Assistant (31.1%), 21 people use Siri (28.4%), and 5 people mentioned other options (6.8%) such as Jibo, Bible or the Xfinity remote control. Smart speaker users are on average 50 years old (M = 50.3; SD = 17.44) with 61% female and 39% male users. The non-users are on average 49 years old (M = 49.02; SD = 14.67) with 55% female and 45% male users.

RESULTS

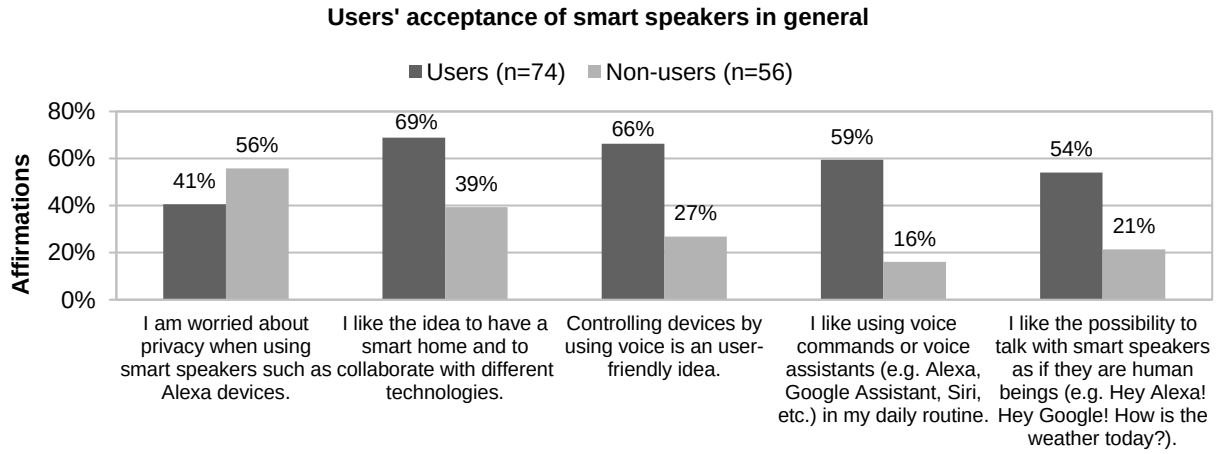

Figure 1. Users' acceptance of smart speakers in general split for Users of Smart Speakers and Non-Users of Smart Speakers

For getting insights into the general acceptance of smart speakers users were asked to rate their perception of privacy concerns on a scale from 1 (do not agree at all) to 5 (highly agree). Figure 1 shows that while only 41% of the participants using smart speakers agreed or highly agreed to be worried about privacy, in the group of participants not using smart speakers 56% agreed or highly agreed.

Users of smart speakers are, as expected, more interested in concepts like smart home, voice interaction and voice assistants and show in general a higher acceptance. Almost 70% of smart speaker users and 40% of non-users currently like the idea of having a smart home and to collaborate with different technologies. 59% of actual users like using voice commands in their daily routine. Only 16% of non-users want to use voice assistants or voice commands daily.

Acceptance of mood-improving smart speakers (based on expectations)

Findings from 120 web survey participants showed tendencies of gender-related differences regarding the anticipated usage of emotionally aware smart speakers. Whereas only 17%

of women would never use smart speakers that are able to improve their mood, it is one quarter of men (27%) who would never use them. Women would also rather prefer to use it (very) frequently (33%) than men (27%).

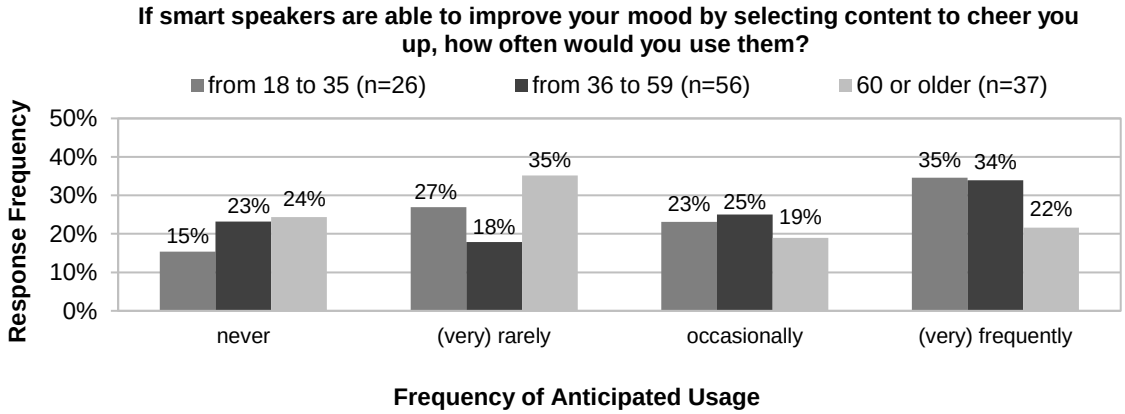

Figure 2. Age-related frequencies of anticipated usage regarding emotionally aware smart speakers

Figure 2 shows the differences between age groups and the anticipated usage of emotionally aware smart speakers. Nearly a quarter of both groups in the age from 36 to 59 (23%) and 60 or older (24%) would never use mood-improving smart speakers. However, 15% of the younger people who are between 18 and 35 would never use that service. Beyond that, the findings show that 35% of the youngest age group as well as 34% of people between 36 and 59 would use emotionally aware smart speakers very frequently or frequently. The group with the highest age likes to use it rather rarely than frequently.

Particularly, the interest in technology takes over a key role in the acceptance of emotion-aware systems, with interest in technology being a significant predictor ($\chi^2_{(5)}=14.767$, $p = 0.011$). 26% of low-technology interested people (n=34) would never use this type of service, whereas 12% would use it very frequently or frequently. 15% of high-technology interested people (n=48) would never use it and 50% would frequently or very frequently use it. Besides that, the findings show 87% of smart speaker users and 67% of non-users would like to use mood-improving voice assistants in the future, at least on a sporadic basis.

Summarized, the acceptance of mood-improving smart speakers, or voice assistants, is rather high. Taking all participants into account, every fifth person (21%) would never use this service. 79% of people would like to use this service in the future.

Potential benefits of mood-improving smart speakers (based on expectations)

Regarding the question when and why users would use this type of mood-improving smart speakers, participants of the web survey replied with rather similar needs. The most frequently stated reasons for using mood-improving smart speakers are stressful situations, where they need relaxation, meditation and motivation. Furthermore, they can imagine frequently using it for listening to music as well as watching movies or "*jokes, stories, uplifting news, reassuring phrases, fun Youtube videos"*. Use case examples of mood-improving services mentioned by users are 1) after bad work days, 2) for demanding situations with household members, 3) for situations where they have sleep problems and 4) if they are

alone and need fun (*"I'm alone a lot and it is a fun integration to my retired life"*). One participant also mentioned it would be useful for *"a time when I was having a strong negative reaction to an event and needed to dissipate the feeling before I did or said something that I would regret"*.

STUDY PART II

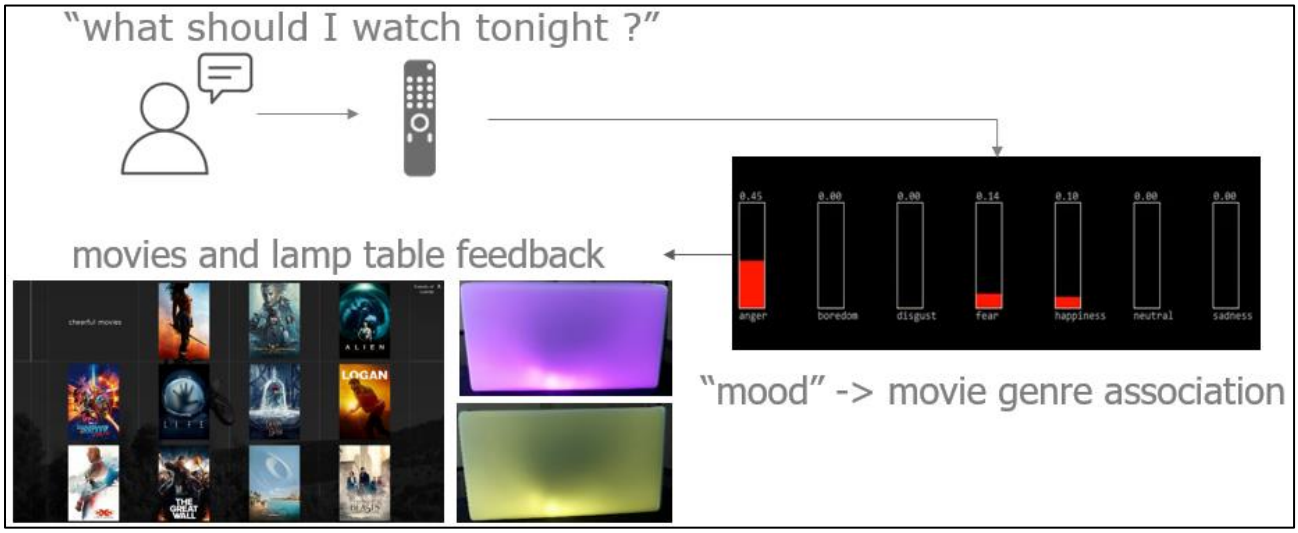

Figure 3: Prototypical System used to investigate voice-based emotion recognition to improve recommendations to enhance mood and display mood via light inside a set-top box

Based on a prototypical system combining voice interaction with an emotion recognition tool, a recommendation system was enhanced to suggest mood improving content. At the same time the user interface indicated the detected mood and content and displayed the mood in terms of color via a set of leds in the set top box.

The sample size (n=20) of the lab study consists of 10 female (7 AT/3 FR) and 10 male (6 AT/ 4 FR) participants and has been balanced regarding the mood condition (10/10). The user experience study sessions lasted about 30 min.

Although it was a Wizard of Oz experiment setting, users trusted the story that the mood detector is able to recognize the current mood state via voice interaction for providing customized movie recommendations on the TV screen. After an introduction at the beginning, the Austrian participants (AT) were asked to read a newspaper report about misleading actions of climate protection and starvation, whereas the French participants (FR) were asked to read a report about child work in palm oil exploitations (sad mood condition). All participants in the happy mood condition were reading a report about the psychological power of friendships. The participants were randomly assigned to each mood condition. To find out the current mood state after the mood induction procedure, the participants rated how they were feeling (9-points Likert-scale; from very sad (1) to very happy (9)). Thereupon, the user task has been a voice request ("What can I watch today?") executed by participants using a remote control that offers the possibility for voice interaction. Followed by that, the participants get displayed their induced mood by means of 1) mood-related light changes of a table lamp (the color switched from the white and neutral state to Yellow for happy mood or Purple for sad mood and back to the standard white color

again after 4 seconds), and 2) mood-related movie recommendations (amusing (fun) and exciting movies (thrill) for people in the happy mood condition; touching (empathic sadness) and amusing (fun) movies for people in the sad mood condition) based on findings from [1, 5]. After displaying the mood state, users were asked to rate the difficulty of the task, the fitting of the light color, the fitting of the movie recommendations, the experienced emotions [4] and the user experience of the mood detector in terms of hedonic quality, pragmatic quality, goodness and beauty [7]. Each session ended by means of an interview about usage of such emotionally aware systems in the future and their perceived effectiveness. All participants were rewarded for their time.

No country-, age- and gender-related differences in user experience were found. Therefore, the following results refer on the difference between pretended sad (M = 3.6; SD = 1.2) and happy (M = 7.3; SD = 1.7) users only, due to the significant difference in mood ratings after mood induction (U = 4, p = .00).

RESULTS

User experience of the mood detector

User Experience Ratings

	Beauty	Goodness	Pragmatic Quality	Hedonic Quality	Ease of Task	Light: Fits to Mood	Movies: Interesting	Movies: Exciting	Movies: Entertaining	Movies: Fit to Mood
happy	89%	90%	84%	90%	96%	90%	92%	82%	92%	86%
sad	84%	81%	78%	82%	96%	78%	80%	78%	74%	64%

Figure 3. User ratings depending on the mood condition

For most of the user experience ratings we didn't find significant differences, although users in the sad mood condition tended to rate, on average, less satisfied than users in the happy mood condition (except for ease of task). Figure 4 shows an overview on the rather high satisfaction ratings in both happy and sad groups. Significant differences between the user groups were found in regard to movies' level of interest, entertainment and fitting to the mood. Also the light's color fitting by mood shows an existing but not significant difference between the user groups. At first sight those findings indicated some users in the sad mood condition had the feeling they had received the wrong feedback. This assumption was confirmed in the interviews, where users were frequently highlighting the importance of a properly functioning feedback of mood and well-customized content, especially for services that are supposed to improve users' mood.

Users were also asked what they felt during the usage of the mood detector. Happy users mentioned the most they were interested (10 out of 10), cheerful (9 out of 10) and fascinated (6 out of 10) and sad users mentioned they were interested (9 out of 10), fascinated (6 out of 10) and cheerful (4 out of 10). No significant differences were found between the user groups and the intensity of those emotions. On average, users rated the perceived intensity of emotions with 91% of interest, 77% of fascination and 75% of cheerfulness during the usage.

Requirements for emotionally aware voice assistants (based on user testings)

18 of 20 users think it is possible to improve a user's mood by means of an emotionally aware voice assistant. While 4 of 20 users (20%) would never use this product at home, 1 user (5%) would use it rarely, 7 users (35%) would use it occasionally and 8 users (40%) would use it (very) frequently. Some users mentioned it would be "*fun*" and "*cool*" showing this type of service to friends or other people (e.g. at parties or dating situations). Most users preferred the idea of getting a mood-based combination of movie as well as music results. Some of them also liked the idea of getting lights to change (light intensity and hue) or different fragrances or the possibility to close the outside blinds as a result they were feeling sad. Most users had similar ideas of use cases to those we had already learnt from the web survey (e.g. it is helpful for 1) retired people, 2) people who are not able to relax or to improve their mood on their own as it could be with mental illnesses, 3) boring situations at home where entertainment is needed).

Taking all users statements together, it is essential for users to have a properly functioning system that is providing accurate results and *"fulfilling their hopes"*. They would need a trustworthy and reliable system *("I need a good interpretation of my current mood state that feels right"*). Due to privacy concerns of some users *"the product must handle the data offline and transparent, if it is possible in any way"* in order to have a good feeling about the system. Besides that, it should be possible to have control over the system without it being enabled on its own. *("I would use it, but only in situations where I have the feeling that I need it. I want to use it, when I am 'ready' to use it").* Otherwise, if the emotionally aware system is not providing an accurate and trustworthy service, "*it could affect my mood negatively"*. Users prefer also to have different options to choose if they are getting service offerings based on their mood (e.g. one user in sad mood condition brought up: *"I only watch amusing movies, I never watch dramas or touching movies as it was recommended on the first page"*). Therefore, most users mentioned they would prefer mood-improving results as a *"nice add-on"* for their content search listed in combination with results of their individual interests. This would be *"really fun and it avoids hours of searching".*

SUMMARY AND CONCLUSIONS

Due to our findings regarding acceptance, it is worth considering factors such as users' age, gender and the interest in technology to develop emotion-aware voice assistants in the context of smart homes. As we have not found any significant differences but tendencies in usage frequencies (excluding the factor of technology interest), our findings would need evidence from studies with larger sample sizes. While the youngest generations and female users tend to be more open for this type of services, older generations and male users might accept mood-improving services as well with the difference being a tendency to use less often. However, almost 80% of users would generally like and accept mood-improving services by usage of voice interaction, because they see benefits for situations, where they need relaxation, meditation and motivation or just fun and entertainment. This evidence promises potential for future innovations in voice search and smart homes, if users' requirements are satisfied well enough.

REFERENCES

1. Bartsch, A., 2012. Emotional gratification in entertainment experience: Why viewers of movies and television series find it rewarding to experience emotions. Media Psychology, 15(3), 267-302.
2. Bernhaupt, R., and Pirker, M., 2013. Evaluating user experience for interactive television: towards the development of a domain-specific user experience questionnaire. In IFIP Conference on Human-Computer Interaction (pp. 642-659). Springer, Berlin, Heidelberg.
3. Broek, E. L., 2013. Ubiquitous emotion-aware computing. Personal and Ubiquitous Computing, 17(1), 53-67.
4. Burkhardt, F., van Ballegooy, M., Englert, R. and Huber, R., 2005. An emotion-aware voice portal. In Electronic Speech Signal Processing Conference, Prague, Czech Republic.
5. Cavalcante, E., Rivero, L. and Conte, T., 2015. MAX: A Method for Evaluating the Post-use User eXperience through Cards and a Board. 27th International Conference on Software Engineering and Knowledge Engineering (SEKE 2015)
6. Fayek, H. M., Lech, M. and Cavedon, L., 2017. Evaluating deep learning architectures for Speech Emotion Recognition. Neural Networks, 92, 60-68.
7. Greenwood, D., 2010. Of sad men and dark comedies: Mood and gender effects on entertainment media preferences. Mass communication and Society, 13(3), 232-249.
8. Hassenzahl, M. and Monk, A., 2010. The inference of perceived usability from beauty. Human–Computer Interaction, 25(3), 235-260.
9. Luger, E. and Sellen, A., 2016. Like having a really bad PA: the gulf between user expectation and experience of conversational agents. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (pp. 5286-5297). ACM.
10. Schuller, B., Rigoll, G., Grimm, M., Kroschel, K., Moosmayr, T. and Ruske, G., 2007. Effects of in-car noise-conditions on the recognition of emotion within speech. In Proceedings of the DAGA 2007, Stuttgart, Germany.
11. Vogt, T., André, E. and Bee, N., 2008. EmoVoice - A framework for online recognition of emotions from voice. In International Tutorial and Research Workshop on Perception and Interactive Technologies for Speech-Based Systems (pp. 188-199). Springer, Berlin, Heidelberg.