Jan 17th, 12:00 AM

The impact of emotional voice assistants on consumers' shopping attitude and behavior

Carolin Kaiser
Institute for Market Decisions, Nuremberg, Germany, carolin.kaiser@nim.org

René Schallner
Institute for Market Decisions, Nuremberg, Germany, rene.schallner}@nim.org

The impact of emotional voice assistants on consumers' shopping attitude and behavior

Carolin Kaiser[1], René Schallner

[1] Institute for Market Decisions, Nuremberg, Germany
{carolin.kaiser,rene.schallner}@nim.org

Abstract. While most recent voice assistants talk in a non-emotional tone, this is expected to change. Since emotions are often contagious in human-human-interactions, the question arises whether consumers also catch the expressed emotion of voice assistants while shopping. In this paper, we present an approach for synthesizing speech in different emotional tones and show that synthesized emotional speech impacts consumers' emotions, shopping attitude and behavior.

Keywords: emotional voice assistants, human-computer interaction, consumer behavior, voice shopping

1 Introduction

Most recent voice assistants talk in a non-emotional tone. However, with technology becoming more humanoid, this is about to change [1]. In human-human-interaction, people often catch the emotion of others. With the increasing use of voice assistants, the question arises whether people also catch the emotion of voice assistants. This is especially important for marketing, as emotions play a crucial role in purchase behavior [2]. Therefore, we develop a speech synthesizer which generates speech in three emotional tones and explore how the emotional tone of a voice assistant impacts consumers' shopping behavior and attitude. We contrast an excited and happy tone which favorably impact consumer shopping behavior when interacting with human shop assistants [2] to an uninvolved tone as baseline.

2 Related Work

Voice assistants change the way people are shopping online. Instead of reading, consumers are talking with voice assistants. Speech is a powerful carrier of information, signaling also social clues such as personality or emotion [1]. Adding a human voice to text enhances peoples' user experience [3], shopping joy and buying intention in online shopping [4].

Synthesized speech must not be expressionless but can manifest personality [1]. In shopping, female synthesized voices can feminize products, while male synthesized voices can masculinize products [1]. Compared to recorded speech, it is perceived less positive and less convincing [1].

17th International Conference on Wirtschaftsinformatik,
February 2022, Nürnberg, Germany

While synthesized speech often lacks emotions, they are essential to human speech. People often catch the emotion of others while speaking. Studies show that positive emotions of service employees enhance the mood of customers [5] and impact their buying behavior [2]. Emotions are also important in human-computer-interaction [1]. However, it is unclear, to which extent different emotional tones of voice assistants impact consumers' attitude and shopping behavior.

3 Emotional Speech Synthesis

Text-to-speech systems, which convert text into speech, have improved rapidly. New deep learning text-to-speech systems produce natural and intelligible speech. To model emotional German speech, we train an enhanced version of the state-of-the-art Tacotron system [6] in the three different emotional tones: excited, happy, and uninvolved.

Deep learning systems require huge training data sets. To efficiently synthesize speech in different emotional tones, we train a base model on a large general German speech data set (29h audio) [7] and fine-tune it to different emotions using 2h audio spoken by a German voice actress.

To assess the quality of the synthesized emotional speech, we conducted listening tests by three different groups of people. Two experts recognized the emotion with an accuracy of 98%, 121 university students with 90% and 224 crowd workers without German skills with 71%.

4 Shopping Experiment

To explore the impact of a voice assistant's tone on consumers' shopping behavior and attitude, we conducted a laboratory shopping experiment with 210 German students hearing either no voice, a recorded voice, or a synthesized voice with an excited, happy, or uninvolved tone.

Participants had to buy 3 offices supplies with a budget of 10 Euros in an experimental audio-visual online-shop (see Figure 1). During check-out they get a special offer. At the end, participants receive their bought products and remaining budget in cash. While shopping, participants' emotional arousal was measured by a heart-rate sensor. After shopping, participants answered a questionnaire about the perception of the voice, their emotion and shopping attitude (likert-scale 1-7).

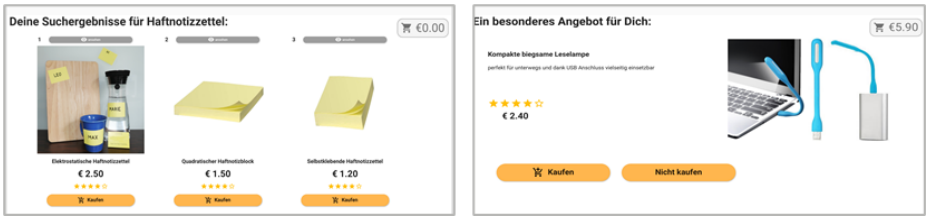

Figure 1. Experimental online-shop

The results reveal that voice shopping increases the impulsiveness and speed of shopping compared to shopping without voice (see Table 1). The synthesized voice was perceived less natural and likable as the recorded voice but did not impact participants' shopping attitude and behavior negatively.

Table 1. Impact of voice

Condition	Impulsive buying score*	Overall time spent shopping in sec. **
No Voice	4.87	142.68
Voice	6.18	72.41
Overall	6.12	82.70

Voice	Liking of voice*	Naturalness of voice****
Synthesized	2.94	2.56
Recorded	3.56	4.17
Overall	3.25	3.35

* p<.1, ** p<.05, *** p<.01, **** p<.001

The emotional tone influenced participants' emotion, attitude, and behavior (see Table 2). Participants hearing the happy voices felt most positive, evaluated the user experience best, and were most satisfied. Participants listening to the excited voice became most aroused and seemed to buy more impulsively.

Table 2. Impact of emotional tones

Emotion	Consumer valence while shopping **	User experience ****	Customer Satisfaction ****	Percentual change of arousal at special offer *	Impulsive buying score
Excited	5.74	5.72	4.67	1.43	6.25
Happy	5.95	5.79	4.72	0.21	6.09
Uninvolved	5.44	5.14	3.80	0.18	6.23
Overall	5.71	5.55	4.40	0.58	6.18

* p<.1, ** p<.05, *** p<.01, **** p<.001

To substantiate these findings, we repeated the experiment online with 1000 students hearing the synthesized voice in an excited, happy, or uninvolved tone. The results were confirmed. The happy voice significantly enhanced the shopping attitude and the excited voice significantly increased the impulsiveness to buy the special offer.

5 Conclusions

This paper presents an approach for synthesizing speech in different emotional tones and explores its impact on consumers' shopping attitude and shopping behavior. The shopping experiment yields valuable insights about German consumers. Further research with participants from different cultures is ongoing to gain further knowledge.

References

1. C. Nass, S. Brave, Wired for Speech. Cambridge: MIT Press, 2005.
2. N. Bessouh, D. O. Belkhir, "The Effect of Mood on Impulse Buying Behavior - Case of Algerian Buyers," The International Journal of Business & Management 6(1), 2018
3. L. Qiu, I. Benbasat, "An investigation into the effects of Text-To-Speech voice and 3D-avatars on the perception of presence and flow of live help in electronic commerce," ACM Trans. Comput. Hum. Interact 12(4), 329–355, 2005.
4. S.-A. Jin, "The Roles of Modality Richness and Involvement in Shopping Behavior in 3D-Virtual-Stores," Journal of Interactive Marketing 23(3), 234–246, 2009.
5. S. D. Pugh, "Service with a Smile: Emotional Contagion in the Service Encounter," The Academy of Management Journal, 44(5), 1018–1027, 2001.
6. D. Stanton, Y. Wang, R. Skerry-Ryan, "Predicting Expressive Speaking Style from Text in End-To-End Speech Synthesis," 2018 IEEE Spoken Language Technology Workshop, 595–602.
7. I. Solak, "The M-AILABS Speech Dataset," Jan. 2019. https://www.caito.de/2019/01/the-m-ailabs-speech-dataset/.