Morphing Vulnerable Machines: Paralinguistic Cues in Digital Voice Assistants Shape Perceptions of Physicality, Vulnerability, And Trust

FOTIOS EFTHYMIOU
University of St Gallen

Cite as:
EFTHYMIOU FOTIOS (2020), Morphing Vulnerable Machines: Paralinguistic Cues in Digital Voice Assistants Shape Perceptions of Physicality, Vulnerability, And Trust. *Proceedings of the European Marketing Academy*, 49th, (64040)

Paper from the 49th Annual EMAC Conference, Budapest, May 26-29, 2020.

Morphing Vulnerable Machines: Paralinguistic Cues in Digital Voice Assistants Shape Perceptions of Physicality, Vulnerability, And Trust

ABSTRACT

Paralinguistic features in the human voice are reliable cues to detect discrete momentary emotions and personality inferences in other humans. By listening to different paralinguistic features in the human voice people infer differences in personality (Mohammadi, Origlia, Filippone, & Vinciarelli, 2012). The current work provides evidence from a series of tightly controlled experiments that altering digital voice assistants along the vibrato dimension (i.e., systematic changes in the pitch of a synthesized voice) causes systematic changes in personality perception and trust, while holding critical dimensions of message content, syntax, and other paralinguistic cues constant. Our results demonstrate that humans attribute greater submissiveness, lower dominance, and reduced perceptions of power to a digital voice assistant with increasing vibrato. We further show that these perceptions are explained by altering the perception of physicality such that greater vibrato results in perceiving the voice assistant as older and smaller. Moreover, we show that these changes in perception cause subsequent attributions of trust and are robust across a broad range of consumer demographics and psychological trait measures.

Keywords

Digital Voice Assistants, Voice Morphing, Vibrato.

Track

Digital Marketing & Social Media

Morphing Vulnerable Machines: Paralinguistic Cues in Digital Voice Assistants Shape Perceptions of Physicality, Vulnerability, And Trust

EXTENDED ABSTRACT

INTRODUCTION

Paralinguistic features in the human voice, such as pitch, speech rate, or loudness, play an important role in nonverbal human-to-human communication (Knapp, Hall, & Horgan, 2014). Specifically, paralinguistic features signal nuances in meaning, convey a person's attitude, or express discrete emotions during speech formation. The current study tests whether changing the vocal features of a digital voice assistant can elicit predictable changes in perception. Specifically, the current research isolates the effect of changes in vibrato (i.e., variations in the pitch of a synthesized voice) and how this increase in vibrato affects the perceptions of physicality of a digital voice assistant, how vulnerable they perceive the assistant to be, and how much they trust in recommendations received by a digital voice assistant.

THEORETICAL BACKGROUND

The current research builds on prior research showing that, people use paraglinguistic cues to judge others' personality traits (Mohammadi et al., 2012), physicality (Scherer, 2003) or attractiveness (Klofstad, Anderson, & Peters, 2012). The majority of the past research in this area has almost exclusively focused on human paralinguistic cues, and the most studied vocal features are pitch (high vs low), intensity, fundamental frequency and voice breaks (Scherer, 2003)

In our work, we investigate both a prevalent but understudied paraglinguistic cue (i.e., vibrato) and how it affects people's perceptions in terms of attributions of physicality, personality traits, and trust.

STUDY 1

In the first study, a total of N = 379 participants were randomly assigned to four conditions differing in the extent of vibrato of the voice assistant. Across all studies, we used the AWS Amazon Polly interface for speech formation and the PRAAT vocal toolkit as an established tool to morph the assistants voice. The first group acted as a baseline whereas the second,

third, and fourth increased systematically in the extent of tremolo such that all voice output was scaled up by one semitone each (from semitone=2 to semitone=4). Participants were asked to listen to seven random-generated sentences with a duration of 2-4 seconds and were subsequently requested to answer questions about the perceptions they formed while listening to the voice. The results in the first study showed that quiver voices were perceived as significantly more powerless ($F(3,375) = 9.663$, $p < 0.01$) more vulnerable (Chi square = 80.1285, $p < 0.01$, df = 3), submissive ($F(3,375) = 17.95$, $p < 0.01$) and greater in neuroticism (Chi square = 68.5104, $p < 0.01$, df = 3) compared to baseline.

STUDY 2

The key objective of Study 2 was to test the underlying process to explain perceptions of vulnerability, lower perceptions of power, and how these perceptions affect downstream consequences in terms of trust on recommendations. Extending Study 1, all vocal output was scaled up by half semitone dependent on condition (from semitone=2 to semitone=4). A total of 466 participants were asked to listen to an advertisement and then were requested to answer questions about the perceptions they formed about the vocal features of the speaker, as well as questions on the physicality of the speaker (in terms of height, weight, and age). The results revealed that greater vibrato led to perceptions of smallness ($F(5,460) = 1.769$, $p < 0.05$) and being older (β=-0.2462, $p< 0.05$), and these perceptions of physicality significantly mediating the effect on vulnerability, perceptions of power, and submissiveness. Furthermore, the results also revealed a non-monotonic effect of vibrato such that at the extreme of the vibrato distribution, people had lower trust and this effect was significantly mediated by perceiving the digital voice assistant as more vulnerable and possessing a lack of power.

GENERAL DISCUSSION

The current findings make three novel contributions to prior work on the use of digital voice assistant in marketing and information systems, the role of paralinguistic cues in mind attribution in human-machine interaction, and the effective design of a machine's "personality". To the best of our knowledge, this is the first line of systematic research demonstrating how variations in vibrato induce perceptions of vulnerability, power, and submissiveness, how these changes are caused by altering the envisioned physicality of a digital voice assistant, and how such changes affect important downstream consequences from such as attributions of trust. We are currently extending the current findings and test to

which extent these changes in perceptions could evoke an increase on donation spendings, or under which conditions more vs. less as powerful perceived voice assistants affect people's willingness to engage on tasks that require greater dominance or power (such as following driving directions or receiving medical advice), and the role of individual differences in people's attachment style. Similarly, building on literature proposing that feelings of power in human-human interaction can induce greater goal-orientation (Guinote, 2007), we further investigate how people would react in challenging vs. non-challenging tasks.

REFERENCES

Corretge, R. (2019). *Praat Vocal Toolkit. Http://www.praatvocaltoolkit.com*.

Guinote, A. (2007). Power and Goal Pursuit. *Personality & Social Psychology Bulletin*, *33*, 1076–1087. https://doi.org/10.1177/0146167207301011

Klofstad, C., Anderson, R., & Peters, S. (2012). Sounds like a winner: Voice pitch influences perception of leadership capacity in both men and women. *Proceedings. Biological Sciences / The Royal Society*, *279*, 2698–2704. https://doi.org/10.1098/rspb.2012.0311

Knapp, M. L., Hall, J. A., & Horgan, T. G. (2014). *Nonverbal communication in human interaction*. Boston, MA: Wadsworth.

Mohammadi, G., Origlia, A., Filippone, M., & Vinciarelli, A. (2012). From speech to personality: Mapping voice quality and intonation into personality differences. *Proceedings of the 20th ACM International Conference on Multimedia - MM '12*, 789. https://doi.org/10.1145/2393347.2396313

Scherer, K. (2003). Scherer, K.R.: Vocal Communication of Emotion: A Review of Research Paradigms. Speech Communication 40, 227-256. *Speech Communication*, *40*, 227–256. https://doi.org/10.1016/S0167-6393(02)00084-5