Haptic Augmentation of Audio and its Effects on Speech Perception

Antoine Weill-Duflos, Pascal E. Fortin, Feras Al Taha, Jeremy R. Cooperstock

Haptic Augmentation of Audio and its Effects on Speech Perception

Antoine Weill-Duflos*[1,2], Pascal E. Fortin†[1,2], Feras Al Taha‡[1], Jeremy R. Cooperstock§[1,2]

[1]Department of Electrical and Computer Engineering, McGill University, Montreal, Canada
[2]Centre for Interdisciplinary Research in Music Media and Technology, Montreal, Canada

Voice characteristics are known to influence people's perception of a speaker's professional abilities, often offering an unfair disadvantage to speakers in position of perceived vulnerability. By using a custom speech to haptics synthesis framework, this paper presents results from a user study investigating the influence of haptic speech enhancement on speaker's characteristic perception. A custom speech to haptic system is used to first replicate a study from Klofstad et al. examining how voice pitch influences perception of speaker strength, competence and trustworthiness, and second to explore the impact of multimodal stimulus presentation on these perceived characteristics. Our preliminary findings suggest that the perceived strength of a speaker with a higher voice pitch is enhanced, whereas the outcome is uncertain for competence. Perceived trustworthiness was not affected by the system at all. This work puts forward the potential positive effect that the use of haptic-enhanced communication system could have in social and professional communications, but also outlines its limitations.

Introduction

Our perception of our peer's professional abilities are based on cues extracted from several information channels. A non-exhaustive list would include proxemics [4], kinesics, vocalics [3] chronemics and haptics [8]. Physical appearance (stature/posture), charisma, smell and speech are all characteristics that affect the impression that a speaker has on their audience. Accordingly, the outcome of work meetings, academic presentations, and videoconferences, can vary for the listeners through their perception of the speakers involved, affecting their judgement of the matter discussed. Focusing on speech characteristics, which also appear in remote meetings, the voice's loudness and resonance will affect how dominant and confident one appears [6, 2, 12, 15], whereas voice pitch and tone was found affect attractiveness [6, 2]. While we acknowledge the contribution of all voice components in forming an impression of a speaker, this work focuses specifically on voice pitch.

Prior literature on the human voice identified that the fundamental frequency (F_0), loudness and resonance of voice are part of the factors that influence the perceived characteristics of speakers [9]. [13] investigates the effects of various characteristic of the voice in the perception of attractiveness of the speaker. On the topic of fundamental frequency, a review of 35 articles suggests that female listeners would prefer lower-pitched male voices and male listener would prefer higher-pitched female voices. [5] showed that speaker would modulate their own speech fundamental frequency in response to their attractiveness to the conversational partner. Another notable contribution from [13] is the highlight on the complexity of this specific research field, behaviors depends on factors including language or social context and published studies have sometimes contradictory conclusions.

[14, 7] highlight the importance of the voice pitch bias as it could influence voting behavior or trust in a speaker [10]. According to [7]:

> men with lower-pitched voices are perceived as "more attractive", physically stronger and "socially dominant". [..] higher-pitched female voices are perceived as more attractive, whereas lower-pitched female voices are perceived to be "socially dominant"

Klofstad *et al*. [7] investigated how voice pitch would influence perception of leadership characteristics, namely "competence", "strength" and "trustworthiness". One of the research questions considered was the influence of gender, both from the listener and the speaker in the perception. The results from their study suggested that female listeners are not affected by pitch differences when judging male voices. On the other hand, male listeners find lower-pitched voices to be more competent and strong. For female voices, both genders found the lower-pitched voices to be more trustworthy, strong and competent. Female listeners perceive lower-pitched male voices to be more attractive while male listeners perceive them as more competent and stronger. Moreover, both male and female listeners perceive lower-pitched female voices to be more competent, stronger and more trustworthy.

Inspired by research on the influence of different vibrotactile stimuli on user arousal [17], we took on the challenge of exploring the use of haptic augmentation to modify the perception of speech without altering the speaker's original voice [16].

[16] presented a system and results from an early user study. The gist of the system is as follows: the higher the pitch of the speaker, the stronger the vibrations were rendered. The objective is to counterbalance the potential lack in dominance and presence associated with higher pitched voice with a greater and hypothesized increased dominance associated with more intense vibrotactile renderings. Results collected using a two alternative forced-choice methodology suggested that the inclusion of haptic reinforcement indeed modified the participants' response patterns. However, due to methodological limitations, it was difficult to determine what properties of the speaker were meaningfully influenced by the use of haptic reinforcement. This paper therefore presents the results from a follow-up study investigating in more details which perceived speakers' characteristics, from perceived competence, confidence and trustworthiness, were impacted by the haptic reinforcement.

Study

This study was adapted from the methodology of [7], whose results will serve as comparison material.

*antoine.weill-duflos@mcgill.ca
†pefortin@cim.mcgill.ca
‡feras.altaha@mail.mcgill.ca
§jer@cim.mcgill.ca

A within-subject experiment design was employed, where each participant was presented with every combination of the two experimental conditions: with or without haptics, and hearing a male or female speaker.

During a trial, participants listened to the low and high-pitched version of a same voice uttering the sentence "I urge you to vote for me this November" through headphones. After listening to each pair, participants were asked to reply to *one* of the following three questions:

1. "Which voice is more competent (e.g. capable, experienced, knowledgeable, effective)?"
2. "Which voice is stronger (e.g. confident, determined, resolute, self-assured)?"
3. "Which voice is more trustworthy (e.g honest, straightforward, reliable, believable)?"

Participants were asked to enter their answer using a regular keyboard's keys labeled *First* and *Second*, corresponding to the first or second voice respectively.

At the end of the experiment, participants completed a short questionnaire consisting of demographic questions.

Each question was asked four times to a participant, using each one of the following condition combination once:

1. Female voice, no haptics
2. Female voice, with haptics
3. Male voice, no haptics
4. Male voice, with haptics

The presentation order of the 12 combinations of questions and pairs of voices was randomized within subject.

Technical Framework

Participants were seated in an office chair augmented with two 100 W haptic actuators (Clark Synthesis TST239 Silver) placed in the lower back and under the seat (see Figure 1). The study took place in a quiet room.

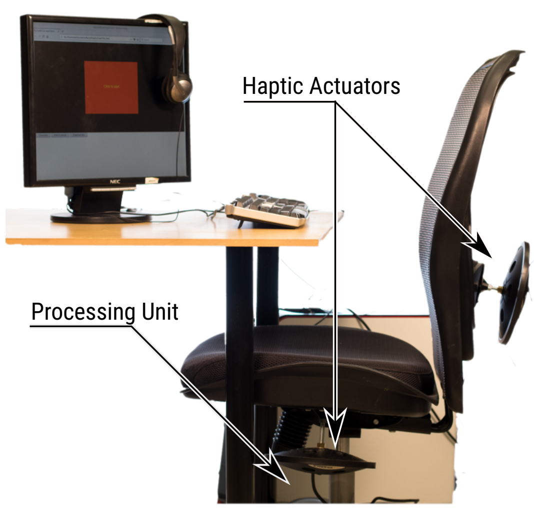

Figure 1: Picture of the real setup

Adapting the methodology presented by Klofstad et al. [7], we kept the phrases and questions they used as well as the acoustic presentation of the samples. In our study, six synthetic voices (3 males, 3 females) were used from Google Wavenet [11] to generate our test speakers. For every voice, lower pitch and higher pitch versions were generated by removing or adding 5 semitones, resulting in a total of 12 synthetic voices. All generated voices were synthesized saying the exact same sentence without changes of emphasis.

PRAAT phonetic analysis program [1] was used to measure the pitch of each recorded voice. The pitch of female and male low-pitched voices ranged from 129.8 Hz to 177.7 Hz ($\bar{x} = 150.9$ Hz) and 106.2 Hz to 144.3 Hz ($\bar{x} = 129.1$ Hz), respectively. The pitch for female and male high-pitched voices ranged from 164.4 Hz to 224.6 Hz ($\bar{x} = 191.3$ Hz) and 131.8 Hz to 179.3 Hz ($\bar{x} = 161$ Hz), respectively. Compared with Klofstad et al.'s study [7], the male voices used are approximately 40 Hz higher in pitch, while the female voices are practically the same pitch.

Using the signal processing pipeline presented in [16] and briefly schematized in Figure 2, the haptic signal was generated by filtering and pitch-shifting the original speech signal such that it rested in haptic-relevant frequency bands of 0 Hz to 500 Hz. The vibrotactile signal was presented synchronously with the speech signal, such that they formed a cohesive multimodal stimulus. In the haptics condition, the voice was augmented by its synchronous haptic version. The signal processing algorithm was designed such that higher pitched voices would be presented with more intense vibrations than their lower pitch opposite, attempting to counterbalance the potential biases caused by differences in voice pitch.

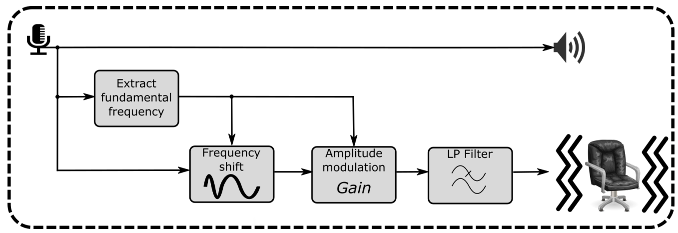

Figure 2: Schematics of the haptic delivery system

Results and Discussion

A total of 14 participants (9F, 5M) took part in the study, aged from 19 to 33 ($\bar{x} = 24$ and SD = 4.2).

Cross analysis with the baseline study

Table 1 reports the differences between listener's and speaker's gender. We first show the results in the condition without the haptic augmentation to allow an unbiased comparison with Klofstad's results.

voice gender	*Female*		*Male*	
listener gender	*F*	*M*	*F*	*M*
competent	75.00	66.67	50.00	20.0
stronger	62.50	60.00	14.29	60.0
trustworthy	55.56	40.00	50.00	50.0

Table 1: Frequency at which participants selected the high-pitched version of a voice without haptic augmentation, segregated by voice and listener gender.

The baseline results from [7] are listed in Table 2 with the difference between the two studies. In contrast with Klofstad's study [7], our participants seemed to be generally more likely to select the high-pitched

voice gender	Female		Male	
listener gender	F	M	F	M
competent	28 (47)	30 (36.67)	45 (5)	42 (-22)
stronger	31 (31.5)	27 (33)	44 (-29.71)	38 (22)
trustworthy	27 (28.56)	32 (8)	50 (0)	44 (6)

Table 2: Percentages of choice of the high-pitched option in the study from [7], the number in parenthesis are the difference between both studies. Frequency at which participants selected the high-pitched version of a voice in [7], segregated by voice and listener gender. In parenthesis is the difference between the current findings and their results.

voices. An average difference of 22.5 % was observed between the two studies, suggesting that a our results fail to replicate their findings with regards to the effect of listener and speaker gender on perceived competence, strength and trustworthiness.

Several factors could explain this dissimilarity; our main hypothesis lies in the distortion created by the speech synthesis system for voices of different pitch. The Klofstad *et al.* study used recorded voices and the pitch was altered ± 0.5 equivalent rectangular bandwidths. This would results in shift of ± 20 Hz. Our study relies on shifting the voice from 5 semitones, equivalent to multiplying/dividing the frequencies by $(\sqrt[12]{2})^5$. Another source of uncertainty is that the pitch was entered as one of the parameter in the speech synthesis as opposed to shifting the pitch after the voice was generated, which could have introduced different effects in the voice. In addition, this difference outlines the potential serious impact of differences in vocal properties for the design of smart voice assistant technologies.

Effects of the Haptic signal

We now turn our attention towards the effects of using the proposed haptic enhancement of the voice on the perception of speakers. A comparison of both with and without haptics conditions are given in Table 3 with the results of both gender of listeners averaged, all results are also visually represented in Figure 3.

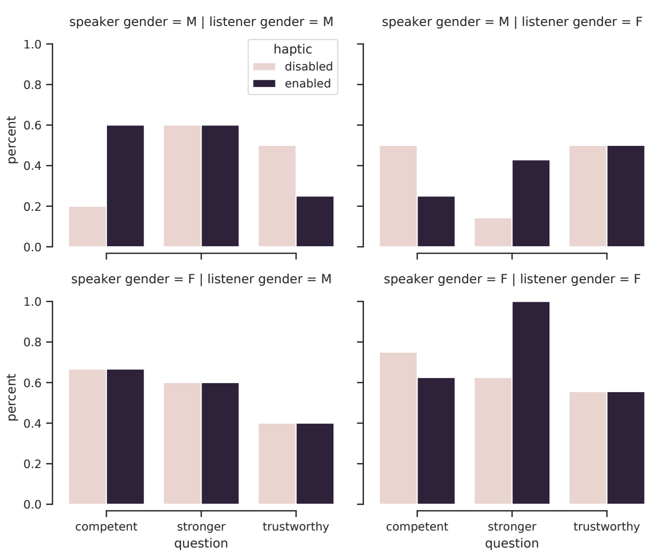

Figure 3: Percentage of choice of high-pitched voice with and without haptic reinforcement depending on speaker and listener gender

voice gender	Female		Male	
haptic	disabled	enabled	disabled	enabled
question				
competent	72.73	63.64	38.46	38.46
stronger	61.54	84.62	33.33	50.00
trustworthy	50.00	50.00	50.00	41.67

Table 3: Frequency at which the high-pitched option was selected for every question. The columns represents the different combinations of speaker gender and haptic effect presentation.

The authors refrain from using inferential statistics due to the relatively small sample size. As such, all reported results should be considered as exploratory.

Our results suggest that the proposed haptic enhancement does influence the perceived competence of the speaker. However, this effect seems to be interacting with the actors' gender. If male listeners were more likely to select the higher-pitched *male* haptic-enhanced voice as being more competent, this effect was not present with *female* voices. On the other hand, the use of haptic enhancement seems to have negatively influenced perceived competence of the high-pitched voice for female listeners regardless of the speaker's gender. The fact that the system had a negative impact with female listeners would reinforce the existing biases which is not against the desired objectives of the proposed system.

In terms of the perceived strength of the speaker, no notable effects were noted for male participants. However, the use of haptic enhancement seems to have positively impacted the female participants' strength perception of both speaker genders. Despite the lack of noticeable difference with male participants, the fact that the system partially countered the bias in female participants is supportive of the proposed speech enhancement approach.

In the case of the perceived trustworthiness of the speaker, the proposed haptic enhancement system failed to counterbalance or attenuate existing biases. Indeed, the only effect visible suggests that male participants found male speakers with a higher-pitched voice less trustworthy with the haptic system than without its assistance.

Conclusion

In this work, results from an exploratory study evaluating the impact of a haptic speech enhancement system on voice-pitch related biases were presented. The study recreated the protocol of previous documented work investigating the impacts of voice pitch on perceived competence, strength and trust.

Our results failed to replicate the results of the study from Klofstad *et al.* which noted statistically significantly inferior competence, strength and trustworthiness in higher-pitched speakers. It is suspected that this difference might have been caused by differences in the voices samples used, which in the current study were synthesized. These observations also outlines the importance of investigating the social and affective impact of vocal properties for synthesized voices, especially as smart voice assistants and becoming increasingly embedded in our daily lives.

The second part of the results on the effects of haptic enhancement were more agreeing with our previous findings [16], as the haptic signal helped reinforce the perceived strength in higher pitched voices. Indeed, these preliminary findings indicate that it could be possible to enhance the leadership perception of speakers with a higher fundamental frequency using a real-time system rendering vibrotactile feedback to audience members, reinforcing the speech. As suggested by the phrase used in our trials, such a system could ultimately be used to equalize the speech perception of speakers in a political context, to force listeners to focus on more important features such as the content/ideas conveyed.

References

[1] P. Boersma and D. Weenink. Praat: Doing phonetics by computer (version 6.0. 37). *Retrieved 14.03. 2018 from http://www.praat.org/*, 2018.

[2] B. Borkowska and B. Pawlowski. Female voice frequency in the context of dominance and attractiveness perception. *Animal Behaviour*, 82(1):55–59, July 2011.

[3] J. K. Burgoon, N. E. Dunbar, and C. Segrin. Nonverbal Influence. In *The Persuasion Handbook: Developments in Theory and Practice*, pages 446–474. SAGE Publications, Inc., 2455 Teller Road, Thousand Oaks California 91320 United States, 2002.

[4] E. T. Hall. The hidden dimension. 1966.

[5] S. M. Hughes, S. D. Farley, and B. C. Rhodes. Vocal and physiological changes in response to the physical attractiveness of conversational partners. *Journal of Nonverbal Behavior*, 34(3):155–167, 2010.

[6] B. C. Jones, D. R. Feinberg, L. M. DeBruine, A. C. Little, and J. Vukovic. A domain-specific opposite-sex bias in human preferences for manipulated voice pitch. *Animal Behaviour*, 79(1):57–62, Jan. 2010.

[7] C. A. Klofstad, R. C. Anderson, and S. Peters. Sounds like a winner: Voice pitch influences perception of leadership capacity in both men and women. *Proceedings of the Royal Society of London B: Biological Sciences*, page rspb20120311, Mar. 2012.

[8] M. L. Knapp and J. A. Daly. *Handbook of Interpersonal Communication*. SAGE, Oct. 2002.

[9] S. J. Ko, M. S. Sadler, and A. D. Galinsky. The Sound of Power: Conveying and Detecting Hierarchical Rank Through Voice. *Psychological Science*, 26(1):3–14, Jan. 2015.

[10] K. J. Montano, C. C. Tigue, S. G. E. Isenstein, P. Barclay, and D. R. Feinberg. Men's voice pitch influences women's trusting behavior. *Evolution and Human Behavior*, 38(3):293–297, May 2017.

[11] A. v. d. Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu. Wavenet: A generative model for raw audio. *arXiv preprint arXiv:1609.03499*, 2016.

[12] D. A. Puts, C. R. Hodges, R. A. Cárdenas, and S. J. C. Gaulin. Men's voices as dominance signals: Vocal fundamental and formant frequencies influence dominance attributions among men. *Evolution and Human Behavior*, 28(5):340–344, Sept. 2007.

[13] A. Suire. *Évolution de la voix humaine : le rôle de la sélection sexuelle*. PhD thesis, Université de Montpellier, Nov. 2019.

[14] C. C. Tigue, D. J. Borak, J. J. M. O'Connor, C. Schandl, and D. R. Feinberg. Voice pitch influences voting behavior. *Evolution and Human Behavior*, 33(3):210–216, May 2012.

[15] K. J. Tusing and J. P. Dillard. The sunds of dominance vocal precursors of perceived dominance during interpersonal influence. *Human Communication Research*, 26(1):148–172, Jan. 2000.

[16] A. Weill-Duflos, F. Al Taha, P. E. Fortin, and J. R. Cooperstock. BarryWhaptics: Towards Countering Social Biases Using Real-Time Haptic Enhancement of Voice. In *2019 IEEE World Haptics Conference (WHC)*, pages 365–370. IEEE, 2019.

[17] Y. Yoo, T. Yoo, J. Kong, and S. Choi. Emotional responses of tactile icons: Effects of amplitude, frequency, duration, and envelope. In *World Haptics Conference*, pages 235–240. IEEE, 2015.