9th International Conference on Speech Prosody 2018
13-16 June 2018, Poznań, Poland

Prosodic entrainment and trust in human-computer interaction

Štefan Beňuš[1,2], Marian Trnka[2], Eduard Kuric[3], Lukáš Marták[3], Agustín Gravano[4], Julia Hirschberg[5], Rivka Levitan[6]

[1]Constantine the Philosopher University in Nitra, Slovakia
[2]Institute of Informatics, Slovak Academy of Sciences, Bratislava, Slovakia
[3]Faculty of Informatics and Inf. Technologies, Slovak Univ. of Technology Bratislava, Slovakia
[4]Depart. de Computación, FCEyN, Universidad de Buenos Aires & CONICET-UBA, Argentina
[5]Department of Computer Science, Columbia University, New York, USA
[6]Department of Computer and Information Science, Brooklyn College CUNY, USA

sbenus@ukf.sk, trnka@savba.sk, {eduard.kuric, lukas.martak}@stuba.sk, gravano@dc.uba.ar, levitan@sci.brooklyn.cuny.edu, julia@cs.columbia.edu

Abstract

Despite the prevalence of findings supporting the positive relationship between entrainment and social aspects, previous research does not provide a clearly converging picture and several recent studies stress the importance of *disentrainment* for smooth spoken interactions. We have developed a novel scenario design to explore how prosodic entrainment relates to the trust of a human user in an avatar's ability to provide good advice. Our results with Slovak subjects suggest that 1) applications might need to boost the entrainment effect to affect trust of humans towards avatars, 2) user's sex plays a role in developing the relationship between trust and entrainment, and 3) females tend to prefer the advice of *disentrainning* avatars.

Index Terms: human-computer interaction, entrainment, trust, Slovak

1. Introduction

Entrainment, also called accommodation or alignment, is the tendency of interlocutors to produce similar forms of behavior. In speech, rich literature supports this tendency in various prosodic aspects such as pitch, intensity, speech rate, or voice quality in multiple studies of entrainment at the local and global levels, or as similarity, convergence, or synchrony [1, 2, 3].

Several theoretical models propose either external or internal motivation for this widespread speech behavior. For example, according to Communication Accommodation Theory [4] people entrain when they (want to) feel closer to and think positively about the interlocutor. The level of verbal entrainment or coordination might also stem from the degree of mutual active engagement between interlocutors [5]. The model in [6] assumes that alignment on linguistic representations in the production-perception loop is an essential internal component of dialogue between humans.

In an effort to improve the naturalness and effectiveness of applications using human-computer (spoken) interaction, several studies have reported promising results when entrainment between the user's and system's voices took place. For example, [7] reported gains in ASR accuracy when entrainment of speech rate was induced, [8] gains in learning when entrainment between humans and a tutoring system occurred in pitch and intensity, [9] improved rapport and naturalness when a system shifted the pitch contour of the synthesized speech by the mean pitch of the user, or [10] the link between entrainment and trust in English.

Despite the prevalent results linking entrainment and positive social aspects, several recent studies point to the relevance of disentrainment and dialogue organization. For example, [11] observed correlations between positively perceived interactions and entrainment only if entrainment was construed as including both positive and negative similarity. [12] showed more prosodic entrainment in collaborative than competitive dialogues only when analyzed within dialogue acts. [13] found that interlocutors who were perceived as likeable tended to disentrain in terms of speech rate. And disentrainment was also linked to trust in Slovak and Spanish human-computer interactions [10]. Hence, both entrainment and disentrainment are present in human interactions and may play diverse, but possibly complementing, functions.

Finally, no consensus regarding the role of biological sex in affecting the degree and functions of (dis)entrainment has been reached. Although some studies have found that women are more likely to entrain than men, stemming possibly from the higher phonetic sensitivity of female speakers [14] or their greater social engagement with the interlocutor or perhaps from lesser perceived dominance, other studies suggest that the pattern is more complex. More entrainment was observed in mixed-sex dyads than dialogues between two females [15] and disentrainment between humans and robots was found irrespective of the user's sex [16].

Given the assumed potential of entrainment for human-computer spoken interactions and the open issues related to the relevance of disentrainment and user's sex, the goal of the current paper is to explore how prosodic (dis)entrainment relates to the trust of humans towards the capability of avatars to provide good advice and whether this relationship is affected by the user's sex and personality. Our contribution to these debates is 1) in describing a novel scenario and procedures for probing the relationship between trust and prosodic entrainment, 2) presenting pilot results suggesting that females tend to prefer the advice of the disentraining avatars, and 3) findings that suggest that this might be linked to the openness to new experience as a personality trait.

2. Methodology

To test the relationship between speech entrainment and trust we developed a novel scenario implemented using the Unity

 10.21437/SpeechProsody.2018-45

Engine. We describe the key elements of the game design, the implementation of the voice interaction between the user and the avatars, and the experimental procedure.

2.1. Game design

Our one-person adventure game is inspired by Harry Potter. The main character is a young wizard whose goal is to obtain a special device to protect the wizards in the school of magic. The wizard is faced with challenges and binary decision points and is assisted by a pair of owl-like helpers (avatars) who provide information and situation-relevant advice. For each binary choice, the two avatars give diverging advice and the prosodic features of the avatar's voices are manipulated to either entrain or disentrain to the user's features (further explained in (e) below and in Section 2.2). The player's preferences regarding whose advice they follow constitute the operationalization of the player's trust toward an avatar. Each decision involves trust since the player is vulnerable to the actions of the avatars [17] as the players always risk that a decision might disadvantage them. In this implementation we concentrate only on trust in the system's skills/ability, and not in its integrity or benevolence.

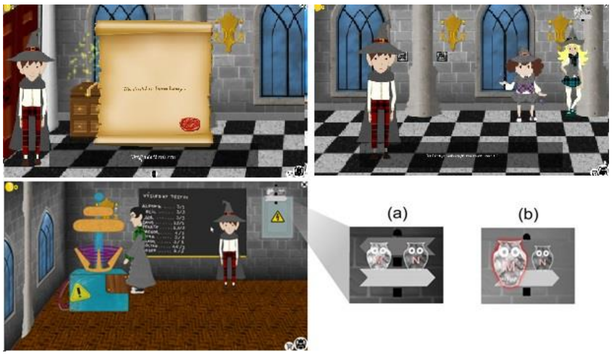

Figure 1: *Examples of game visualizations; the oath in top-left, a scene with other characters in top-right, the final scene in bottom-left, and avatar visualization when quiet (a) and when one avatar speaks (b).*

The key challenges and requirements and our approach to them in the game design and implementation were as follows.

a) *The player should not be biased externally to prefer either avatar nor to confound voice manipulation as the primary factor for entrainment~trust analysis*. The quality of advice is not revealed because the player is not allowed to return to the decision point once a decision has been made to check the other option and no immediate feedback is given to the player regarding their decisions. The two pieces of advice for each decision are controlled for length, the avatar's expressed confidence, and presentation order. The avatars are designed to be visually similar; see Fig. 1 bottom-right.

b) *Each subject should have a comparable game experience to facilitate between-subject analysis*. The progression of the game is fixed. Tasks requiring motor skills or affecting cognitive load are avoided to minimize the effect of players' differences in these aspects on the game progression.

c) *The player-avatar spoken interaction should be varied and natural*. To extend the scenario for entrainment~trust analysis in [10], in which players could only ask wh-questions and avatars could only answer by providing advice, avatars here ask questions and provide advice and players respond to questions and select advice either verbally or by clicking. Each decision sequence contains a mini-dialogue which is managed through a decision tree-based conversational model with key-word spotting. This is initiated by an avatar's question relevant to the current decision (e.g. '*please describe the girls in the room'* in Fig 1 top-right), then the player responds, and, if the relevant keywords have been recognized, the avatars proceed to provide advice (e.g. which girl should the player talk to), and the player decides to follow one of the two pieces of advice. The 'push-to-talk' mode for players' speech lowers the demands on Automatic Speech Recognition (ASR).

d) *The game should be engaging.* Despite the visually simple interface, the player is faced with different types of tasks and challenges, is motivated to be competitive with a promise of a financial bonus dependent on the game outcome. The design includes simple fun games and jokes.

e) *Initial baseline values for the player's natural utterances are needed*. After familiarizing themselves with the interface, players must take an Oath; see Fig 1 top-left. The player is asked to respond to easy questions (e.g. *How old are you*), to read short sentences (e.g. *I swear to uphold the rules*…), and to finish proverbs (e.g. *One apple a day*…). In this way we elicit utterances that are similar in length, spontaneity, and prosody to those expected in the game.

f) *Vary the stakes for some decisions*. This is achieved by manipulating the mini-dialogue that initiates each decision by including avatars' verbal remarks such as *think hard, you will not have another possibility*, or *you might die*.

g) *All relevant information should be logged*. Each action of the player is logged, all speech is recorded, and the log also include the feature vectors used for the (dis)entrainment manipulation (section 2.2).

2.2. Voice interaction: ASR, TTS, and Entrainment

The voice interaction between the player and the game relied on customized Text-to-speech (TTS) and ASR systems. For base TTS we used a female voice trained on 8k phonetically rich sentences (5h) using the HMM-based speech synthesis system [18, 19]. The ASR engine was based on Open-Source Large Vocabulary CSR Engine Julius [20]. We trained the n-gram language model with a manually created set of sentences and words expected to occur in the players' utterances. We also enriched the language model with the actual players' responses during the pilot testing. The n-gram model is more suitable for this task than grammar construction due to great variability in the player's responses and the importance of correct key words recognition for the game to proceed successfully.

The mean f0 and intensity values were extracted from each recorded utterance with a customized Praat [21] script. The speech rate in terms of syllables per second was extracted from the ASR output since this method proved more reliable over a signal-based method in pilot testing. To achieve a reliable fit between the measured and perceived speech rate, we adjusted the measured rate by normalizing several factors known to affect the discrepancy between the measured and perceived rates [22]. Specifically, at the syllable level we normalized for vowel quality and the number and type of consonants, and on the utterance level for the number of phones, syllables (containing phonemically short and long vowels) and words, and the number of silent and filled pauses. Since a database with annotated perceived speech rate was not available, we used TTS for creating sets of corpora in which only a single factor from the above list was varied. For each such sub-corpus we obtained a function, or a weighing table, minimizing the variability of the measured speech rate (since the TTS was assumed to produce speech of the stable rate). The output speech rate of an utterance then combined the raw syllable rate and these weighing factors.

The prosodic entrainment, operationalized here as relative turn-based synchrony, was done by adjusting the base TTS in the following way. The baseline values for the subject's pitch, intensity and speech rate were obtained from their responses to the 'Oath' at the beginning of the game; e) in 2.1. For each utterance of the player during the game and after taking the Oath, we extracted and stored the means of these three values. For the speech of the entraining avatar we took the relative difference between the most recent utterance of the player and her baseline, and adjusted the pitch, intensity and speech rate of the base TTS by this factor in the same direction. The same was done for the disentraining avatar in the opposite direction.

The pilot testing showed little variability in the distribution of the players' utterance means for the three features varied for entrainment (f0, intensity, speech rate). Consequently, the adjustments for the (dis)entraining voice were often miniscule and the differences between the two voices were thus minimal. To separate the two voices more while remaining within the reasonable regions of variability, we decided to 'boost' the (dis)entraining feature of the voices by increasing the adjustment coefficients while keeping the same direction following the conversion Table 1.

Table 1: *Conversion for boosting (dis)entrainment.*

Current utterance – baseline (%)	
Measured	**Implemented to TTS**
1-3	7
4-6	12
7-10	15
11-15	18
16-21	22
>22	25

2.3. Subjects and procedure

Native speakers of Slovak (N=32, 18F, 14M) participated in the experiment. The majority were undergraduate students and received small payment. Within each sex group, half the subjects played the game in which avatar M was entraining and avatar N was disentraining, and for the other half the entraining behavior of the avatars was switched. Subjects wore over-ear headphones fitted with close-talk microphones

After providing informed consent subjects filled out a sociometric questionnaire and the Ten Item Personality Inventory (TIPI, [23]). Then they familiarized themselves with the setup of the game and were instructed to try to speak in complete sentences. Importantly, subjects were told that the two avatars are similar but use slightly different AI algorithms regarding their ability (not integrity or benevolence), and thus if the subjects wanted to do well, they should try to figure out who tends to give better advice. Doing well means maximizing the number of coins earned, which will be the basis for calculating their bonus payment. In reality, the quality of the advice was never revealed. After the game subjects filled out a post-test questionnaire probing their perception of the avatars and their voices.

Due to unforeseen problems with the design, 8 subjects did not complete the game and thus fewer than the planned 37 decisions were recorded for these subjects. Of the 8 subjects, 6 were females, 2 males, 6 made more than 2/3 of the decisions, for 3 subjects avatar M was entraining for 5 avatar N. These data were kept and the findings hold for the subset of 24 subjects who completed the full game.

3. Results

The core question of this study was whether the manipulation of the acoustic-prosodic features (pitch, intensity and speech rate) in the avatar's voices to either entrain or disentrain to the user's voice affects the user's trust toward the advice provided by these avatars, and if this trust-entrainment relationship is affected by the sex and the personality of the subject.

There were two dependent variables assessing the trust of the subjects toward the avatars' advice intrinsically: *Raw_score* is the sum of the decisions in which the user followed the advice of the entraining and disentraining avatars during the game, *Raw_Trust_Index* is then the difference between these two numbers; positive values reflect the preference for the entraining avatar while negative values indicate a preference for the disentraining avatar. *Weighted_score* is similar to *Raw_score* but indicates the sum of the decision's ordinal numbers in which the user followed the advice of the entraining and disentraining avatar. *Weighted_score* thus tests the hypothesis that the relationship between speech entrainment and trust develops over the course of the game and becomes stronger toward the game's end since in this metrics later decisions are weighted more than earlier ones.

To test the effects of entrainment behavior of avatars and subject's sex on subject's trusting preferences we constructed a regression model for both *Raw_Score* and *Weighted_Score*. The models showed that subjects follow the advice from the disentraining avatar more often ($t(60) = -3.4$, $p = 0.001$) and that the interaction between sex and entrainment preferences was also significant ($t(60) = 3.04$, $p = 0.004$) in that the main effect is mostly due to females while males do not show a strong preference ($F(1,30) = 6.36$, $p = 0.0172$). This is illustrated in the left panel of Fig. 2.

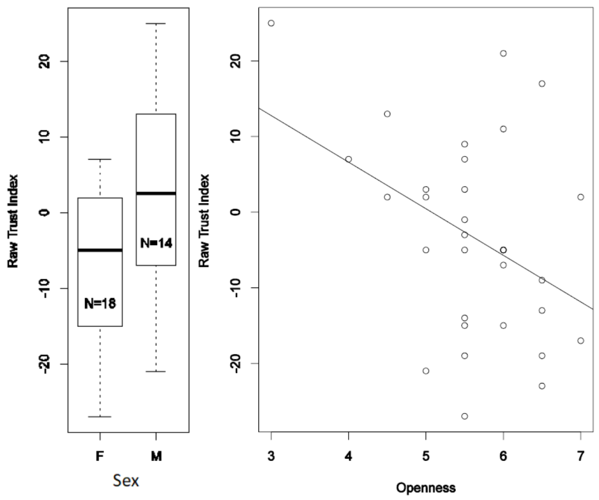

Figure 2: *Effect of entrainment on trust for based on sex (left) and its relationship with openness; see text for the explanation of Raw Trust Index.*

The same effect holds also for *Weighted_Score* although the t-values are lower and thus the effect is less robust ($t(60) = -2.1$, $p = 0.04$ and $t(60) = 2.4$, $p = 0.02$ respectively).

Finally, recall that the final decision carried the highest stakes since pulling the wrong lever may have resulted in the player's 'death' in the game. In both sex groups, subjects who reached this final decision (13F, 11M) followed the advice of the entraining avatar slightly more often (7 vs. 6 for females, 7 vs. 4 for males) but this preference is certainly not significant and can be taken as a trend at most.

We also examined subjects' self-reported responses to post-test questions about avatars and their voices in the three questions: Which avatar did you like more? Which avatar's voice did you like more? Which avatar gave better advice? The percentages of the binary choice (entraining vs. disentraining) split for sex is shown in Figure 2.

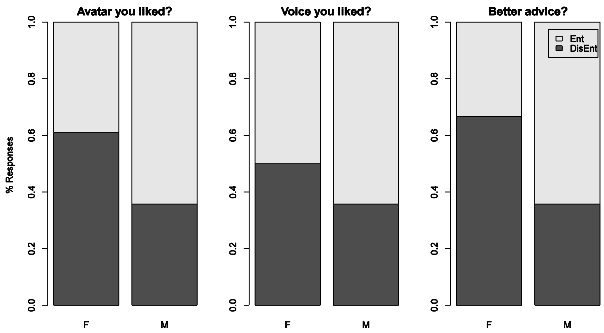

Figure 2: *Responses to the post-test questions probing selfreported preference for the avatars based on their general features, voice, and the quality of advice.*

First, we see a possible link between trust behavior expressed implicitly through following the advice during the game and explicitly as perceiving an avatar as more likeable and giving better advice: females who followed the advice of the disentraining avatar more often also perceived this avatar as more likeable and preferred its advice (NB: the design prevented the subject to have any game-internal evidence for the quality of advice and the two avatars looked extremely alike). Second, the females' answers to the 2^{nd} question (Whose voice did you like more) does not show any preference despite voice being the only avatar's characteristic that was varied in the experiment. This might suggest either a possible dissociation between the general preference for the avatar and the perception of its voice, or that the liking of the entraining avatar's voice for some subjects was not strong enough to affect their advice following behavior. Third, the males liking of the entraining avatar did not result in significantly greater trust toward the advice given by this avatar. Moreover, for males the three questions were treated as probing the same perception (most males responded to all three questions with the identical avatar identity) while females showed a clear divergence between their perceptions for avatar's voice on the one hand and general liking and advice preference on the other hand.

Testing the relationship between personality and the trust-entrainment link we found that openness to experience was the only dimension that showed a significant relationship with the Trust index. As shown in the right panel of Fig. 1, subjects who score higher on openness favored the disentraining avatar more. This relationship held for females ($t[16] = -2.27$, $p = 0.038$) as well as for the pooled data ($t[30] = -2.4$, $p = 0.023$). Despite literature suggesting a rather strong correlations between Openness and Extroversion, this association was rather weak in our data ($r = 0.31$; $t[30] = 1.76$, $p = 0.09$).

4. Discussion & Conclusion

We tested a possible link between subject's trust toward an avatar, expressed as the tendency to follow its advice, and acoustic-prosodic entrainment of the avatar's voice towards (or away from) the subject's characteristics. Our main finding is that females tend to trust the avatar whose mean pitch, intensity and speech rate in a turn are locally disentraining (i.e. move in the opposite direction) from the values in the preceding subject's turn. This result contributes to the growing recent literature suggesting the complex role of speech entrainment in both human-human and human-machine spoken interactions.

Furthermore, the effect of entrainment on trust was observed only when the manipulation of the prosodic features was boosted to counter the low variability of these features in the current data. This finding opens further possibilities for designing applications with speech entrainment capabilities.

Regarding the link between personality and trust toward the (dis)entraining avatar, we found that people who score higher on openness to experience are more likely to trust the avatar that disentrains to their own vocal features. This might stem from openness being related to preference for variety [24], and the link between openness and high tolerance for ambiguity.

The similarity of the effects for weighted and raw scores suggests that the relationship between trust and speech entrainment changes minimally during roughly 45 minutes of game playing. However, two tendencies showing a slightly weaker link between trust and disentrainment toward the end of the game also allow us to speculate that females initially preferred the disentraining avatar but this preference weakened over the course of the game (less robust results for weighted scores and the tendency from the last decisions with the highest stakes). This in turn might mean that preference for disentrainment is an early 'automatic' effect [6, 25] and preference for entrainment takes some time to develop and is under more subconscious control and used for negotiating social ties as suggested by CAT [4].

At least there limitations of the study may have contributed to the lack of the link between trust and prosodic entrainment. First, avatars looked 'dehumanizing' and the subjects did not have a prior chance to develop a relationship and emotional attachment to them. Yet, social interaction has been found to positively affect the development of trust toward robots in other research [26]. Second, 'noise' introduced by the voices of other characters in the game could affect the development of trust toward the avatar voices. Third, we manipulated only means that has a limited effect on prosody and adapting other aspects of the contours might facilitate entrainment-trust link more.

Finally, another intriguing hypothesis for future research is that a relatively weak link between entrainment and trust in Slovak and the preference for disentrainment in females in particular, might be connected to cultural differences. Comparisons between Slovak and English suggest that Slovak subjects display less entrainment in human-human conversations [27, 28] and may also, for some prosodic features, (e.g. speech rate) reveal a link between trust and disentrainment [10].

5. Acknowledgements

This material is based upon work supported by the Air Force Office of Scientific Research, Air Force Material Command, USAF under Award No. FA9550-15-1-0055.

6. References

[1] S. Gregory, S. Webster, and G. Huang, “Voice pitch and amplitude convergence as a metric of quality in dyadic interviews,” *Language & Communication*, vol. 13, no. 3, pp. 195–217, 1993.

[2] J. H. Manson, G. A. Bryant, M. M. Gervais, and M. A. Kline, “Convergence of speech rate in conversation predicts cooperation,” *Evolution and Human Behavior*, vol. 34, no. 6, pp. 419–426, 2013.

[3] R. Levitan and J. Hirschberg, “Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions”, in *Proc. Interspeech*, 2011.

[4] H. Giles, N. Coupland, and J. Coupland, “Accommodation theory: Communication, context and consequence,” in *Contexts of Accommodation*, H. Giles, N. Coupland, and J. Coupland, Eds. Cambridge University Press, 1991, pp. 1–68.

[5] K. G. Niederhoffer and J. W. Pennebaker, “Linguistic style matching in social interaction,” *Journal of Language and Social Psychology*, vol. 21, no. 4, pp. 337–360, 2002.

[6] M. Pickering and S. Garrod, "An integrated theory of language production and comprehension," Behavioural and Brain Sciences, vol. 36, no. 4, pp. 329-347, 2013.

[7] L. Bell, J. Gustafson, and M.Heldner, „Prosodic adaptation in human-computer interaction,“ in *Proceedings of International Congress of Phonetic Sciences*, pp. 2463-2466, 2003.

[8] J. Thomasson, H. V. Nguyen, and D. Litman, “Prosodic Entrainment and Tutoring Dialogue Success,” in *Artificial Intelligence in Education. AIED, Lecture Notes in Computer Science, vol. 7926* H. C. Lane, K. Yacef, J. Mostow and P Pavlik Eds. Berlin, Heidelberg: Springer, pp. 750-753, 2013.

[9] N. Lubold, H. Pon-Barry, and E. Walker, “Naturalness and rapport in a pitch adaptive learning companion,” in IEEE Automatic Speech Recognition and Understanding Workshop, 2015.

[10] R. Levitan, Š. Beňuš, R. H. Gálvez, A. Gravano, F. Savoretti, M. Trnka, A. Weise, and J. Hirschberg, “Implementing acoustic-prosodic entrainment in a conversational avatar,” Proc. Of Interspeech 2016, pp. 1166–1170.

[11] J.M. Perez, R.H. Galvez, and A. Gravano, “Disentrainment may be a positive thing: A novel measure of unsigned acoustic-prosodic synchrony, and its relation to speaker engagement”, in Proc. of Interspeech 2016, pp. 1270-74.

[12] J. Cole and U. Reichel, “Prosodic entrainment: The cognitive encoding of prosody and its relation to discourse function,“ Talk presented at Speech Prosody 2016.

[13] A. Schweitzer and M. Walsh, ”Exemplar Dynamics in Phonetic Convergence of Speech Rate,” in *Proc. Interspeech*, 2016.

[14] L.L. Namy, L.C. Nygaard, and D. Sauerteig, “Sex differences in vocal accomodation: the role of perception,” Journal of Personality and Social Psychology, vol. 21, no. 4, pp. 422-432, 2002.

[15] R. Levitan, A. Gravano, L. Willson, Š Beňuš, J. Hirschberg, and A. Nenkova, “Acoustic-prosodic entrainment and social behavior,“ in *Proceedings of the 2012* NAACL HLT, pp. 11-19.

[16] E. Strupka, O. Niebuhr, and K. Fischer, “Influence of Robot Sex and Speaker Sex on Prosodic Entrainment in HRI,” in IEEE International Symposium on Robot and Human Interactive Communication, 2016.

[17] R. C. Mayer, J. H. Davis, and F. D. Schoorman, “An Integrative Model of Organizational Trust,” The Academy of Management Review, vol. 20, no. 3, pp. 709-734, 1995.

[18] M. Sulír, J. Juhár, and M. Rusko, “Development of the Slovak HMM-Based TTS System and Evaluation of Voices in Respect to the Used Vocoding Techniques,” *Computing and Informatics*, vol. 35, no. 6, pp. 1467-1490.

[19] H. Zen, T. Nose, J. Yamagishi, S. Sako, T. Masuko, A.W. Black, and K. Tokuda, “The HMM-based speech synthesis system version 2.0,” Proc. of ISCA SSW6, Bonn, Germany, Aug. 2007. [http://hts.sp.nitech.ac.jp/]

[20] A. Lee and T. Kawahara. "Recent Development of Open-Source Speech Recognition Engine Julius" Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), 2009.

[21] P. Boersma and D. Weenink, “Praat: doing phonetics by computer” [Computer program, http://www.praat.org/]

[22] D. Gibbon, K. Klessa, and J. Bachan, “Duration and speed of speech events: A selection of methods,” *Lingua Posnaniensis*, vol. 56, no 1, pp. 59–83, 2014.

[23] S. Gosling, P. Rentfrow, and W. Swann, “A very brief measure of the big-five personality domains,” *Journal of Research in Personality*, vol. 37, no.6, pp. 504–528, 2003.

[24] P. T. Costa, and R. R. McCrae, *NEO Personality Inventory Professional Manual*. Odessa, FL: Psychological Assessment Resources, 1992.

[25] T. L. Chartrand, and J. A. Bargh, “The chameleon effect: The perception-behavior link and social interaction,” Journal of Personality and Social Psychology, vol. 76, no. 6, pp. 893–910, 1999.

[26] M. Lohani, C. Stokes, M. McCoy, C. A. Bailey, and S. E. Rivers, "Social interaction moderates human-robot trust-reliance relationship and improves stress coping," *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 471-472.

[27] R. Levitan, Š. Beňuš, A. Gravano, and J. Hirschberg, “Acoustic-prosodic entrainment in Slovak, Spanish, English and Chinese: A cross-linguistic comparison,” in Proc. 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue, pp. 325-334, 2015.

[28] Š. Beňuš, R. Levitan, J. Hirschberg, A. Gravano, and S. Darjaa, “Entrainment in Slovak collaborative dialogues,” in Proc. 5th IEEE Conference on Cognitive Infocommunications, pp. 309-313, 2014.