AFRL-AFOSR-UK-TR-2018-0041

Relationship between trust and entrainment in speech

Stefan Benus
Univerzita Konstantina Filozofa v Nitre
1 Tr. A. Hlinku
Nitra, Nitra, 94974
SK

09/25/2018
Final Report

Air Force Research Laboratory
Air Force Office of Scientific Research
European Office of Aerospace Research and Development
Unit 4515 Box 14, APO AE 09421

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
25-09-2018	Final	15 Dec 2014 to 14 Dec 2017

4. TITLE AND SUBTITLE
Relationship between trust and entrainment in speech

5a. CONTRACT NUMBER

5b. GRANT NUMBER
FA9550-15-1-0055

5c. PROGRAM ELEMENT NUMBER
61102F

6. AUTHOR(S)
Stefan Benus

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Univerzita Konstantina Filozofa v Nitre
1 Tr. A. Hlinku
Nitra, Nitra, 94974 SK

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
EOARD
Unit 4515
APO AE 09421-4515

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL/AFOSR IOE

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-AFOSR-UK-TR-2018-0041

12. DISTRIBUTION/AVAILABILITY STATEMENT
A DISTRIBUTION UNLIMITED: PB Public Release

13. SUPPLEMENTARY NOTES

14. ABSTRACT
We have designed experimental scenarios and collected data from Slovak and Argentinian Spanish testing the effect of manipulating acoustic-prosodic features of synthesized voices, to either entrain or disentrain with respect to the human user's voice, on trust of users toward the avatars with these synthesized voices. The results of these experiments and our continued exploration of human-human entrainment in spoken dialogue corpora provide a rather complex and multidimensional picture of prosodic entrainment and its potential for calibrating trust in human-machine spoken interactions. The key results include: 1) not only entrainment but also disentrainment in terms of acoustic-prosodic features might facilitate trust of humans towards computer avatars, 2) different prosodic features affect the trust-entrainment relationship differently, 3) most of our results were consistent for both Slovak and Argentinian Spanish, which points to complex, yet systematic and cross-culturally valid, nature of the trust-entrainment relationship, and 4) the observations from human-human dialogues are generally consistent with human-machine experiments and offer also novel factors, such as biological sex and task-role, that affect speech entrainment and the entrainment-trust relationship.

15. SUBJECT TERMS
EOARD, Entrainment, Trust

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			RIECKEN, RICHARD
Unclassified	Unclassified	Unclassified	SAR		19b. TELEPHONE NUMBER *(Include area code)* 703-941-1100

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Report Type: Final Report

Primary Contact E-mail: sb513@nyu.edu

Primary Contact Phone Number: +421911275318

Organization / Institution name: Univerzita Konstantina Filozofa v Nitre

Grant/Contract Title: Relationship between trust and entrainment in speech

Grant/Contract Number: FA9550-15-1-0055

Principal Investigator Name: Stefan Benus

Program Manager: Benjamin A. Knott

Reporting Period Start Date: 12/15/2014

Reporting Period End Date: 12/14/2017

Abstract

We have designed experimental scenarios and collected data from Slovak and Argentinian Spanish testing the effect of manipulating acoustic-prosodic features of synthesized voices, to either entrain or disentrain with respect to the human user's voice, on trust of users toward the avatars with these synthesized voices. The results of these experiments and our continued exploration of human-human entrainment in spoken dialogue corpora provide a rather complex and multidimensional picture of prosodic entrainment and its potential for calibrating trust in human-machine spoken interactions. The key results include: 1) not only entrainment but also disentrainment in terms of acoustic-prosodic features might facilitate trust of humans towards computer avatars, 2) different prosodic features affect the trust-entrainment relationship differently, 3) most of our results were consistent for both Slovak and Argentinian Spanish, which points to complex, yet systematic and cross-culturally valid, nature of the trust-entrainment relationship, and 4) the observations from human-human dialogues are generally consistent with human-machine experiments and offer also novel factors, such as biological sex and task-role, that affect speech entrainment and the entrainment-trust relationship.

Report

Our goal was to test the hypothesis that automated systems (e.g. avatars) with synthesized voices that are modified to become more prosodically entrained (similar) to the voices of their human interlocutors are more trustworthy for these interlocutors than the avatars whose voices are not manipulated this way. Furthermore, by using data primarily from Slovak and Argentinian Spanish, and comparing them to the available data from English, we were interested in finding the aspects of the entrainment-trust relationship that might be cross-culturally valid and those

that might vary across various languages and cultures. Our hypothesis is motivated by several studies of human-human dialogues showing positive effects of speech entrainment on several aspects of these conversations.

Our approach was based on 1) designing novel scenarios for probing the hypothesized link between prosodic entrainment and trust from different angles, 2) using the developed scenarios for collecting and analyzing the data in experiments in Slovakia and Argentina, and 3) continuing the investigations of human-human conversations with the goal of better understanding the factors underlying human-human entrainment in spoken interactions.

Regarding 1) we have created four games and associated protocols to use as experimental tools: *GoFish-v2*, *True/False*, *NavGame*, and *GuessWho*. *GoFish-v2* is a card game modified from the original GoFishWithHelpers design of R. Levitan in which the subject plays against the computer and elicits cards from the computer's hand to collect sets of four cards of identical value. Before each turn, the subject asks one of two avatars for advice which card to ask for and has to follow this advice. One of the avatars entrains its voice to the subset of prosodic features extracted from each request for advice of the user. The quality of advice is controlled and the **trust measure** thus corresponds to **which avatar the subject asks more often** during the game; the one whose voice entrains to the subject or the other one.

True/False is a game in which the subject hears statements of unknown veracity and guesses if the statement is true or false. The statements are synthesized by using three levels of prosodic features: mean pitch, intensity and speech rate. We can thus examine if the operationalization of **trust as belief in the veracity of the statement** is affected by prosodic features. Moreover, having baseline values for the speaker's pitch, intensity and speech rate, a post-hoc analysis can test the relationship between the prosodic similarity of the subject's prosody and the prosody of the statements on the one hand and the subject's belief in their veracity on the other hand.

NavGame is a one-person adventure game motivated by the Harry Potter character. The subject is faced with multiple decisions and is advised before each one which choice to take by two owl-like helpers. The voices of the helpers are modified in the same way as in *GoFish-v2*. The game involves much more spontaneous and varied speech from the subject and the avatars compared to *GoFish-v2*, and the option to vary the stakes in the decision. The **trust measure** corresponds to **whose advise the subject follows more often** during the game.

Finally, *GuessWho* is a game in which the subject and a computer avatar take turns to ask yes-no questions (e.g., *does your character have curly hair?*) in order to eliminate irrelevant faces from their own board, and ultimately guess the opponent's target face. To avoid ASR errors, we use a Wizard-of-Oz setup, in which a hidden human operator (the *wizard*) is responsible for all speech recognition and understanding. In our protocol, the subject plays two games against each of the two avatars; one with entraining and one with disentraining voice features. Subsequently, the subject is asked to choose one of the two avatars, against whom they will have to play one more game for extra monetary bonus if they win. Hence, the **trust measure** concerns the **selection of the avatar for the final game**: we expect subjects should select the avatar whose skills they trust less.

In all of the games we designed an online system for (dis)entrainment modifications to the text-to-speech synthesized voices testing the concept of dynamically adjusting TTS during the interaction rather than modifications of pre-recorded utterances. For some games (*GoFish-v2* and *NavGame*) also automatic speech recognition (ASR) system was required and it was either designed or an existing one was implemented. In both Slovakia and Argentina we collected data from *GoFish-v2* and *True/False* while *NavGame* was tested in Slovakia and *GuessWho* in Argentina only.

The results of experiments using these research tools and continued exploration of entrainment in spoken dialogue corpora provide a **rather complex and multidimensional picture of prosodic entrainment and its relationship to trust** in human-machine spoken interactions. The key findings can be summarized as follows:

1. Not only entrainment but also disentrainment in terms of acoustic-prosodic features might facilitate trust of humans towards computer avatars.
2. Different prosodic features affect the trust-entrainment relationship differently.
3. Most of our results were consistent for both Slovak and Argentinian Spanish, which points to complex, yet systematic and cross-culturally valid, nature of the trust-entrainment relationship.
4. The observations from human-human dialogues are generally consistent with human-machine experiments and offer also novel factors, such as biological sex and task-role, that affect speech entrainment and the entrainment-trust relationship.

Our original hypothesis was that acoustic-prosodic entrainment of avatar's voices positively affects trust of humans toward this avatars. This hypothesis was supported with *GoFish-v2* with both Slovak and Argentinian subjects when the modification of the base TTS system adjusted all three investigated features: pitch, intensity, and speech rate (Benus et al., in prep.). Moreover, the prosodic similarity between the subject's speech and the statements in the *True/False* game suggested a facilitatory effect of this similarity in guessing the veracity of statements (Gálvez et al. 17). On the other hand, the beneficial effects of acoustic-prosodic **dis-entrainment** on the development of human trust toward avatars was observed a) in *GoFish-v2* when only speech rate was manipulated for both Slovaks and Argentinians (Levitan et al. 16), b) in Slovak females playing *NavGame* when prosodic dis-entrainment was boosted to counter low variability of the subject' speech (Benus et al. 18), and c) in Argentinian *GuessWho* when all three features were modified (Gauder et al. 18).

The **importance and relevance of both entrainment and dis-entrainment** (for trust and other aspects of conversations) were corroborated in analyses of **human speech corpora** containing task-oriented dialogues in Slovak, Spanish, and English. In English, the perceived level of engagement, encouragement, control, or likability in dialogues, all plausibly linked to trust, correlated with a discrete operationalization of prosodic entrainment (Gravano et al. 15). However, in the same data, a measure combining **both** entrainment and disentrainment showed a robust significant correlation with the degree of speaker engagement (Perez et al. 16). In a comparative Slovak corpus, features extracted with a novel approach to the bottom-up parametrization of prosodic characterization of utterances revealed a complex interplay of entrainment and disentrainment on the global and local levels and in different sets of prosodic

features (Reichel et al, accepted with revisions). Hence, the multifaceted cognitive system underlying the way we cue discourse, pragmatic, and interpersonal meanings by systematic variations in our speech prosody is a rich ground for further explorations.

The **consistency of results for both Slovak and Argentinian Spanish** was observed in both rounds of *GoFish-v2* as well as in the *True/False* games (Gálvez et al. 16). Additionally, the same pattern of association of trust and prosodic disentrainment was observed in Slovak *NavGame* and Spanish *GuessWho* (Benus et al. 18, Gauder et al. 18). This consistency was corroborated by finding similar patterns of acoustic-prosodic entrainment in human-human dialogues in Spanish and Slovak observed in comparable corpora and analyzed with identical methods for feature extraction and statistical analyses (Levitan et al. 15a). This suggests that when the implementation of the human-machine entrainment is consistent across language and culture, the relationship between prosodic entrainment, in terms of mean pitch, intensity and speech rate, and trust seems complex, yet cross-linguistically and cross-culturally stable.

Another novel (and unexpected) observation is that speakers' **biological sex** and their **role** in task-oriented dialogues might affect both entrainment itself as well as its relationship to trust (Reichel et al, accepted with revisions). In Slovak human-human corpus, females were found to entrain more than males and people describing the situation (describers) entrained more than those who needed to act based on that description (followers). Yet, males were not worse in solving the task than females and we hypothesize that entrainment might be used differently across genders in cooperative solution-oriented interactions: Male describers might tend to mark hierarchy/authority by disentrainment, which can be as beneficial as the female strategy of common ground creation by entrainment. As followers, females and males were similar. In other words, male followers seem to entrain more, maybe to signal that they accept the describer's authority, and female followers appear to entrain less, to signal it is less their but rather the describer's responsibility to establish common ground.

Finally, in the last year of the grant we started probing the entrainment-trust relationship when humans communicate with **robots**. In collaboration with Angelo Cangelosi (PI for another research sponsored by AFOSR) and his students we implemented the *GoFish-v2* game to replace static avatar pictures with embodied humanoid Nao robots. Although the pilot testing has not shown a significant association between trust and prosodic modification in general, we did find an interaction between this relationship and personality traits (disentrainment seems to correlate with openness to new experience and agreeableness, Benus et al. 18b). Additionally, we have acquired the social robotic head *Furhat* and are in the process of modifying our game protocols for testing the entrainment-trust relationship when people interact with robots capable of social signalling.

In future, we plan to modify our scenarios to involve more **engagement** and **emotion** since the ones we used present relatively little variability and thus a decreased potential for producing and perceiving speech entrainment. We also plan to explore the manipulation of **turn-taking behavior** in addition to prosodic features since our research in human-human dialogues suggest positive effects of entrainment on turn-taking such as backchanneling and latency (Levitan et al. 15b, Benus 16). Additionally, we will expand on our initial exploration of the

potential links between entrainment-trust and **personality** (Benus et al. 18a, Benus et al. 18b). Finally, using a technique known as voice morphing, we are developing tools for exploring how avatar's voices synthesized with an **accent** (e.g. an avatar speaking English with 'Slovak' accent to resemble Slovaks who speak English) might affect trust of humans towards such avatars.

Archival Publications:

1. Beňuš, Š. (2016). The prosody of backchannels in Slovak. In: *Proceedings of 8th International Conference on Speech Prosody, pp. 75-79*. Boston, MA, May 2016. [DOI: 10.21437/SpeechProsody.2016-85]

2. Beňuš, Š., Trnka, M., Kuric, E., Matrák, L., Gravano, A., Hirshberg, J., Levitan, R. (2018a). Prosodic entrainment and trust in human-computer interaction. Proceedings of the 9th Speech Prosody. Poznan, Poland, May 2018.

3. Beňuš, Š., Patacchiola, M., Trnka, M., Zanatto, D., Sabo, R., Cangelosi, A. (2018b). Do people trust robots whose prosody synchronizes with the user? Proceedings of Cognition and Artificial Life conference. Brno, Czech Republic, June 2018.

4. Gálvez, R., Beňuš, Š., Gravano, A., Trnka, M. (2017). Prosodic facilitation and interference while judging on the veracity of synthesized Statements. *Proceedings of INTERSPEECH*, pp. 2331-2335. Stockholm, Sweden, Aug 2017. [DOI: 10.21437/Interspeech.2017-453]

5. Gauder, L., Reartes, M., Gálvez, R., Beňuš, Š., Gravano, A. (2018). Testing the Effects of Acoustic/Prosodic Entrainment on User Behavior at the Dialog-Act Level. Proceedings of the 9th Speech Prosody. Poznan, Poland, May 2018.

6. Gravano, A., Beňuš, Š., Levitan, R., Hirschberg, J. (2015). Backward Mimicry and Forward Influence in Prosodic Contour Choice in Standard American English. *Proceedings of INTERSPEECH*, pp. 1839-43. Dresden, Germany, Sep 2015. [http://www.isca-speech.org/archive/interspeech_2015/i15_1839.html]

7. Levitan, R., Beňuš, Š., Gravano, A., Hirschberg, J. (2015a). Acoustic-prosodic entrainment in Slovak, Spanish, English and Chinese: A cross-linguistic comparison. *Proceedings of 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pp. 325-334. Prague, Czech Rep., Sep 2015. [DOI: 10.18653/v1/W15-4644]

8. Levitan, R., Beňuš, Š., Gravano, A., Hirschberg, J. (2015b). Entrainment and Turn-Taking in Human-Human Dialogue. Proceedings of AAAI Spring Symposium on Turn-Taking and Coordination in Human-Machine Interaction. Stanford, CA, USA, Mar 2015. [http://www.aaai.org/ocs/index.php/SSS/SSS15/paper/viewFile/10226/10106]

9. Levitan, R., Beňuš, Š., Gálvez, R., Gravano, A., Savoretti, F., Trnka, M., Weise, A., Hirschberg, J. (2016). Implementing Acoustic-Prosodic Entrainment in a Conversational

Avatar. In: *Proceedings of INTERSPEECH*, pp. 1166-1170. San Francisco, CA, Sep 2016. [DOI: 10.21437/Interspeech.2016-985]

10. Perez, J.M., Gálvez, R.H., Gravano, A. (2016). Disentrainment may be a positive thing: A novel measure of unsigned acoustic-prosodic synchrony, and its relation to speaker engagement. In: *Proceedings of INTERSPEECH*, pp. 1270-74. San Francisco, CA, Sep 2016. [DOI: 10.21437/Interspeech.2016-587]

11. Reichel, U., Beňuš, Š., Mády, K. (accepted with revisions). Entrainment profiles: Comparison by gender, role, and feature set. Speech Communication.

Invited talks

- Štefan Beňuš: Relationship between trust and entrainment in speech. Institute for Natural Language Processing, University of Stuttgart, Germany, SFB guest talk series. November 2017.
- Agustín Gravano: Modeling coordination in dialogue. Unidad de Neurobiología Aplicada, CEMIC, Buenos Aires, August 2017.
- Štefan Beňuš & Uwe Reichel: Entrainment profiles: Comparison by gender, role, and feature set. Workshop of SimPhon.Net, funded by the German Research Foundation (DFG), Mátraháza, Hungary. July 2017.
- Agustín Gravano: Modeling coordination in dialogue. Instituto de Cálculo, Universidad de Buenos Aires, July 2017.
- Štefan Beňuš: Speech entrainment: Approaches and case studies. University of Texas El Paso, Departments of linguistics and computer science. January 2017.
- Štefan Beňuš: Backchannels, turn-taking, and entrainment in Slovak. CoFee (Conversational feedback) workshop, Aix-en-Provence,France, November 2015.

Publicity

- Interview with Stefan Benus regarding the development and design of NavGame for testing the relationship between trust and entrainment (Slovak national television)
- Interviews with Agustin Gravano on Spoken Dialogue Systems and Artificial Intelligence (Argentinian radio, newspapers)
- Interview with Stefan Benus about exploring trust-entrainment relationship when people play game with Nao robots (Slovak national radio)
- Press report appearing in various Slovak newspapers regarding the experiment testing trust of humans towards Nao robots when the prosody of their voices matches the prosody of their human interlocutors

Changes in research objectives (if any):

NA

Change in AFOSR Program Manager, if any:

NA

Extensions granted or milestones slipped, if any:

NA