Speech Prosody 2016
31 May - 3 Jun 2106, Boston, USA

Behavioural mediation of prosodic cues to implicit judgements of trustworthiness

Ilaria Torre[1], *Laurence White*[1], *Jeremy Goslin*[1]

[1]School of Psychology, Plymouth University, U.K.

ilaria.torre@plymouth.ac.uk

Abstract

Prosodic information is known to play a role in personality attributions, such as judgements of trustworthiness. Research so far has focused on assessing the determinants of such attributions in static contexts, very often in the form of questionnaires, and not much is known about their dynamics, in particular, how direct experience of behaviour over time influences the interpretation of vocal characteristics. We used the investment game, an innovative methodology adapted from game theory studies, to assess how trust attributions – to virtual players acting more or less cooperatively – are affected by the prosodic characteristics of speakers of a range of British English accents. Regression analysis shows that speaker accent, mean pitch, and articulation rate all influence participants investment decisions, our implicit measure of trust. Furthermore, participants interpretations of these prosodic characteristics interact with how the virtual players behave over time. Our findings are discussed with reference to "Size/Frequency Code" and "Effort Code" accounts of prosodic universals.

Index Terms: game theory, trust, accent, mean pitch, articulation rate

1. Introduction

We make personality inferences spontaneously [1, 2, 3], and evidence suggests that the voice of a person just encountered affects the judgements formed almost immediately in the mind of a listener [4]. Here we consider the role of voice characteristics, and prosody in particular, with respect to attributions of trustworthiness, for which there are limited and inconsistent data. In an interview scenario, where participants had to rate the perceived trustworthiness of embodied conversational agents, high f0 early in the interaction predicted low trustworthiness attributions [5]. This finding was consistent with an earlier study based on questionnaire responses, where participants rated speakers with high f0 as "less truthful" [6]. The same study also found that speakers with slow rate of speech were rated as "less truthful" and "more passive" [6]. On the other hand, smiling when speaking (which tends to raise f0 [7]) was found to boost trustworthiness judgements [5]. Finally, accents have long be linked to trustworthiness attributions [8, 9, 10] and were recently shown to affect implicit judgements of trustworthiness, with standard accents being associated with higher trust levels than regional accents [11].

Studies researching specifically the effect of voice characteristics on trustworthiness are scarce, but inferences can be drawn from the literature on the role of voice in deception and in attributions of personality more in general, e.g., the "Big Five" traits [12]. Findings are not wholly consistent, however. Concerning pitch, for example, participants in one deception study consistently raised their f0 when lying [13], while participants in another study *perceived* a lower pitch in the deceptive messages they heard [14]. A further study failed to find any acoustic differences – including in f0 – in the production of deceptive and truthful messages [15]. From a slightly different perspective, two studies found that actors generally had a lower f0 and slower speech rate when acting sarcastic voices than when they were acting sincere voices [16, 17]. Similarly, sincerity in a synthetic voice was associated with greater pitch range and faster articulation rate [18]; slow speech rate, by contrast, was associated with less competence [19]. Finally, higher f0 was associated with higher agreeableness, which also relates indirectly to trust [20].

We are still far from consensus regarding the direction of association between prosodic features – in particular, speech rate and f0 – and trustworthiness. Methodological differences may account, at least in part, for observed discrepancies between studies. Trustworthiness is a complex and dynamic social attribution, difficult to capture fully through explicit questionnaire rating, the predominant method used so far.

We present a study designed to calibrate listeners' dynamically responsive trust attributions to voice characteristics, with a game theory methodology that produces implicit measures of trustworthiness. While questionnaires might give rise to essentially attitudinal responses – e.g. based on social stereotypes linked to accents instead of voice characteristics *per se* – the investment game allows us to study the impact of voice characteristics on trustworthiness without participants realising that this is the focus of the experiment. To our knowledge, no other study has used such an implicit measure to assess the effect of prosody on trust attributions. Our method has the further advantage of taking into account the development of personality attributions with experience: the participants' learning about the other player's behaviour simulates (although in a somewhat contrived manner) the dynamics of personality judgements in real life situations.

2. Method

2.1. The investment game

We used an iterated trust game – "the investment game" – to test how voice-based trust attributions change with experience. Based in game theory [21], it involves making simulated monetary investments to other players, who decide whether to reciprocate the investment or not. The money invested, the dependent variable, can therefore be used as the implicit measure of trust. Experiments using trust games have been carried out to investigate trust attributions to a range of characteristics, including gender [22, 23], race, emotion and reputation [24, 25] and facial expressions [26]. This paper presents a follow-up experi-

 doi: 10.21437/SpeechProsody.2016-167

ment to [11], in which participants made monetary investments with virtual players associated with different behaviours and accents.

2.2. Participants

Participants were 83 native British English speakers (52 females, 32 males) aged 18-67 (median = 21, SD = 11). They were university undergraduate students who received course credit for participation or members of the public who received monetary compensation. Self-reports on participants' geographical origins were: southwest England (n = 44), southeast England (n = 20), Midlands (n = 7), Wales (n = 5), northwest England (n = 3), East Anglia (n = 2) and northeast England (n = 1) . One English-speaking subject from Canada was eliminated from the data-set, as we needed the participants to have had some contact with a wide range of British English accents.

2.3. Stimuli

The utterances used in the experiment were recorded in a sound-attenuated booth. Research is inconclusive regarding which gender elicits the highest trustworthiness judgement [27], and we eliminated this factor from our experimental design, recording twelve female native British English speakers in their twenties. Three speakers had a Plymouth accent, three a Birmigham accent, three a London accent and three spoke standard southern British English (SSBE). Each speaker read four blocks of 20 sentences (one for each round of the game), all approximately the same length (mean number of syllables per sentence 16.95, SD 1.08). Apart from the first utterance of each block, which served for the virtual player to introduce herself, all other utterances were about strategies to follow in the game, e.g.: "I'm going to return more money now, if you invest more as well"; "The goal of the game is to earn as much money as possible". A noise-removal filter was applied to the recorded utterances, which were then amplitude normalised.

2.4. Procedure

In the investment game, one player has a sum of virtual money which can be invested, in whole or in part, with another player in the hope of receiving an improved return. The amount of money invested is taken as an implicit measure of trust (e.g. [25, 26]). In our version of the game, the participant started with a notional sum of 8 at the beginning of each of the 20 rounds of the game. Each round began with a prompt utterance from the virtual player (see above), and then the participan decided whether to invest all, part, or none of it with the virtual player, who received three times the invested amount. The virtual player was programmed to have one of two behaviours, either returning 120% to 240% of the invested money to the participant (Generous condition) or 0% to 120% (Mean condition), for every round.

Each participant played the investment game four times – twice in the generous condition and twice in the mean condition - with a different virtual player each time. Each virtual player in these four games was associated with a different accent, from one of the 12 pre-recorded voices. The order of behaviours and accents was counterbalanced across participants. The 20 utterances were always in the same position within a block, with random association of utterances to block. Each participant therefore heard all 80 of the prompt sentences, read by four speakers with different accents, across the four game

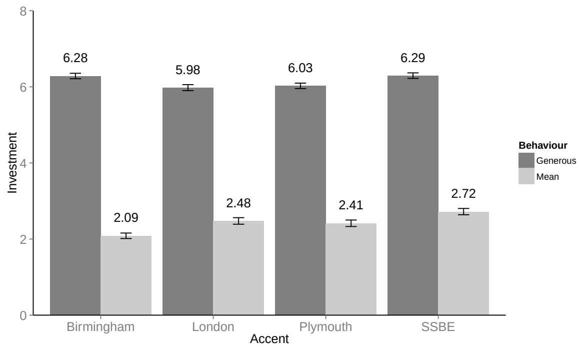

Figure 1: *Mean investments to the four target accents, in the two behaviour conditions*

blocks. Within an accent, speakers were randomly assigned to participants and behaviours in a counterbalanced fashion.

2.5. Prosodic measures

Segmentation and labelling of the individual sound files was done with the MAUS General Web service [28]. The transcriptions thus obtained were then used to extract prosodic measures in Praat [29] and MATLAB, specifically, mean pitch, pitch range, voice quality and articulation rate. Mean pitch was calculated as the mean f0 value for each vowel, then averaged across individual utterances. Pitch range, in order to eliminate potential outliers, was calculated as the difference between the 10th and the 90th percentiles of the mean f0 value for each vowel, as in [30], then averaged across individual utterances. Finally, we used H1-H2 – the difference between the first and second harmonic – as a measure of voice quality, as in [31, 32]. This was calculated using VoiceSauce [33].

3. Results

To determine the effects of game behaviour and vocal characteristics on investments, a mixed-effects regression linear model was fitted to the data using forward stepwise selection, selecting each successive predictor according to the lowest AIC (Akaike Information Criterion [34]). Investment was the dependent variable, behaviour, accent, game turn (i.e. ordinal numbers of the rounds within each game), pitch range, f0, articulation rate and H1-H2 were predictors, and subject was a random factor.

3.1. Effect of game behaviour

The virtual player's behaviour explained the biggest portion of the variance ($\chi^2(4)$ = 4269.1, $p < 0.001$). This was expected, since participants learn very quickly to invest higher amounts of money with Generous virtual players than with Mean virtual players. We also found a main effect of game turn ($\chi^2(5)$ = 96.02, $p < 0.001$) and a significant interaction between behaviour and game turn ($\chi^2(6)$ = 191.88, $p < 0.001$): as previously noted, investments increase in the Generous condition and decrease in the Mean condition as the game progresses.

3.2. Effect of accent

We found an effect of accent, $\chi^2(9)$ = 30.59, $p < 0.001$. Post-hoc pairwise comparisons show that average investments to SSBE speakers were higher than the investments to Plymouth (p = 0.005), London (p = 0.022) and Birmingham ($p < 0.002$)

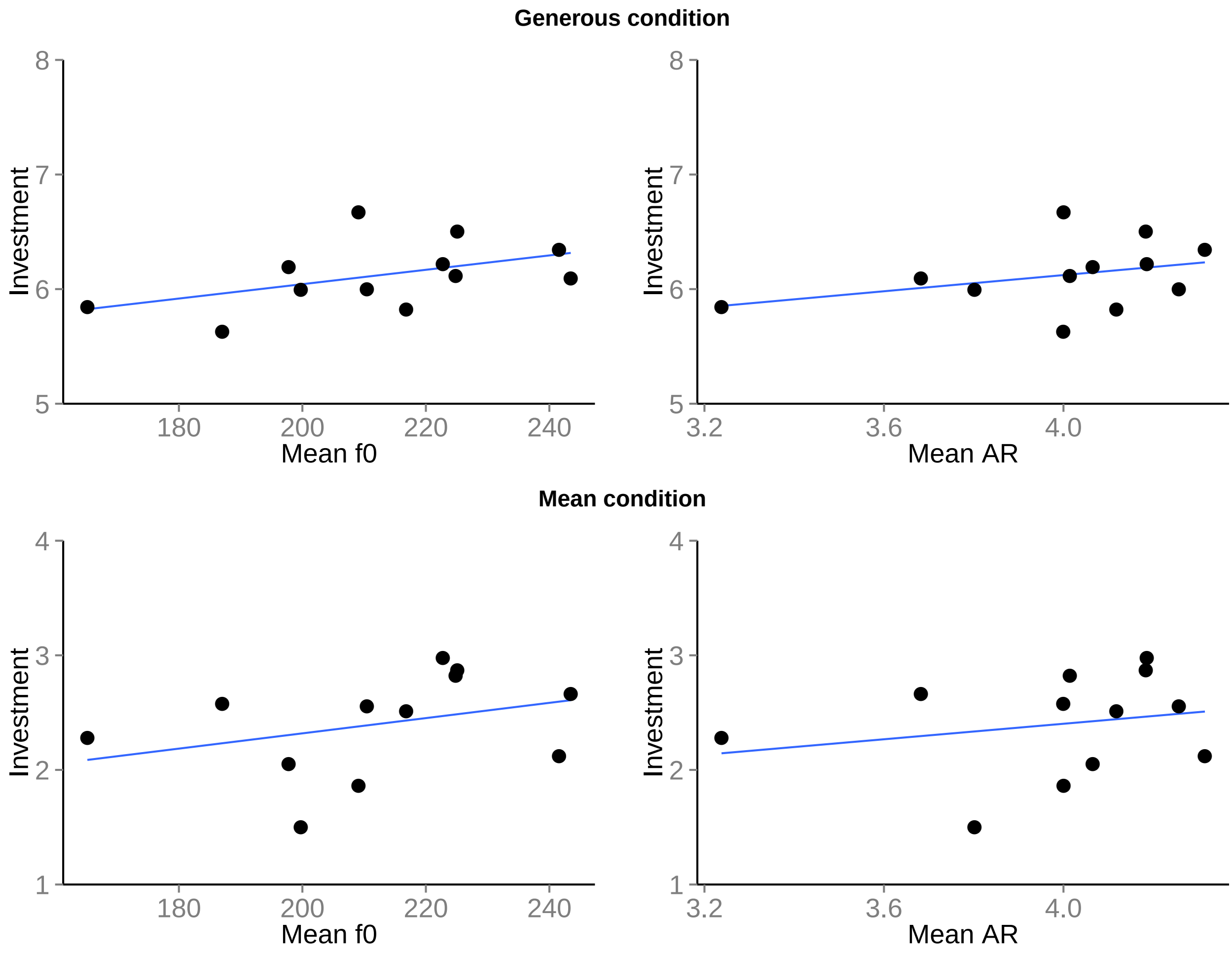

Figure 2: *Scatterplot of mean f0 and articulation rate (AR) against investment in the Generous behaviour condition (top) and Mean behaviour condition (bottom)*

speakers, with no other pairs showing significant differences. There was also an interaction between accent and behaviour, $\chi^2(12) = 13.63$, $p = 0.003$. Figure 1 shows investment according to accent and behaviour. Pairwise comparisons showed that the investments in the Generous condition were significantly higher with SSBE speakers than with speakers from Plymouth (p = 0.04) and London ($p = 0.014$), and that investments to Birmingham speakers were significantly higher than investments to London ($p = 0.017$) and Birmingham speakers ($p = 0.041$). By contrast, in the Mean condition, investments to SSBE, Plymouth and London speakers were all higher than the investments to Birmingham speakers ($p < 0.001$, $p = 0.017$, $p = 0.004$ respectively), and investments to SSBE speakers were also higher than investments to Plymouth speakers ($p = 0.025$).

3.3. Effect of prosody

The regression model showed an effect of f0 ($\chi^2(13) = 5.86$, p = 0.015), while articulation rate approached significance ($\chi^2(15)$ = 3.46, $p = 0.063$). Overall, as shown in Figure 2 , higher pitch and faster articulation rate were associated with higher investments. As we interpret the participants' monetary investments in the game as an implicit measure of trust, these results are largely consistent with previous findings (e.g. [5]). There were no effects of pitch range and H1-H2 on investments.

We found an interaction between f0 and game turn ($\chi^2(14)$ = 12.69, $p < 0.001$), with higher f0 more strongly associated with higher investment earlier in the game. There was an interaction between f0 and articulation rate ($\chi^2(16) = 5.38$, p = 0.02), with greater investment for high f0 at fast rate. There was an interaction between articulation rate and behaviour ($\chi^2(17)$ = 4.58, $p = 0.032$), with the effect of faster rate on boosting investment being somewhat greater in the Mean condition (Figure 2). We discuss the interpretation of these effects of prosody on trustworthiness below.

4. General Discussion

As previously found, the investment game allows us to assess the effect of voices on implicit trust judgements. Standard southern British English attracted higher overall investment than the other British English accents, but – in accordance with our earlier work [11] – there were differential effects of virtual player accents according to their behaviour. In particular, the Birmingham accent attracted relatively high investment in the Generous condition, but the lowest investment in the Mean condition. We are still exploring the underlying reasons for this contrast, which may relate to social stereotypes, individual voice quality or a combination of the two.

Two prosodic features previously associated with trust – high pitch and fast articulation rate – influenced our participants' monetary investments. This is consistent with some, but not all previous studies on attributions of trustworthiness and associated characteristics [5, 14, 16, 17, 18, 19, 20], although the methodology we employed to establish an implicit measure of trust is radically different.

The greater trustworthiness of voices with higher pitch may be related to the "Size/Frequency Code" theory [35], based on higher f0 being generally associated with a smaller larynx and therefore smaller body size. As a consequence, we tend to associate lower f0 with dominance and assertiveness, and higher f0 with friendliness and cooperativeness [36, 37]. In a type of interaction where trust is required, listeners might be more inclined to attribute trustworthiness to a speaker who is perceived as friendly rather than dominating. We note that this interpretation can only be applied to female speakers on the basis of our methodology; it remains to be seen if higher f0 also accords with trustworthiness in male speakers. Furthermore, the relationship between f0 and trust attributions is manifestly not likely to be linear through the range of possible values, partially evidenced in our study by how f0 effects on investment interact with game turn and articulation rate.

The "Effort Code" account [38] may help to explain the association we found between faster articulation rate and higher trustworthiness attributions. Speaking faster can be seen as a corollary of greater conversational effort and hence more engagement in the interaction; manifestly, perceived engagement is likely be seen as a sign of trustworthiness. As with f0, there must be limits to any trust-boosting effects of rate however, and we found differential effect of rate according to whether the virtual player was behaving generously or meanly. It may be considered whether a faster rate impacts on participants' sense of response urgency and therefore on investment decisions. Previous studies have shown that, under conditions of time pressure, a faster rate in spoken prompts may encourage quicker responding [39, 40], although such pressure was not explicitly present in our context. In addition, a fast speech rate of a synthesized voice did not influence the persuasiveness of a commercial message [41]. Naturally, given the complex interactions that emerged among speech rate, accent and our other variables of interest, future studies tailored to unravel them are needed.

There are likely to be voice characteristics contributing to trustworthiness attributions that have not been considered in this study. Furthermore, the effect of the listener's personality and their context-driven expectations should also be examined in a comprehensive account. However, our key findings with respect to prosody concur with most of the foregoing literature: higher f0 and faster rate are associated with higher speaker trustworthiness. Given that we have used an implicit measure with no explicit reference to speaker voice in the experimental procedure, we can be confident that the findings reflect genuine paralinguistic effects rather than stereotype-based attributions.

5. Acknowledgements

This work was supported by CogNovo (FP7-PEOPLE-2013-ITN-604764), a project funded by the EU Marie Curie programme.

6. References

[1] J. S. Uleman, S. A. Saribay, and C. M. Gonzalez, "Spontaneous inferences, implicit impressions, and implicit theories," *Annual Review of Psychology*, vol. 59, pp. 329–360, 2008.

[2] C. Nass, Y. Moon, B. Fogg, B. Reeves, and D. C. Dryer, "Can computer personalities be human personalities?" *International Journal of Human-Computer Studies*, vol. 43, no. 2, pp. 223–239, 1995.

[3] J. Willis and A. Todorov, "First impressions: making up your mind after a 100-ms exposure to a face." *Psychological Science*, vol. 17, no. 7, pp. 592–598, 2006.

[4] P. McAleer, A. Todorov, and P. Belin, "How do you say Hello? Personality impressions from brief novel voices." *PLoS ONE 9(3): e90779*, 2014.

[5] A. C. Elkins and D. C. Derrick, "The sound of trust: voice as a measurement of trust during interactions with embodied conversational agents." *Group Decision and Negotiation*, vol. 22, no. 5, pp. 897–913, 2013.

[6] W. Apple, L. A. Streeter, and R. M. Krauss, "Effects of Pitch and Speech Rate on Personal Attributions," *Journal of Personality and Social Psychology*, vol. 37, no. 5, pp. 715–727, 1979.

[7] V. C. Tartter and D. Braun, "Hearing smiles and frowns in normal and whisper registers," *Journal of the Acoustical Society of America*, vol. 96, no. 4, pp. 2101–2107, 1994.

[8] H. Bishop, N. Coupland, and P. Garrett, "Conceptual accent evaluation: Thirty years of accent prejudice in the UK." *Acta Linguistica Hafniensia: International Journal of Linguistics*, vol. 37, no. 1, pp. 131–154, 2005.

[9] J. A. Dixon, B. Mahoney, and R. Cocks, "Accents of guilt? effects of regional accent, race, and crime type on attributions of guilt," *Journal of Language and Social Psychology*, vol. 21, no. 2, pp. 162–168, 2002.

[10] J. N. Fuertes, W. H. Gottdiener, H. Martin, T. C. Gilbert, and H. Giles, "A meta-analysis of the effects of speakers accents on interpersonal evaluations," *European Journal of Social Psychology*, vol. 42, pp. 120–133, 2012.

[11] I. Torre, J. Goslin, and L. White, "Investing in accents: How does experience mediate trust attributions to different voices?" in *Proceedings of the 18th International Congress of Phonetic Sciences*, Glasgow, U. K., 10-14 August 2015.

[12] L. A. Pervin, *Personality: Theory and Research*. New York: Wiley, 2001.

[13] L. Anolli and R. Ciceri, "The voice of deception: vocal strategies of naive and able liars," *Journal of Nonverbal Behavior*, vol. 21, no. 4, pp. 259–284, 1997.

[14] M. Zuckerman, R. S. DeFrank, J. A. Hall, D. T. Larrance, and R. Rosenthal, "Facial and vocal cues of deception and honesty," *Journal of Experimental Social Psychology*, vol. 15, no. 4, pp. 378–396, 1979.

[15] C. Kirchhbel and D. M. Howard, "Detecting suspicious behaviour using speech: Acoustic correlates of deceptive speech - an exploratory investigation," *Applied Ergonomics*, vol. 44, pp. 694–702, 2013.

[16] H. S. Cheang and M. D. Pell, "The sound of sarcasm," *Speech Communication*, vol. 50, pp. 366–381, 2008.

[17] R. Rao, "Prosodic consequences of sarcasm versus sincerity in Mexican Spanish," *Concentric: Studies in Linguistics*, vol. 39, no. 2, pp. 33–59, November 2013.

[18] J. Trouvain, S. Schmidt, M. Schrder, M. Schmitz, and W. J. Barry, "Modelling personality features by changing prosody in synthetic speech," in *Proceedings of the third International Conference on Speech Prosody*, Dresden, Germany, 2-5 May 2006.

[19] B. Smith, B. Brown, W. Strong, and A. Rencher, "Effects of speech rate on personality perception," *Language and Speech*, vol. 18, pp. 145–152, 1975.

[20] M. Imhof, “Listening to voices and judging people,” *The International Journal of Listening*, vol. 24, pp. 19–33, 2010.

[21] W. Poundstone, *Prisoner’s Dilemma*. New York: Doubleday, 1992.

[22] A. Boenin and D. Serra, “Gender pairing bias in trustworthiness,” *The Journal of Socio-Economics*, vol. 38, pp. 779–789, 2009.

[23] A. Chaudhuri, T. Paichayontvijit, and L. Shen, “Gender differences in trust and trustworthiness: Individuals, single sex and mixed sex groups.” *Journal of Economic Psychology*, vol. 34, pp. 181–194, 2013.

[24] M. I. Tortosa, J. Lupiez, and M. Ruz, “Race, emotion and trust: An ERP study.” *Brain Research*, vol. 1494, pp. 44–55, 2013.

[25] D. A. Stanley, P. Sokol-Hessner, D. S. Fareri, M. T. Perino, M. R. Delgado, M. R. Banaji, and E. A. Phelps, “Race and reputation: perceived racial group trustworthiness influences the neural correlates of trust decisions,” *Philosophical Transactions of the Royal Society B*, vol. 367, pp. 744–753, 2012.

[26] E. Krumhuber, A. S. R. Manstead, D. Cosker, D. Marshall, P. L. Rosin, and A. Kappas, “Facial dynamics as indicators of trustworthiness and cooperative behavior,” *Emotion*, vol. 7, no. 4, pp. 730–735, 2007.

[27] C. Nass and S. Brave, *Wired for speech: How voice activates and advances the human-computer relationship.* Cambridge, MA: MIT Press, 2005.

[28] F. Schiel, “Automatic phonetic transcription of non-prompted speech,” in *Proceedings of the 14th International Congress of Phonetic Sciences*, San Francisco, United States, 1-7 August 1999, pp. 607–610.

[29] P. Boersma and D. Weenink, “Praat: doing phonetics by computer,” 2010.

[30] D. Patterson and D. R. Ladd, “Pitch Range Modelling: Linguistic Dimensions of Variation,” in *Proceedings of the 14th International Congress of Phonetic Sciences*, San Francisco, United States, 1-7 August 1999, pp. 1169–1172.

[31] K. Johnson, *Acoustic and Auditory Phonetics*, 2nd ed. Oxford: Blackwell, 2002.

[32] M. Garellek, P. Keating, C. M. Esposito, and J. Kreiman, “Voice quality and tone identification in White Hmong,” *Journal of the Acoustical Society of America*, vol. 133, pp. 1078–1089, 2013.

[33] Y.-L. Shue, P. Keating, C. Vicenik, and K. Yu, “Voicesauce: A Program for Voice Analysis,” in *Proceedings of the 17th International Congress of Phonetic Sciences*, Hong Kong, China, 17-21 August 2011, pp. 1846–1849.

[34] K. Aho, D. Derryberry, and T. Peterson, “Model selection for ecologists: the worldviews of AIC and BIC,” *Ecology*, vol. 95, no. 3, pp. 631–636, March 2014, paper reference for AIC selection.

[35] J. J. Ohala, “Cross-language use of pitch: An ethological view,” *Phonetica*, vol. 40, pp. 1–18, 1983.

[36] ——, “The acoustic origin of the smile,” *Journal of the Acoustical Society of America*, vol. 68, 1980.

[37] J. Hirschberg, “The Pragmatics of Intonational Meaning,” in *Proceedings of the first International Conference on Speech Prosody*, Aix-en-Provence, France, 11-13 April 2002.

[38] C. Gussenhoven, “Intonation and interpretation: phonetics and phonology,” in *Proceedings of the first International Conference on Speech Prosody*, Aix-en-Provence, France, 11-13 April 2002.

[39] H. Shintel and H. C. Nusbaum, “The sound of motion in spoken language: Visual information conveyed by acoustic properties of speech,” *Cognition*, vol. 105, no. 3, pp. 681–690, 2007.

[40] ——, “Moving to the speed of sound: Context modulation of the effect of acoustic properties of speech,” *Cognitive science*, vol. 32, no. 6, pp. 1063–1074, 2008.

[41] C. Jones, L. Berry, and C. Stevens, “Synthesized speech intelligibility and persuasion: Speech rate and non-native listeners,” *Computer Speech & Language*, vol. 21, no. 4, pp. 641–651, 2007.