Eye Gaze Affects Vocal Intonation Mimicry

Marie Postma-Nilsenová (M.Nilsenova@tilburguniversity.edu)
Tilburg center for Cognition and Communication (TiCC), Warandelaan 2, 5037 AB Tilburg
Tilburg, The Netherlands

Niek Brunninkhuis (n.brunninkhuis@gmail.com)
Tilburg center for Cognition and Communication (TiCC), Warandelaan 2, 5037 AB Tilburg
Tilburg, The Netherlands

Eric Postma (E.O.Postma@tilburguniversity.edu)
Tilburg center for Cognition and Communication (TiCC), Warandelaan 2, 5037 AB Tilburg
Tilburg, The Netherlands

Abstract

Eye gaze and behavioral mimicry are important foundations of social interaction. Inspired by recent studies on eye-gaze mediated spontaneous behavioral mimicry of gestures, we studied the effect of eye gaze direction on vocal mimicry. Participants were instructed to repeat digits spoken by a virtual agent with a direct or averted eye gaze. As a measure of imitation, the vocal pitch was recorded and analyzed in order to determine if and to what extent vocal mimicry was modulated by eye gaze. The results showed that eye gaze direction affects vocal mimicry as measured by pitch slope. That is, when participants were exposed to an agent that gazed at them directly, they accommodated their intonation more to that of the agent, than when they were exposed to an agent that averted its gaze. These results suggest that in social interaction with a virtual agent, humans mimic vocal intonation and that the degree of mimicry depends on the eye-gaze direction of the agent. The implications for studies of social interaction are discussed.

Keywords: Eye Gaze; Vocal Mimicry; Virtual Agent

1. Introduction

In social interaction, eye gaze direction and behavioral mimicry are powerful nonverbal social signals (Stass & Willis Jr, 1967; Kendon, 1967; Scherer, 1974; Cook & Smith, 1975; Fukayama, Ohno, Mukawa, Sawaki, & Hagita, 2002; Mason, Tatkow, & Macrae, 2005; Baaren, Janssen, Chartrand, & Dijksterhuis, 2009; Wang, 2012). Among different types of nonverbal cues, vocal pitch mimicry appears to play a fundamental role. The results of a range of experimental studies suggest that speakers effortlessly imitate and converge to the phonetic properties of recently heard speech (Delvaux & Soquet, 2007; Gentilucci & Bernardis, 2007; Natale, 1975; Pardo, 2006; Shockley, Sabadini, & Fowler, 2004), including pitch (Babel & Bulatov, 2012; Goldinger, 1998; Gorisch, Wells, & Brown, 2012). Pitch – the perceptual correlate of fundamental frequency (F_0) – is, arguably, the most important vocal source of information regarding emotions, stands and attitudes of the speaker. The F_0 region thus provides acoustic information for imitation exploited in promoting social convergence and status accommodation (Gregory, 1983; Gregory, Webster, & Huang, 1993; Gregory & Webster, 1996; Gregory, Dagan, & Webster, 1997; Haas & Gregory, 2005; Pardo, 2006) and expressing ingroup-outgroup bias (Babel, 2009; Pardo, Gibbons, Suppes, & Krauss, 2012).

According to the Communication Accommodation Theory (CAT), in social interaction, people adjust their vocal characteristics to accommodate to each other (Giles, Coupland, & Coupland, 1991). Support for CAT came from a study by (Gregory & Webster, 1996), who analyzed *Larry King Live* television interviews. The results revealed that depending on the relative status of the interviewed guest, Larry King mimicked the vocal characteristics of his guests (in case of high status guests) or the guests mimicked the vocal characteristics of Larry King (in case of low status guests). In general, speakers who are perceived as attractive, likable and/or dominant influence listeners' pitch output, and pitch convergence can be seen as an indicator of cooperative behavior in communication dyads (Okada, Lachs, & Boone, 2012). Pitch divergence, on the other hand, suggests that speakers may wish to be viewed as dissimilar and increase social distance between themselves (Babel, 2009).

Interestingly, empirical studies have shown that in social interactions, the direction of eye gaze influences the degree of behavioral mimicry (Kleinke & Pohlen, 1971; Chartrand & Bargh, 1999). A striking demonstration of the direct link between eye gaze direction and behavioral mimicry is due to Wang, Newport, and Hamilton (2011). In their study, participants were presented with a movie of an actress that either looked directly at the camera or averted her gaze from the camera. In both conditions there were movies of the actress opening her hand and movies in which she closed her hand. At the beginning of each trial, participants were instructed to either open or close their hand. The instructed hand movements could be congruent or incongruent with the displayed hand movements of the actress. After receiving the instruction, participants had to make the hand movement as quickly as possible. They were *not* instructed to mimic the hand movement of the actress. Not surprisingly, participants were significantly faster in making congruent hand movements than in making incongruent ones. Interestingly, though, the congruent hand movements in the direct gaze condition were considerably faster than those in the averted gaze condition, whereas for the incongruent hand movements gaze direction had no effect. These findings reveal that eye contact has a quick and specific effect on action mimicry.

Inspired by these results, we expect eye-gaze modulated mimicry in other response modalities, such as the vocal modality. Although vocal mimicry is a well-known phenomenon, the modulating effect of eye gaze has not yet been explored experimentally. The aim of this study is to investigate if vocal mimicry is modulated by eye gaze direction.

Instead of exploring the effect of eye gaze on reaction times, we determined its effect on the degree of vocal mimicry. We employed an experimental setting in which a virtual agent with either an averted or direct gaze utters single words with one of three pitch contours. The participants were instructed to repeat the words, but were *not* instructed to mimic the pitch contours (instead, they were distracted with another task).

2. Experiment

Participants and Design

Forty-seven Dutch native speakers (24 male; mean age 21;4) were recruited from the Tilburg University student population. The experiment had a one-way within-subjects design with Eye Gaze Direction (direct, averted) as the independent variable and Vocal Mimicry as the dependent variable.

Material

The stimulus material consisted of 8 visually presented words during which we measured speakers' baseline pitch, followed by 48 videos (16 experimental trials + 32 distractors). Eight of the 16 experimental trials involved direct gaze of the agent (top figure 1), during the other 8, the agent either averted his gaze left or right (4 times each). In half of the distractor movies and only in these movies, the agent blinked his eyes; the ratio of gaze directions was 8:4:4 for *direct:left:right* in the blinking distractor group as well as in the no-blinking distractor group. In every stimulus, the agent expressed a single word (all Dutch monosyllabic digits between 0-10, in order to prevent possible emotional associations to the stimuli that might affect the speaker's pitch), followed by a blank screen.

The virtual agent was created with Poser (Smitch Micro Software inc, Aliso Viejo, California, U.S.), see Figure 1. The agent's lip movements were matched to the pre-recorded words and subtle head and eye movements were added to enhance the realism and to prevent the agent from being perceived as threatening (Ellsworth, 1975; Cook & Smith, 1975; Argyle, Lefebvre, & Cook, 1974). A film strip of a sample movie is shown in Figure 2.

The sound files used for the agent's voice were pre-recorded in a sound attenuated booth by the second (male) author. His pitch values represented average vocal values for an adult male speaker in the Netherlands (i.e., in the 70–250 Hz range). Three different intonations were used in these recordings (a falling, a rising, and a late-rising tune, see Figure 3). Each movie lasted approximately 5 s, including a 0.5 s fade-from-black and fade-to-black to (i) smooth the transition between consecutive movies, (ii) mark the beginning and end of each stimulus, and (iii) avoid an unnatural and potentially threatening gaze duration.

Procedure

The experiment was set up in E-Prime (Psychological Software Tools Inc, Pittsburgh, Pennsylvania, USA) and presented with the help of a Dell Latitude E5510 laptop and a Trust HS-2100 headset. The distance between the participant's mouth and the headset's microphone was kept constant. For the baseline recordings, participants were presented with a random sequence of eight consecutive digits displayed

Figure 1: Impression of the virtual agent used in the experiment. From top to bottom: direct gaze, right averted gaze (while blinking in a distractor movie), and left averted gaze (while speaking).

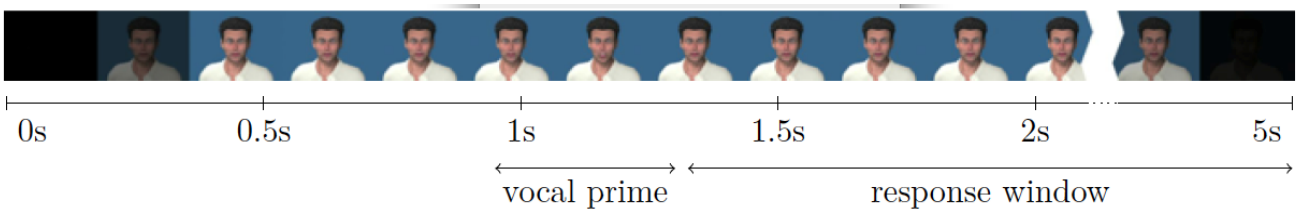

Figure 2: Film strip of a sample movie as used in the experiment.

in white against a blue background. They were instructed to read each word out loud. Subsequently, they were shown a randomized sequence of videos in which the agent pronounced a digit and were instructed to repeat it. In order to ensure that the participants fixated the agent's eye region and did not focus on the imitation task, they were given the additional instruction to press the space bar whenever the agent blinked. The full length of the experiment was 10 minutes on average.

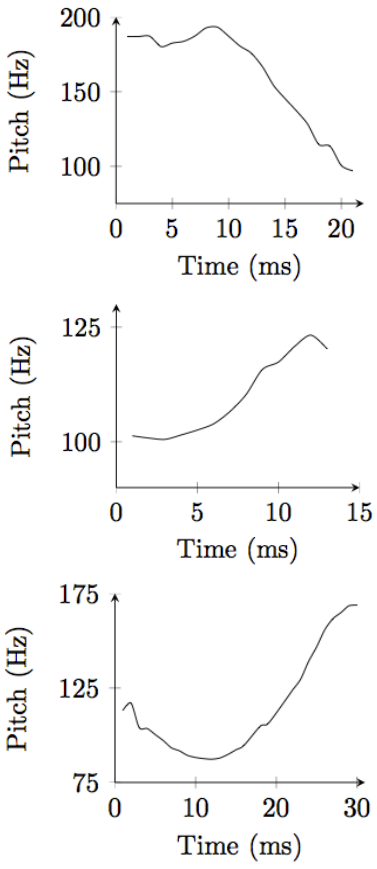

Figure 3: Graphical representations of the three different intonation patterns of the spoken digits as pronounced by the EIA (second experimenter's voice). From top to bottom: falling intonation, rising intonation, and late-rising intonation (in Hz).

Measurements

The experimental audio files collected during the experiment (1128 in total) were manually preprocessed to remove unvoiced speech and silent segments. After establishing the appropriate pitch threshold and ceiling for each individual voice, by auditory and visual inspection of the audio file and spectrogram, respectively, we extracted the pitch contour using the standard autocorrelation-based pitch detection function of *Praat* (Boersma, 2001). The frequency values were converted to semitones to allow for a comparison of male and female speakers (Borden & Harris, 1980). For each audio file, we determined the values of two measures of pitch contour: pitch slope (P_{deriv}) and pitch regression ($P_{regline}$). The pitch slope is defined as the average difference of adjacent frequencies in the pitch contour. The pitch regression is defined as the slope of the linear regression line through the points making up the pitch contour.

Our measure of vocal mimicry is based on two variables. The first variable $\Delta_{P_{baseline}}$ is the absolute difference between the agent's pitch, P_{agent} and the participant's baseline pitch, $P_{baseline}$, i.e., the pitch of the participant before repeating the agent.

$$\Delta_{P_{baseline}} = |P_{agent} - P_{baseline}| \quad (1)$$

The second variable $\Delta_{P_{rep}}$ is the absolute difference between the agent's pitch and the participant's pitch while repeating the agent, P_{rep}.

$$\Delta_{P_{rep}} = |P_{agent} - P_{rep}| \quad (2)$$

By subtracting the values of $\Delta_{P_{baseline}}$ and $\Delta_{P_{rep}}$, we obtain our measure of vocal mimicry M_V.

$$M_V = \Delta_{P_{baseline}} - \Delta_{P_{rep}} \quad (3)$$

A positive value of the vocal mimicry M_V indicates vocal mimicry, whereas a negative value indicates vocal complementarity. The value of vocal mimicry was calculated both for the pitch slope and the pitch regression separately.

3. Results

We start by reporting the vocal mimicry results regardless of eye gaze direction. Subsequently, we report the gaze-dependent vocal mimicry results. Non-parametric tests were used for data that were not normally distributed.

Vocal Mimicry Results Independent of Gaze Direction

A Wilcoxon Signed-rank test was performed to establish if vocal mimicry occurred. Effect size estimates were computed using $r\ (= |Z/\sqrt{N}|)$, where N equals the number of samples. The results for P_{deriv} and $P_{regline}$ indicated a significant difference between the baseline and the experimental trial measurements, with a shift in the direction of the agent's pitch (see Table 1). Figures 4 and 5 illustrate the results: Both figures show the median values of the absolute differences between the agent's and the participant's pitch in the baseline and the repetition trial, with the results for males and females plotted separately.

Table 1: Results of Wilcoxon Signed-rank test for mimicry regardless of eye gaze direction (pitch measurements reported in semitones).

	$\Delta_{P_{baseline}}$	$\Delta_{P_{rep}}$	Z	r	p
P_{deriv}					
Females	41.473	37.926	-2.829	0.417	0.005
Males	41.012	38.252	-2.543	0.367	0.011
Total	41.813	37.926	-3.852	0.397	<0.001
$P_{regline}$					
Females	0.837	0.301	-3.041	0.448	0.002
Males	0.847	0.255	-3.114	0.449	0.002
Total	0.841	0.668	-4.360	0.450	<0.001

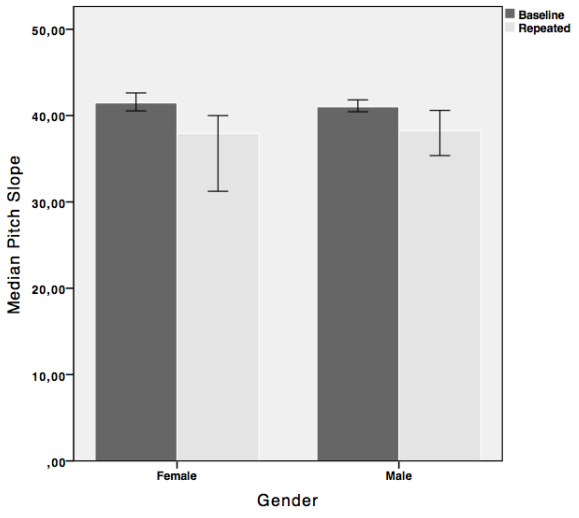

Figure 4: Plot of the results obtained for P_{deriv} absolute differences with error bars (95% CI).

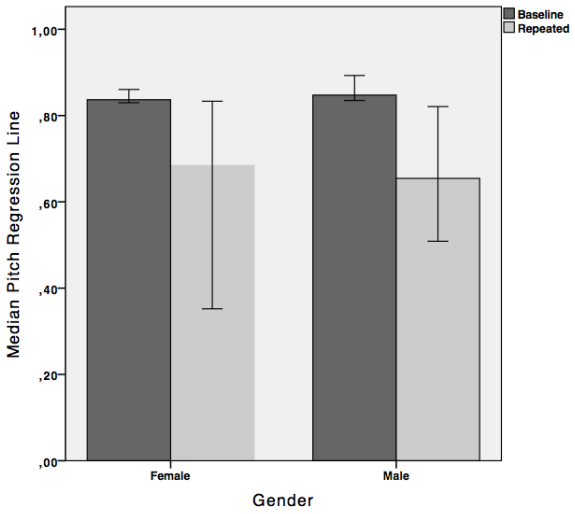

Figure 5: Plot of the results obtained for $P_{regline}$ absolute differences with error bars (95% CI).

These results indicate that participants accommodated the slope and regression line of their pitch contours to that of the agent. On both measures, when the dataset was split by gender, male and female participants showed similar effects.

Vocal Mimicry Results Dependent on Gaze Direction

The results of the Wilcoxon Signed-rank test (Table 2) show a significant effect of gaze: compared to the participant's baseline, the slope of the pitch contour in the participant's repetition is more similar to that of the agent gazing towards the participant, than when the agent averted its gaze. A split-file analysis by gender showed a significant effect for male participants only. The medians of P_{deriv} are visually presented in Figure 6, indicating cases of divergence in the condition with averted gaze in the male participant group.

Table 2: Results of Wilcoxon Signed-rank test for mimicry as measured by P_{deriv} depending on eye gaze direction.

	Median M_V Direct	Median M_V Averted	Z	r	p
P_{deriv}					
Females	6.339	2.203	-1.612	0.238	0.107
Males	5.013	2.394	-1.971	0.285	0.049
Total	5.754	2.203	-2.529	0.261	0.011

The results of a mixed within-between analysis of variants as measured by $P_{regline}$ are listed in Table 3. The results indicate no significant main effect of either gaze or gender, as well as no interaction effect of the two variables.

To explore possible individual variations in mimicry (taking pitch slope as the representative measure), we computed

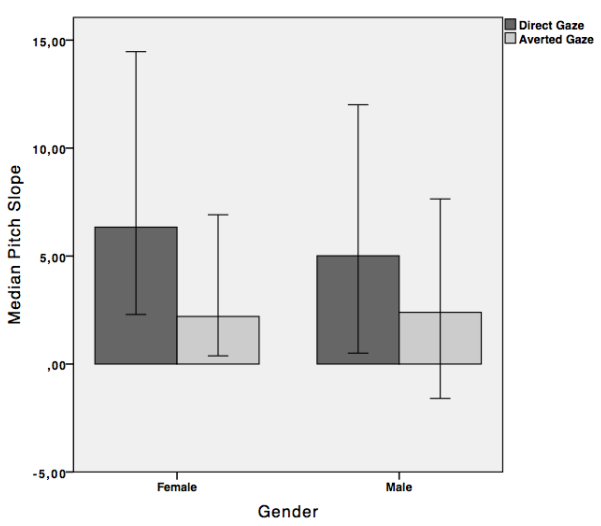

Figure 6: The median values of P_{deriv} absolute differences between the agent's and the participant's pitch for direct and averted gaze including error bars (95% CI).

Table 3: Results of ANOVA test for mimicry as measured by $P_{regline}$ depending on eye gaze direction.

	df	F	p	η_p^2
$P_{regline}$				
Gender	1	0.083	0.775	0.002
Gaze	1	1.772	0.190	0.038
Gender × Gaze	1	0.386	0.537	0.009

for each participant the average difference between M_V in the averted gaze condition and M_V in the direct gaze condition. The resulting scores reflect the individual effect of gaze on vocal mimicry; positive scores are associated with vocal mimicry, negative scores with vocal complementarity. Figure 7 is a graphical display of the scores, sorted from smaller to larger scores. Each bar represents a participant, and the height of the bar represents the magnitude of mimicry (positive) or complementarity (negative).

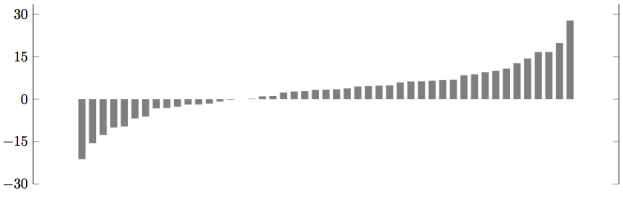

Figure 7: Individual differences in vocal mimicry

To sum up, we found that participants exhibit verbal mimicry as measured by P_{deriv} and $P_{regline}$ and that at least in the case of P_{deriv}, the mimicry effect is stronger when the agent gazes at the participant. In addition, we observed individual variations in the degree of vocal mimicry. The implications of our findings are discussed in the next section.

4. General Discussion and Conclusion

Pitch is arguably the most important source of information regarding emotions, stances and attitudes of the speaker (Juslin & Laukka, 2003) and pitch mimicry plays an important role in human interaction in that it reflects the closeness of the social bond between two individuals. Our findings indicate that pitch mimicry in social interactions may be modulated by eye gaze. The results of our experiment extend and generalize the findings obtained by Wang et al. (2011) for the visual (gesture) modality. The existence of eye-gaze modulated vocal mimicry underscores the importance of eye gaze as a social signal and lends further support to the close relation of eye contact and behavioral mimicry in social interaction.

The potential impact of eye gaze on the social bond between virtual agents and humans is of relevance to the development of future human-computer interfaces that display an interactive embodied agent and sense vocal and visual cues of the human interacting with the agent. Software controlling an interactive embodied agent, may confirm the establishment of a social bond with the human by instructing the agent to eye gaze and vocally address the human and subsequently sense the concomitant vocal mimicry.

In our experiment, participants were instructed to repeat the digit pronounced by the agent. It is not clear to what extent the mimicry observed depends on the type of instruction. Future work may experiment with alternative types of responses. For instance, participants could be instructed to complete a partial sentence uttered by the agent or to respond to a statement. In this way the dependency between an explicit instruction to repeat an utterance and behavioral mimicry can be determined.

Our use of a male human voice and a male virtual agent, may have caused gender effects that can be further explored in future studies. According to the CAT (Giles et al., 1991), talkers modify their speech to reinforce valued and socially meaningful differences between themselves and their interaction partners. Since male voices are lower pitched than female voices (Sachs, Lieberman, & Erickson, 1973), females possibly reinforced the gender difference between themselves and the agent by deviating from the agents relatively low pitched voice and produce a higher pitched voice. Another issue to be explored in the future concerns the effect of joint attention. As well known, interlocutors are likely to follow each others gaze direction. It remains to be seen if contexts eliciting joint attention support vocal and other types of behavioral mimicry.

References

Argyle, M., Lefebvre, L., & Cook, M. (1974). The meaning of five patterns of gaze. *European Journal of Social Psychology*, *4*(2), 125–136.

Baaren, R., van, Janssen, L., Chartrand, T., & Dijksterhuis, A. (2009). Where is the love? The social aspects of

mimicry. *Philosophical Transactions of The Royal Society B*, *364*(1528), 2381–2389.

Babel, M. (2009). *Phonetic and social selectivity in speech accommodation*. Unpublished doctoral dissertation, University of California.

Babel, M., & Bulatov, D. (2012). The role of fundamental frequency in phonetic accommodation. *Language and Speech*, *55*(2), 231–248.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International*, *5*(9/10), 341–345.

Borden, G., & Harris, K. (1980). *Speech science primer: Physiology, acoustics and perception of speech*. Baltimore, Maryland, USA: Williams & Wilkins.

Chartrand, T., & Bargh, J. (1999). The chameleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology*, *76*, 893–910.

Cook, M., & Smith, M. (1975). The role of gaze in impression formation. *British Journal of Social and Clinical Psychology*, *14*, 19–25.

Delvaux, V., & Soquet, A. (2007). The influence of ambient speech on adult speech productions through unintentional imitation. *Phonetica*, *64*, 145–173.

Ellsworth, P. (1975). Direct gaze as a social stimulus: The example of aggression. In L. Krames, T. Alloway, & P. Pliner (Eds.), *Nonverbal communication of aggression.* New York, New York, USA: Plenum Press.

Fukayama, A., Ohno, T., Mukawa, N., Sawaki, M., & Hagita, N. (2002). Messages embedded in gaze of interface agents — impression management with agent's gaze. In *Proceedings of the sigchi conference on human factors in computing systems: Changing our world, changing ourselves* (pp. 41–48). New York, New York, USA: ACM.

Gentilucci, M., & Bernardis, P. (2007). Imitation during phoneme production. *Neuropsychologia*, *45*, 608–615.

Giles, H., Coupland, J., & Coupland, N. (1991). Accommodation theory: Communication, context, and consequence. In H. Giles, J. Coupland, & N. Coupland (Eds.), *Contexts of accommodation. developments in applied sociolinguistics* (pp. 1–68). Cambridge University Press.

Goldinger, S. D. (1998). Echoes of echoes? an episodic theory of lexical access. *Psychological Review*, *105*, 251–279.

Gorisch, J., Wells, B., & Brown, G. (2012). Pitch contour matching and interactional alignment across turns: an acoustic investigation. *Language and Speech*, *55*, 57–76.

Gregory, S. (1983). A quantitative analysis of temporal symmetry in microsocial relations. *American Sociological Review*, 129–135.

Gregory, S., Dagan, K., & Webster, S. (1997). Evaluating the relation of vocal accommodation in conversation partners' fundamental frequencies to perceptions of communication quality. *Journal of Nonverbal Behavior*, *21*(1), 23–43.

Gregory, S., & Webster, S. (1996). A nonverbal signal in voices of interview partners effectively predicts communication accommodation and social status perceptions. *Journal of Personality and Social Psychology*, *70*, 1231–1240.

Gregory, S., Webster, S., & Huang, C. (1993). Voice pitch and amplitude convergence as a metric of quality in dyadic interviews. *Language & Communication*.

Haas, A., & Gregory, S. (2005). The impact of physical attractiveness on womens social status and interactional power. *Sociological Forum*, *20*, 449–471.

Juslin, P., & Laukka, P. (2003). Communication of emotions in vocal expression and musical performance: different channels, same code? *Journal of the Acoustical Society of America*, *129*, 770-814.

Kendon, A. (1967). Some Functions of Gaze-direction in Social Interaction. *Acta Psychologica*, *26*, 22–63.

Kleinke, C., & Pohlen, P. (1971). Affective and emotional responses as a function of other person's gaze and cooperativeness in a two-person game. *Journal of Personality and Social Psychology*, *17*(3), 308–313.

Mason, M. F., Tatkow, E., & Macrae, C. (2005). The look of love: gaze shifts and person perception. *Psychological Science*, *16*, 236–239.

Natale, M. (1975). Social desirability as related to convergence of temporal speech patterns. *Perceptual and Motor Skills*, *40*(3), 827–830.

Okada, B., Lachs, L., & Boone, B. (2012). Interpreting tone of voice: Musical pitch relationships convey agreement in dyadic conversation. *Journal of Acoustical Society of America*, *132*(3), EL208–EL214.

Pardo, J. (2006). On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America*, *119*, 2382-2393.

Pardo, J., Gibbons, R., Suppes, A., & Krauss, R. (2012). Phonetic convergence in college roommates. *Journal of Phonetics*, *40*, 190–197.

Sachs, J., Lieberman, P., & Erickson, D. (1973). Anatomical and cultural determinants of male and female speech. In R. Shuy & R. Fasold (Eds.), *Language attitudes: Current trends and prospects* (pp. 74–84). Washington DC, USA: Georgetown University Press.

Scherer, S. (1974). Influence of proximity and eye contact on impression formation. *Perceptualand Motor Skills*, *38*, 538.

Shockley, K., Sabadini, L., & Fowler, C. (2004). Imitation in shadowing words. *Perception & Psychophysics*, *66*, 422–429.

Stass, J., & Willis Jr, F. (1967). Eye-contact, pupil dilation,and personal preference. *Psychonomic Science*, *7*, 375–376.

Wang, Y. (2012). *The control of mimicry by social signals*. Unpublished doctoral dissertation, University of Nottingham.

Wang, Y., Newport, R., & Hamilton, A. (2011). Eye contact enhances mimicry of intransitive hand movements. *Biology Letters*, *7*(1), 7–10.