Speaker Attitude and Sexual Orientation Affect Phonetic Imitation

Alan Yu
University of Chicago

Carissa Abrego-Collier
University of Chicago

Rebekah Baglini
University of Chicago

Tommy Grano
University of Chicago

Martina Martinovic
University of Chicago

See next page for additional authors

Speaker Attitude and Sexual Orientation Affect Phonetic Imitation

Abstract

Numerous studies have documented the phenomenon of phonetic convergence: the process by which speakers alter their productions to become more similar on some phonetic or acoustic dimension to those of their interlocutor. Though social factors have been suggested as a motivator for imitation, few studies have established a tight connection between these extralinguistic factors and a speaker's likelihood to imitate. The present study explores the effects of perceived sexual orientation and speaker attitude toward the interlocutor on the likelihood of imitation for extended VOT. Experimental results show that the extent of phonetic convergence (and divergence) depends on the perceived sexual orientation of the talker as well as whether the speaker is positively disposed to the interlocutor.

Authors

Alan Yu, Carissa Abrego-Collier, Rebekah Baglini, Tommy Grano, Martina Martinovic, Charles Otte III, Julia Thomas, and Jasmin Urban

Speaker Attitude and Sexual Orientation Affect Phonetic Imitation

Alan C. L. Yu, Carissa Abrego-Collier, Rebekah Baglini, Tommy Grano, Martina Martinovic, Charles Otte III, Julia Thomas, and Jasmin Urban

1 Introduction

Imitation (also convergence or accommodation) is central to human behavior, and has been observed at many levels, including postures, gestures, and facial expressions (Dijksterhuis and Bargh 2001). Within the domain of speech, imitation has been observed with respect to lexical and syntactic alignment (Pickering and Garrod 2004), speech rate (Webb 1970), pause and utterance duration (Jaffe and Feldstein 1970), vocal intensity (Natale 1975), VOT (Nielsen 2007, 2008), and vowel quality (Babel 2007, 2009). Numerous studies have documented in particular the phenomenon of phonetic convergence: the process by which speakers alter their productions to become more similar on some phonetic or acoustic dimension to those of their interlocutor. Subjects, for example, shifted their speech production in the direction of speech they were asked to shadow (Goldinger 1998). Shockley et al. (2004) found a significant VOT imitation effect in single-word shadowing for voiceless stops with artificially extended VOTs. In a non-shadowing task, Nielsen (2007) demonstrated a significant effect of implicit phonetic imitation for extended VOTs that was generalized to novel tokens. While the ability to imitate is assumed to be innate, phonetic imitation is not an entirely automatic or unrestricted process (Dijksterhuis and Bargh 2001). For example, subjectst would imitate lengthened VOTs but not shortened ones (Nielsen 2008). Men were found more likely to converge in a speech task than women and conversation role also affects a speaker's likeliness to imitate (Pardo 2006). Social factors have been suggested as important motivators for imitation (Dijksterhuis and Bargh 2001, Babel 2009). Gender difference is one that is most commonly observed, although there are conflicting results regarding which gender is more likely to imitate. For example, as mentioned earlier, Pardo (2006) found that men were more likely to converge in a map task than women, yet Namy et al. (2002) found female participants converged more than male participants in a shadowing experiment. The fact that such conflicting results exist suggests that gender might not be the right factor in mediating likelihood of imitation. Building on Pardo (2006) and Nielsen (2007), the present study explores how listeners' evaluation of interlocutor affects the likelihood of phonetic convergence for extended VOT. In particular, two evaluative factors are examined: speaker attitude toward the interlocutor and perceived sexual orientation. Our experimental results show that the extent of phonetic convergence (and divergence) depends on the perceived sexual orientation of the talker as well as whether the speaker is positively disposed to the interlocutor.

2 Methodology

2.1 Procedure

The experiment contains three phases: A *baseline* block where subjects produced a list of 72 /p, t, k/-initial target words in the carrier sentence "say ___again". The target words were selected from the CELEX lexical database and are evenly distributed by frequency and by place of articulation. A *test* block consisted of the same words plus an additional 72 words similarly balanced for frequency and place of articulation. In between the two production tasks, subjects listened to a constructed narrative where the 72 words from the *baseline* block were embedded; the narrative details the narrator's bad date from the previous night and contains no other stressed syllable-initial voiceless aspirated stops aside from the target words. Two minimally different versions of the narrative were created: one in which the narrator's date was female ("straight" condition), and one in which the narrator's date was male ("gay" condition). All subjects took a post-experiment survey which included questions about the subject's age (mean = 20.63, Range = 18–26), second language knowledge, assessment of own sexual orientation (1–7; 1 for exclusively heterosexual and 7 for exclusively homosexual), attitude towards the story narrator's behavior (1–7; 1 for very positive and 7 for very negative), likelihood

Condition	"gay"	"straight"
Number of subjects	9	7
AGE	20.67 (18–25)	20.57(18–26)
SEXUALITY	3.17 (1–7)	1.43 (1–3)
ATTITUDE	3.78 (2–6)	3.57 (1–6)

Table 1: Descriptive statistics (mean and range) of subjects' age and sexuality and attitude scores

of behaving in the same way in a similar situation, and whether anything unusual was noticed in the narrator's speech. Twenty undergraduate males who received either course credits or nominal cash took part in the study. Subjects were randomly assigned to each condition. VOTs of subjects' tokens from the *baseline* and *test* blocks were measured in Praat using both waveforms and spectrograms.

2.2 Stimuli

An adult male speaker of American English was recorded reading both "gay" and "straight" versions of the story. The "straight" recording was then manipulated in Praat to extend the initial VOT of each target word by 100%. The narrative used in the "gay" condition was created by replacing and splicing in appropriate names and pronouns from the "gay" recording to the extended-VOT recording.

3 Results

Three recordings from the "straight" condition and one recording of the "gay" condition were lost due to equipment malfunction. A total of sixteen sets of recordings were analyzed. VOT measurements, which were *z*-normalized, were modeled using three mixed-effects linear regression models. Descriptive statistics of subjects' age and sexuality and attitude scores are given Table 1

3.1 Effects of Speaker Attitude on Phonetic Accommodation

The first model focused on only the target words in the *baseline* and *test* blocks; novel words in the *test* blocks were not analyzed here. The model contains eight fixed variables: TRIAL (1–144), PLACE (p, t, k), BLOCK (baseline vs. test), CONDITION ("straight" vs. "gay"), subject's ATTITUDE toward the narrator (1–7), all possible two-way and three-way interactions between ATTITUDE, BLOCK and CONDITION.[1] Additionally, the analysis includes a by-subject random slope for TRIAL as well as a by-item random slope.[2] To eliminate collinearity, scalar variables, fixed and random, were centered, while BLOCK and CONDITION were sum-coded (i.e., gay = 0.5, straight = -0.5; test = 0.5, baseline = -0.5). A summary of the parameter estimates for the fixed effects of the first regression model and their significance is given in Table 2. As expected, the place of articulation of a consonant has a significant effect on VOT. In particular, labials have significantly shorter VOT than velars and alveolars (see Figure 1a). Consistent with Nielsen (2007)'s finding, (log-transformed) word frequency was not a significant predictor of VOT. A likelihood ratio test comparing a model with LOGFREQUENCY as a predictor and one without it shows that the added predictor does not

[1]Due to the randomness of assigning subjects to each condition, no subjects in the "straight" condition score higher than 3 on the sexuality scale, while the subjects in the "gay" condition span the whole sexuality spectrum. Subject's sexuality was not considered in this overall model due to complications with collinearity. Subject's sexuality will be considered when the data from each test condition is considered separately.

[2]The initial model included the effects of the control variables (TRIAL, subject AGE, subject's SEX, PLACE of articulation of consonants measured, LOG-FREQUENCY of words), BLOCK (baseline vs. test), CONDITION ("straight" vs. "gay"), speaker ATTITUDE, and all two-way and three-way interactions between speaker ATTITUDE, BLOCK, and CONDITION as fixed factors, as well as a by-item random slope and a random slope with TRIAL nested within SUBJECT. The final model was obtained by backward elimination, dropping in a stepwise process all of the nonsignificant effects. The results presented here are not affected by collinearity.

Predictor	Coef.β	SE(β)	t
Intercept	0.1293	0.0757	1.815
TRIAL	0.0017	0.0011	1.486
PLACE = *labial*	-0.5844	0.1071	-5.456 ***
PLACE = *coronal*	-0.0090	0.1054	-0.086
BLOCK	-0.0896	0.0364	-2.460 *
CONDITION	0.0361	0.0360	1.002
ATTITUDE	-0.0120	0.0134	-0.893
BLOCK X CONDITION	-0.0693	0.0710	-0.976
BLOCK X ATTITUDE	0.1003	0.0266	3.768 ***
CONDITION X ATTITUDE	0.0083	0.0269	0.308
BLK X COND X ATTITUDE	-0.1283	0.0532	-2.411 *

Table 2: Result summary: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. All statistical significance are determined via likelihood ratio tests comparing a model with a predictor and one without it.

significantly improve model log-likelihood ($\chi^2 = 1.2757$, $df = 1$, $p = 0.2587$). There is a significant difference between VOT values in the *baseline* block and those in the *test* block, suggesting that the *exposure* block has an effect on the VOT realization in the *test* block. In particular, VOT in the *test* block is significantly shorter than those in the *baseline* block, suggesting that, overall, the subjects are diverging, rather than converging, towards the narrator's speech. A significant interaction between BLOCK and ATTITUDE suggests the degree of divergence is mediated by the attitude of the subject toward the narrator. As shown in Figure 1b, the more negative the subject is toward the narrator, the larger the divergence effect is in the *test* block. There is a significant three-way interaction between BLOCK, CONDITION, and ATTITUDE suggesting that subjects' divergence patterns differ depending on whether they were exposed to the "straight" or the "gay" condition. We consider this three-way interaction further in the next section.

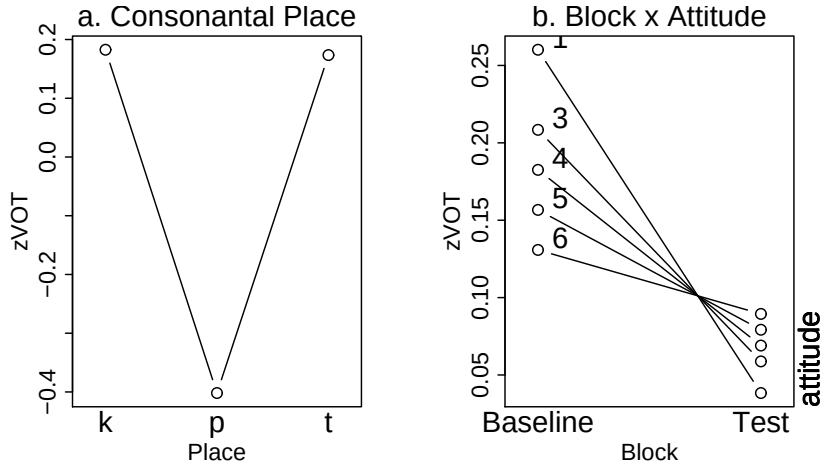

Figure 1: a. The effect of place of articulation on VOT; b. Interaction between test blocks and speaker attitude. The predictor variables were back-transformed to their original scales in the figure.

3.2 Discussion

Our findings so far show that phonetic accommodation is not only not automatic (Namy et al. 2002, Pardo 2006, Nielsen 2007, 2008), it is also strongly dependent on the evaluative judgment of the speaker. In this present case, whether the subject views the narrator favorably or not (which likely reflects whether the subject agrees with the narrator's behavior in the story) affects the subject's ten-

Predictor	Coef.β	SE(β)	t
Intercept	0.2060	0.1066	1.933
TRIAL	0.0019	0.0016	1.190
PLACE = *labial*	-0.6014	0.1158	-5.195 ***
PLACE = *coronal*	-0.0067	0.1136	-0.059
BLOCK	-0.5528	0.1372	-4.031***
SEXUALITY	-0.0117	0.0504	-0.233
ATTITUDE	-0.0182	0.0231	-0.788
BLOCK X SEXUALITY	-0.2862	0.0972	-2.943 **
BLOCK X ATTITUDE	0.2471	0.0444	5.571 ***

Table 3: Result summary: "straight" condition only; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

dency toward phonetic convergence or divergence. Of interest is the fact that whether the narrator is perceived to be straight or gay appears to play a role in determining the subject's convergence pattern as well. To further examine the role perceived sexuality plays in influencing the likelihood of phonetic convergence and divergence, in the next section, we consider whether the perceived sexuality of the narrator and subjects' own sexuality interacts to influence the phonetic accommodation pattern.

3.3 Effects of Perceived Sexuality on Phonetic Accommodation

Two additional regression models were constructed for each narrative condition. Like the general model, the two sub-models have the following fixed factors: TRIAL (1–144), PLACE (/p, t, k/), BLOCK (baseline vs. test), and subject's ATTITUDE toward the narrator (1–7) and the same random factors (i.e., a by-subject random slope for TRIAL as well as a by-item random slope). Subject's SEXUALITY (1–7) was also included in the sub-models as a fixed factor. The "straight" condition model includes two-way interactions of BLOCK with SEXUALITY or with ATTITUDE. The "gay" condition model has an additional two-way interaction between SEXUALITY and ATTITUDE and a three-way interactions between BLOCK, ATTITUDE, and SEXUALITY.[3] Summaries of the parameter estimates for the fixed effects of the regression models for the "straight" and "gay" conditions and their significance are given in Table 3 and Table 4 respectively.

Recall that the overall model reveals a significant three-way interaction between BLOCK, CONDITION, and ATTITUDE. Here, the "straight" condition regression model shows a significant interaction between BLOCK and ATTITUDE but this interaction does not reach significance in the "gay" condition model. As illustrated in Figure 2, which compares the two two-way interactions between BLOCK and ATTITUDE in the "gay" vs. "straight" condition, subjects in the "gay" condition show either no effect or were mildly divergent after the *exposure* block regardless of how favorable they were toward the narrator. On the other hand, subjects in the "straight" condition showed convergence after the *exposure* block if they thought poorly of the narrator.

Focusing now on the role of the subject's own sexual orientation in affecting the likelihood of phonetic convergence, the "straight" condition model reveals a significant interaction between BLOCK and SEXUALITY. As illustrated in Figure 3, subjects in the "straight" condition show sign of convergence when they are exclusively heterosexual (i.e., when the sexuality score is 1). Otherwise, phonetic divergence is prevalent.

While the two-way interaction between BLOCK and SEXUALITY did not reach statistical significance in the "gay" condition, there is a significant three-way interaction between BLOCK, SEXUALITY, and ATTITUDE. To examine further this three-way interaction, the "gay" condition data is subdivided into two bins for further analysis. Subjects were binned by their sexuality score; subjects are classified as "gay" if their sexuality scores range from 4–7, otherwise, they are classified as "straight". As illustrated in Figure 4, regardless of their views of the narrator, straight subjects

[3] BLOCK was sum-coded in both sub-models. To avoid collinearity, SEXUALITY and ATTITUDE were centered in the "gay" condition model. Centering was not necessary for the "straight" condition model.

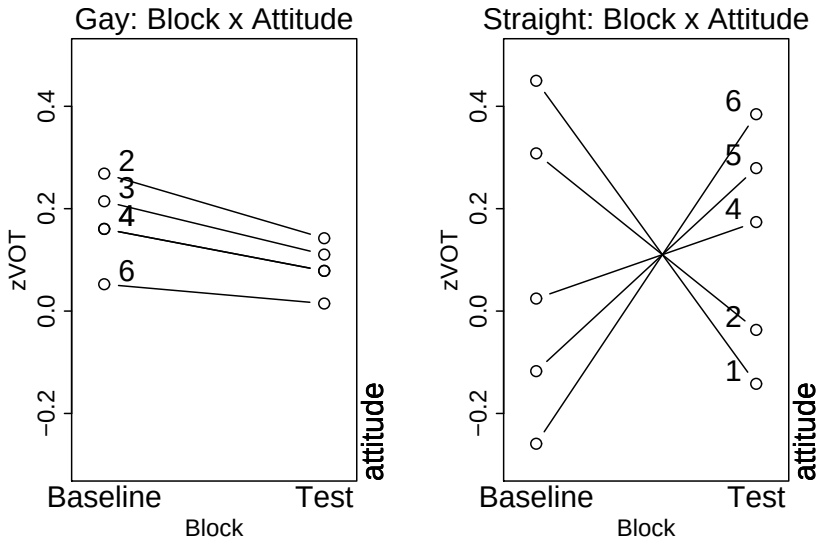

Figure 2: Interaction between test blocks and speaker attitude in the "gay" and "straight" test conditions

Predictor	Coef.β	SE(β)	t
(Intercept)	0.1366	0.0877	1.558
TRIAL	0.0014	0.0014	1.018
PLACE = *labial*	-0.5892	0.1119	-5.266 ***
PLACE = *coronal*	-0.0151	0.1099	-0.138
SEXUALITY	-0.0041	0.0120	-0.341
BLOCK	-0.1280	0.0478	-2.678 **
ATTITUDE	-0.0041	0.0273	-0.151
SEXUALITY X BLOCK	-0.0334	0.0201	-1.664
SEXUALITY X ATTITUDE	0.0328	0.0133	2.476
BLOCK X ATTITUDE	0.0830	0.0454	1.827
SXLTY X BLK X ATTITUDE	0.0514	0.0219	2.344 *

Table 4: Result summary: "gay" condition only; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

uniformedly diverge, albeit weakly, after being exposed to the narrator's extended VOT. On the other hand, gay subjects who are most adverse to the narrator (i.e., attitude score = 5) show a tendency for convergence; those who are more ambivalent toward the narrator (attitude score = 3 or 4) generally show divergence after exposure to the narrator's speech. Our findings suggest that, when the perceived sexual orientation of the narrator is gay, gay subjects might vary their accommodation pattern depending on how positive they are toward the narrator. Straight subjects, on the other hand, show a consistent, though weak, divergence pattern regardless of the subject's attitude toward the narrator.

Concerning the issue of whether the phonetic accommodation effect is extended to novel contexts, we tested for potential differences between the familiar forms (i.e., words that appear in both *baseline* and *test* blocks) and the novel forms in the *test* block. A repeated measures ANOVA shows that novel words were pronounced with a significantly longer VOT than familiar words ($F(1, 2201) = 9.61$, $p > 0.01$). Unfortunately since the novel forms were not present in the *baseline* block, we are unable to ascertain whether the longer VOT is due to a failure to generalize the phonetic accommodation effect to novel contexts or whether this is a general effect of pronouncing novel words for the first time.

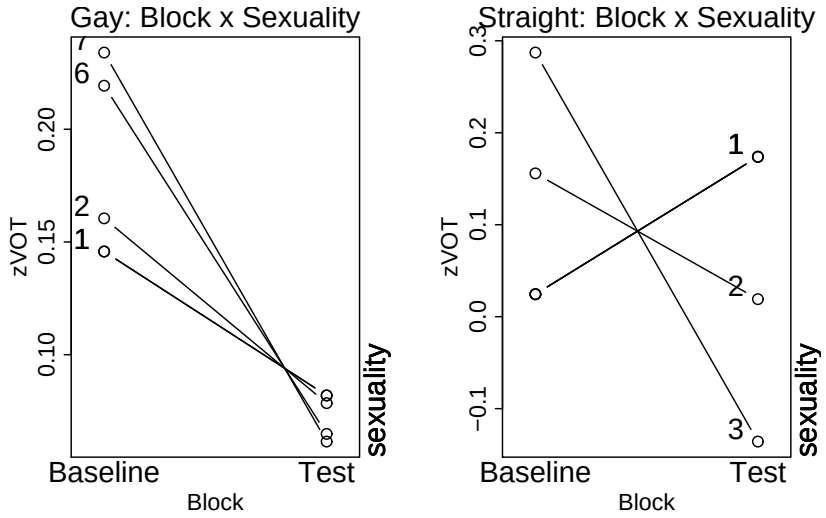

Figure 3: Interaction between test blocks, condition, and subject's sexuality within the "straight" condition.

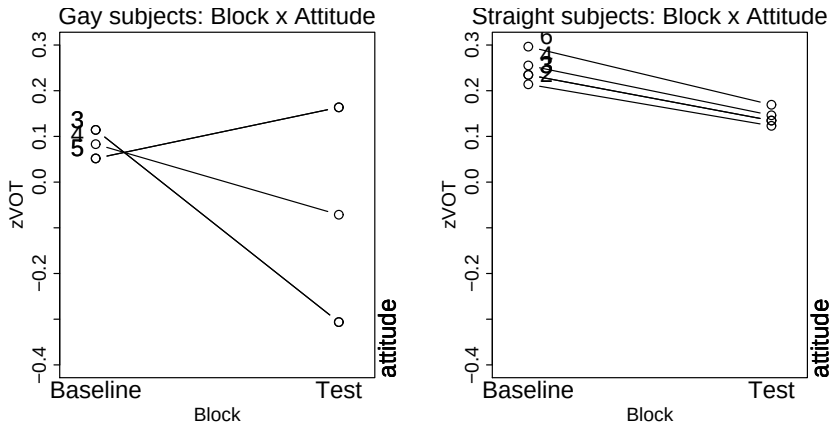

Figure 4: Interaction between test blocks, subject's attitude and subject's sexuality.

4 Conclusion

A growing number of studies has shown that phonetic accommodation is not an entirely automatic process. Yet, little is known still regarding the factors that mediate the likelihood of phonetic accommodation. This study suggests that an individual's evaluative judgement toward the interlocutor plays a significant role in affecting the likelihood and the directionality of phonetic accommodation. Phonetic convergence is not guaranteed in phonetic accommodation. Phonetic divergence is likely when the speaker is not positively disposed toward the interlocutor. Other social perceptual factors might also come into play. For example, while subjects in the "straight" condition may converge or diverge phonetically after being exposed to the narrator's VOT pattern, subjects show little movement or weak divergence when the narrator is perceived to be gay. However, the observed convergence and divergence effects might be further mediated by the subject's sexual orientation. It is worth noting, however, that the generalizability of the present findings is inherently limited given the gender-specific nature of the subject pool. It is not clear, for example, whether women would be more repulsed by the narrator's action and would consequently diverge even more than our male subjects. Also, the lack of "gay" subjects in the "straight" condition also raise concerns about the

validity of the three-way interaction between BLOCK, CONDITION, and subject's SEXUALITY. Further investigation is obviously needed.

It is not clear at this point why subjects exhibit phonetic convergence mainly when they thought poorly of the narrator. The prevalence of phonetic divergence in this study contrasts sharply with the convergence effects observed in Nielsen (2007, 2008). The exposure materials in Nielsen (2007, 2008) were English words presented isolation, while our exposure materials were embedded in a meaningful narrative. The marked difference in experimental results might be attributable to the decontextualization of the exposure materials in Nielsen's (2007, 2008) studies; imitation might be more automatic in a sterile context where the words presented have no meaning beyond the context of the experiment. The narrative in the present study, on the other hand, invites the listeners to make evaluative judgements on the narrator as the narrator recounts his blind date the night before where he left the blind date at the dinner table to go after a person sitting across the hall at the restaurant. The rude behavior of the narrator might give our subjects pause when "deciding" whether or not to phonetically imitate the narrator's speech pattern.

References

Babel, Molly E. 2007. Social judgements and their acoustic cues in read speech. *UC Berkeley Phonology Lab Annual Report* 226–264.

Babel, Molly E. 2009. Phonetic and social selectivity in speech accomondation. Doctoral Dissertation, University of California, Berkeley.

Dijksterhuis, Ap, and John A. Bargh. 2001. The perception-behavior expressway: Automatic effects of social perception on social behavior. In *Advances in experimental social psychology. Vol. 33*, ed. M. P. Zanna, 1–40. San Diego: Academic Press.

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

Jaffe, Joseph, and Stanley Feldstein. 1970. *Rhythms of dialogue*. New York: Academic Press.

Namy, Laura L., Lynne C. Nygaard, and Denise Sauerteig. 2002. Gender differences in vocal accommodation: The role of perception. *Journal of Language and Social Psychology* 21:422–432.

Natale, Michael. 1975. Convergence of mean vocal intensity in dyadic communication as a function of social desirability. *Journal of Personality and Social Psychology* 32:790–804.

Nielsen, Kuniko. 2008. Word-level and Feature-level Effects in Phonetic Imitation. Doctoral Dissertation, UCLA.

Nielsen, Kuniko Y. 2007. Implicit phonetic imitation is constrained by phonemic contrast. In *Proceedings of the 16th International Congress of the Phonetic Sciences, Saarbrcken, Germany*, 1961–1964. Saarbrcken, Germany.

Pardo, Jennifer S. 2006. On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America* 119:2382–2393.

Pickering, M.J., and S. Garrod. 2004. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences* 27:169–225.

Shockley, Kevin, Laura Sabadini, and Carol A. Fowler. 2004. Imitation in shadowing words. *Perception & Psychophysics* 66:422–429.

Webb, James T. 1970. Interview synchrony: An investigation of two speech rate measures in an automate standardized interview. In *Studies in dyadic communication: Proceedings of a research conference on the interview*, ed. A. W. Siegman and B. Pope, 115–133. New York: Pergamon.

Department of Linguistics
University of Chicago
1010 E 59t Street
Chicago, IL 60637
aclyu@uchicago.edu
carissa@uchicago.edu
rebekahbrita@uchicago.edu

tgrano@uchicago.edu
martinam@uchicago.edu
cjotte@uchicago.edu
jmthomas@uchicago.edu
jasmindu@mac.com