EFFECTS OF SPEAKER EVALUATION ON PHONETIC CONVERGENCE

Carissa Abrego-Collier, Julian Grove, Morgan Sonderegger & Alan C. L. Yu

Phonology Laboratory, Department of Linguistics, University of Chicago, USA
carissa@uchicago.edu; juliang@uchicago.edu;
morgan@cs.uchicago.edu; aclyu@uchicago.edu

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, a relatively smaller body of studies has established a tight connection between extralinguistic factors and a speaker's likelihood to imitate. The present study explores the effects of a speaker's attitude toward an interlocutor on the likelihood of imitation for extended VOT. Experimental results show that the extent of phonetic convergence (and divergence) depends on the speaker's disposition towards an interlocutor, but not on more "macro" social variables, such as the speaker's gender.

Keywords: imitation, convergence, divergence, sociophonetics, VOT, speaker attitude

1. INTRODUCTION

Imitation has been observed in many domains of human behavior, including postures, gestures, and facial expressions [5]. In the domain of speech, imitation has been observed for many properties, such as speech rate [16], pause and utterance duration [7], vocal intensity [10], vowel quality [2], and voice on- set time (VOT) [11, 12, 13]. When speakers alter their productions to become more similar on some phonetic or acoustic dimension to those of their interlocutor, phonetic *convergence* obtains; phonetic *divergence* refers to the reverse process.

For example, many results using a "shadowing" paradigm (e.g., [6]) show that subjects shift their speech production (evaluated using perceptual measures) in the direction of speech they are asked to repeat as quickly as possible. Several previous studies consider imitation of VOT in particular. Subjects showed a significant VOT imitation effect in a single-word shadowing task using words with artificially-lengthened initial VOTs [15]. Recently, Nielsen [11, 12] showed that VOT imitation is observed even when subjects were exposed only passively to stimuli with extended VOTs (i.e., they were not asked to immediately imitate these stimuli), and that subjects also generalized the extended VOT pat- tern to novel tokens. While the ability to imitate is assumed to be innate, phonetic imitation is not an entirely automatic or unrestricted process [5]. For example, one of Nielsen's experiments showed that subjects would imitate lengthened VOTs, but not shortened ones [12].

Situational variables, such as a speaker's role in a particular conversation, also affects the degree of imitation [14]. "Macro" social factors, such as gender, have been suggested as important mediators for imitation [2, 5], although the exact nature of this mediation is not clear. In the case of gender, men were found to imitate more than women in the context of a map task [14], but less than women in the context of a shadowing task [9]. These conflicting results suggest that gender may not be the appropriate predictive factor in mediating likelihood of imitation.

Building on previous work on VOT imitation, and how imitation is mediated by situational variables and social factors, the present study explores how both types of social variables affect the extent of imitation of extended VOT. Two situational variables (narrative outcome, and subject attitude towards the narrator) and two social variables (subject gender, and perceived sexual orientation of the talker) are examined. Our results show that the extent of VOT imitation is largely a function of whether a subject is positively disposed towards his/her interlocutor.

2. METHODOLOGY

2.1. Procedure

The experiment contained three phases: First, there was a baseline production block where subjects produced a list of 72 p/t/k-initial target words (randomized order) in the carrier sentence "say __ again". Target words were selected from CELEX2 [1], evenly distributed by frequency quartile and by initial consonant. A subsequent test block consisted of subjects producing the same word list again in a different randomized order. In between the two production tasks was a listening phase where subjects heard a constructed first-person narrative in which

the same 72 p/t/k words were embedded. VOTs for the target words in the story were extended by 100% using Praat. The narrative described a male talker's blind date from the previous night and contained no other stressed syllable-initial voiceless aspirated stops aside from the target words.

Two versions of the narrative were created: one in which the talker abandons his date and goes home alone ("negative" version), and one in which the date goes well and they leave together ("positive" version). For each version, there were two conditions: one in which the talker's date was female ("straight" condition), and one in which the talker's date was male ("gay" condition). This resulted in a total of 4 possible conditions. The narrative used in the "gay" condition was created by replacing and splicing in appropriate names and pronouns from the "straight" recording to the extended-VOT recording. All subjects also took a post-experiment survey which included questions about the subject's age, second language knowledge, assessment of own sexual orientation (from 1=exclusively heterosexual to 7=exclusively homosexual), feelings towards the talker (from 1=very positive to 7=very negative), likelihood of behaving in the same way in a similar situation (yes/no), and whether anything unusual was noticed in the talker's speech.

Fifty-eight subjects took part in the study, and received either course credit or a nominal cash payment. Participants were assigned to one of the four conditions. Approximately equal numbers of subjects participated in each of the conditions (positive/negative x gay/straight; see Table 1). The deviation from a fully-balanced design is of no consequence for the mixed-effects regression used in our analysis. VOTs of subjects' tokens from the *baseline* and *test* blocks were manually measured in Praat using both waveforms and spectrograms.

3. RESULTS

While fifty-eight subjects were recorded, eight subjects (at least 1 but no more than 3 per condition) were lost due to equipment malfunction. One subject did not give an ATTITUDE score, and was thus excluded from the analysis. One additional subject was classified as an outlier, due to an extremely high mean difference in VOT between blocks (>3 s.d. from the mean, considering all subjects' mean VOT differences), and was also excluded. The following analysis was performed on the remaining 48 sets of recordings. Descriptive statistics of subjects' age, sexuality, and attitude scores are given in Table 1.

Table 1: Median & range of subject age, and sexuality and attitude scores.

condition	**gay**	**straight**
positive	14 subjects	14 subjects
AGE	19 (18–36)	20 (18–24)
SEXUALITY	2 (1–7)	2 (1–7)
ATTITUDE	3 (1–5)	3 (1–6)
negative	11 subjects	11 subjects
AGE	20 (18–38)	20 (19–32)
SEXUALITY	2 (1–4)	2 (1–7)
ATTITUDE	4 (1–7)	4 (1–7)

3.1. Model

Subjects' VOTs are analyzed using a linear mixed-effects model fitted in `R`, using the `lmer()` function from the `lme4` package [3].

Predictors The model contains several types of predictors. BLOCK (2 levels) indexed whether a measurement from the *baseline* or *test* block, and TRIAL (1–72) the within-block position of its host word. The model included 4 *social predictors*: SUBJECT GENDER (male vs. female), NARRATOR SEXUALITY (gay vs. straight), subject ATTITUDE towards the talker (1–7), and narrative OUTCOME (positive vs. negative). CONSONANT (/p/, /t/, /k/) indexed which stop the host word began with, SYLLABLES its length in syllables (range: 1–4), and FREQUENCY its log-transformed CELEX frequency. Continuous predictors (TRIAL, FREQUENCY, SYLLABLES, ATTITUDE) were z-scored; two-level factors (BLOCK, GENDER, SEXUALITY, OUTCOME) were sum-coded; CONSONANT was Helmert-coded (contrasts: p vs. t, p/t vs. k). Finally, two predictors indexed the SPEAKER (48 levels) and WORD (72 levels) associated with each measurement.

Random effects: To allow for word-specific and speaker-specific variation in VOT, the model included by-SPEAKER and by-WORD random intercepts. Exploratory data analysis suggested that some speakers' VOTs increased or decreased steadily over the course of each block, and that the slope of this change could differ by block. To control for this possibility, we included by-SPEAKER random slopes of BLOCK, TRIAL, and BLOCK:TRIAL. In the final model, all random slopes and intercepts made significant contributions to model likelihood (p<0.001). All random effect terms were assumed to be uncorrelated; this led to an extremely similar model to one where this was not assumed, and allowed us to obtain p-values calculated by MCMC sampling.

Fixed effects: Main effect terms for CONSONANT, SYLLABLES, and FREQUENCY, were included, to

control for the well-known effect of place of articulation on VOT (p<t<k), and to allow for the possibility that VOT is negatively correlated with the number of syllables and frequency of the host word. To test for the effect of BLOCK on VOT, as well as its interaction with social predictors, terms were included for the interactions of BLOCK with GENDER, SEXUALITY, and ATTITUDE, as well as a main effect term for each of these predictors. To test for the possibility that the interactions of social predictors with BLOCK are not independent, we tested the effect of adding each BLOCK:X:Y interaction (separately), where X and Y are social predictors, along with the X:Y term required by the hierarchy principle. No such interaction significantly improved data likelihood (p>0.1). The distributions of responses for subject's sexual orientation, age, and likelihood of behaving similarly were all heavily skewed, and were thus not included in the analysis here.

Finally, main effects of BLOCK and TRIAL, as well as a BLOCK:TRIAL interaction, were included because of the corresponding random slope terms included in the model.[1]

Table 2: Estimates for all fixed-effect predictors in the mixed-effect model.

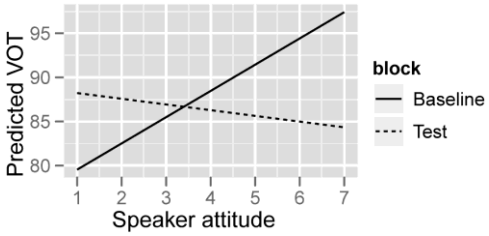
	Coef β	SE(β)	t	p_{MCMC}
Intercept	82.54	2.47	33.36	**<0.001**
FREQUENCY	1.80	1.20	1.49	>0.09
SYLLABLES	−2.68	1.23	−2.19	**<0.05**
CONSONANT (P/T)	8.10	1.34	6.05	**<0.001**
CONSONANT (PT/K)	3.27	0.75	4.34	**<0.001**
BLOCK	−2.18	0.89	−2.44	**<0.05**
ATTITUDE	1.16	2.31	0.50	>0.6
OUTCOME	−2.71	4.70	−0.58	>0.3
SUBJECT GENDER	3.39	4.34	0.78	>0.2
NARRATOR SEXUALITY	1.90	4.51	0.42	>0.5
TRIAL	0.31	0.33	0.92	>0.3
BLK:ATTITUDE	−3.62	0.95	−3.83	**<0.001**
BLK:OUTCOME	−2.25	1.88	−1.20	>0.2
BLK:SUBJECT GENDER	−2.11	1.80	−1.17	>0.2
BLK:NARR. SEXUALITY	1.22	1.80	0.68	>0.4
BLK:TRIAL	0.02	0.68	0.03	>0.9

3.2. Discussion

We omit discussion of random effect terms due to space constraints. Table 2 lists estimated values for all fixed-effect predictors, with p-values computed by MCMC sampling. We note that, because of how we have coded predictors, a fixed-effect term (e.g., BLOCK) which participates in significant higher-order interactions (BLOCK:ATTITUDE) can be interpreted as the effect of a unit change in the predictor when the other variables involved in those inter- actions (ATTITUDE) are held at their average values across the dataset.

Both CONSONANT contrasts are highly significant, confirming that the expected p<t<k ordering holds. There is a significant negative effect of SYLLABLES (p<0.05): VOT is shorter for words containing more syllables, in line with previous work (e.g., [8]). However, we did not find a significant effect of word frequency on VOT (*contra* [11]).

Figure 1: Predicted VOT as a function of ATTITUDE and BLOCK, with all other predictors held constant. 1=very positive; 7=very negative.



Of primary interest are the effects of BLOCK and its interactions with social predictors. There is a significant negative effect of BLOCK (p<0.05), indicating that subjects show divergence, on average: VOT is slightly shorter (by 2.2 msec, the coefficient of BLOCK) after listening to the story. However, the effect of BLOCK is strongly mediated by subject attitude, as reflected in the significant BLOCK:ATTITUDE interaction (p<0.001). Fig. 1 shows the model's predicted VOT, as a function of these variables: subjects with a positive attitude towards the talker (lower ATTITUDE) show convergence, while those with a negative attitude show divergence. The model shows no significant interaction of BLOCK with narrative outcome (p>0.2), subject gender (p>0.2), or talker sexuality (p>0.4).

4. GENERAL DISCUSSION

The significant interaction between BLOCK and ATTITUDE establishes that the likelihood of phonetic imitation is mediated by participants' evaluation of the narrator. Two evaluation factors were considered here, the participant's attitude toward the talker and the outcome as depicted by the narrative. Recall that there were two possible outcomes to the blind date as recounted by the narrator during the listening phase of the experiment. In the positive scenario, the narrator and his date went on well, while in the negative scenario, the narrator behaved rudely by leaving the blind date in a lurch.

Importantly, although there is some correlation between participant attitude and narrative outcome (participants who hear the positive outcome have a more positive attitude towards the narrator), it is weak (Spearman's ρ^2 =0.085, p<0.05). That is, participants do not all react negatively toward the talker under the negative scenario; similarly, not all participants are positively disposed toward the talker in the positive scenario. The only factor which

influences convergence is attitude towards the speaker: on average, participants' show a decrease in VOT between blocks, but speakers with a positive opinion of the narrator show an increase in VOT. Neither "macro" social variable—subject gender nor narrator sexuality—was found to play a role.

Our results suggest that the dynamics of phonetic imitation is mediated by factors such as speaker attitude that are constructed situationally instead of "macro" social variables such as speaker gender and perceived sexual orientation of an interlocutor. This finding, in line with other recent sociophonetic studies (e.g., [4]), highlights the importance of taking into account social variables—such as those indexing attitudes and "stances"—which are defined relative to a particular social situation.

The prevalence of phonetic *divergence* in this study contrasts sharply with the convergence effects observed by Nielsen [11, 12]. The exposure materials in Nielsen's studies were English words presented in isolation, while our exposure materials were embedded in a meaningful narrative. The marked difference in experimental results might be partly attributable to the decontextualization of the exposure materials in Nielsen's studies; imitation might be more automatic in a context where the words are presented in isolation from a social context. The narrative in the present study, in contrast, allows participants to make evaluative judgments on the narrator as he recounts his blind date. Another possibility not explored here is that subject's evaluation of the narrator's speech itself might have played a role in the direction and extent of imitation. A majority of subjects reported noticing unusual features of the narrator's speech, describing it as "articulate", "aspirated", or "robotic". The overall divergence observed may be due to subjects moving away from speech they find unusual.

5. CONCLUSION

In sum, the present study shows that an individual's evaluative judgement toward the interlocutor plays a significant role in affecting the likelihood and the directionality of phonetic accommodation. Crucially, unlike many early studies of phonetic imitation, phonetic *divergence* is found as well as *convergence*, depending upon the speaker's disposition towards the interlocutor. This suggests that phonetic imitation might be influenced by cognitive as well as social factors simultaneously.

6. ACKNOWLEDGEMENTS

Authors' names are in alphabetical order. We thank James Kirby for valuable comments.

7. REFERENCES

[1] Baayen, R., Piepenbrock, R., Gulikers, L. 1996. *CELEX2 (CD-ROM)*. Philadelphia: LDC.

[2] Babel, M.E. 2009. *Phonetic and Social Selectivity in Speech Accommodation*. Ph.D. thesis, University of California, Berkeley.

[3] Bates, D., Maechler, M., Bolker, B. 2011. *Lme4*. R package version 0.999375-38.

[4] Campbell-Kibler, K. 2010. Sociolinguistics and perception. *Language and Lingusitics Compass* 4(6), 377-389.

[5] Dijksterhuis, A., Bargh, J. 2001. The perception-behavior expressway: Automatic effects of social perception on social behavior. In Zanna, M.P., (ed.), *Advances in Experimental Social Psychology*. San Diego: Academic Press 33, 1-40.

[6] Goldinger, S. 1998. Echoes of echoes? An episodic theory of lexical access. *Psych. Rev*. 105, 251-279.

[7] Jaffe, J., Feldstein, S. 1970. *Rhythms of Dialogue*. New York: Academic Press.

[8] Klatt, D. 1975. Voice onset time, frication, and aspiration in word-initial consonant clusters. *JSLHR* 18(4), 686-706.

[9] Namy, L.L., Nygaard, L.C., Sauerteig, D. 2002. Gender differences in vocal accommodation: The role of perception. *Journal of Language and Social Psychology* 21(4), 422-432.

[10] Natale, M. 1975. Convergence of mean vocal intensity in dyadic communication as a function of social desirability. *Journal of Personality and Social Psychology* 32(5), 790-804.

[11] Nielsen, K.Y. 2007. Implicit phonetic imitation is constrained by phonemic contrast. *Proc. 16th ICPhS*, 1961-1964.

[12] Nielsen, K.Y. 2008. *Word-level and Feature-level Effects in Phonetic Imitation*. Ph.D. thesis, UCLA.

[13] Nielsen, K.Y. 2011. Specificity and abstractness of VOT imitation. *J. Phon.* 39.

[14] Pardo, J. 2006. On phonetic convergence during conversational interaction. *JASA* 119(4), 2382-2393.

[15] Shockley, K., Sabadini, L., Fowler, C.A. 2004. Imitation in shadowing words. *Perception & Psychophysics* 66(3), 422-429.

[16] Webb, J.T. 1970. Interview synchrony. In Siegman, A. W., Pope, B. (eds.), *Studies in Dyadic Communication: Proceedings of a Research Conference on the Interview*. New York: Pergamon, 115-133.

[1] The model formula in lme4-style is: VOT ~ FREQUENCY + SYLLABLES + CONSONANT + BLOCK * (ATTITUDE + OUTCOME + GENDER + SEXUALITY + TRIAL) + (1|SUBJECT) + (-1+BLOCK|SUBJECT) + (-1+TRIAL|SUBJECT) + (-+BLOCK:TRIAL|SUBJECT) + (1|WORD), where predictors are coded as described in the text.