THE EFFECTS OF TALKER VARIABILITY ON PHONETIC ACCOMMODATION

Molly Babel and Grant McGuire

University of British Columbia and University of California, Santa Cruz
molly.babel@ubc.ca, gmcguir1@ucsc.edu

ABSTRACT

This paper compares spontaneous phonetic accommodation in high and low variability talker conditions. If multi-talker processing is a kind of increased cognitive load, we predict an increase in imitation in the low variability contexts. Our results suggest that, indeed, phonetic accommodation is greater in low variability contexts, but that this higher level of accommodation is moderated as perceptual experiences unfold in the task. That is, while phonetic accommodation in the low variability context stays relatively constant across the task, participants in the high variability context imitate less across the course of the experiment.

Keywords: phonetic accommodation, talker variability, speech perception

1. INTRODUCTION

High talker variability learning contexts make listeners learn and generalize more effectively [6, 4, 11]. These results suggest that variable environments lead to increased abstraction or improved generalizations, which are advantageous when listeners are later presented with a novel voice. This is predicted by Minerva 2 [9] and shown through experimentation by [8] as comparisons of repetitions of the same items (words or sentences) by different voices will be forced to generalize via their comparison.

At the same time, talker variability also has the consequence of reducing performance in speech processing tasks. Exposure to multiple voices in an experimental paradigm slows responses and lowers performance across a variety of tasks (e.g., [16, 13, 18]). The contextual tuning hypothesis suggests this is a relatively active process that is somewhat under the control of listeners, as the slowing and decrease in performance is susceptible to expectations regarding talker variability [12, 5]. The idea that this variability is somewhat controlled by the listener melds nicely with the concept that there is considerable inter-listener variability in listening strategies [20].

The effects of talker variability on speech perception are similar to the effects induced by an increase in cognitive load. Given that cognitive load decreases listeners' ability to do low-level sensory analyses [14], we expect talker variability to negatively effect phonetic accommodation.

Phonetic accommodation offers a way to investigate the role of talker variability in speech perception and production as it illustrates that perceived phonetic experiences affect speech production – simply, the phenomena reveals that there is some bridge or connection between speech perception and production [7]. This bridge is subject to social moderation [2, 1] with social preferences increasing the magnitude of the effects. Phonetic accommodation has been shown in auditory-naming tasks (e.g., [8]) and more naturalistic dyadic interactions [19, 10]. While it has been suggested that phonetic accommodation only targets phonologically meaningful phonetic detail [15], others have shown that subphonemic phonetic details are subject to imitation [2, 1, 3, 17].

In this experiment we use a phonetic accommodation paradigm to probe the kind of phonetic detail available to listeners/speakers in contexts of high talker variability – where phonetic detail is predicted to be sparser – and low talker variability – where phonetic detail should be more readily available.

2. EXPERIMENTS

We describe the auditory naming task in which participants shadow the productions of model talkers and the AXB similarity judgment task which quantified phonetic imitation.

2.1. Auditory naming task

2.1.1. Stimuli

Five female participants were recorded reading word lists on a head-mounted microphone at a 22kHz sampling rate. We refer to these participants as model talkers. Words were presented visually on a screen in a random order to prevent list intonation. Fifteen unique words were selected for each model talker for use as stimuli in the auditory naming task. That is, there were a total of 75 words in the task, with each model having her own set of 15 words.

2.1.2. Participants

Twenty female participants (n=10 in each in a High and Low Variability condition) completed the task. These participants will be referred to as shadowers. Shadowers were native speakers of Western American English with no known speech, language, or hearing disorders. They were compensated for their time with course credit.

2.1.3. Procedures

Shadowers were tested individually and were seated in a sound-attenuated booth in front of a computer screen. Sound files were played and recorded through an AKG model HSC 271 MKII headset microphone and M-Audio Audio Buddy pre-amplifier. Productions were digitally recorded onto the hard drive of a windows PC at a 22kHz sampling rate.

The task proceeded as follows. Baseline productions for each word were recorded. Shadowers read each word as presented in the middle of the computer screen in 36 point font in a randomized order; they were instructed to read each word aloud as naturally and clearly as possible. Following the baseline productions, shadowers were presented with the auditory tokens for shadowing. Those in the High Variability condition were presented with each of the 15 words from each of the 5 model voices (a total of 75 trials) in a random order, while those in the Low Variability condition were presented with the same 75 items, but blocked by model voice. Within each model voice the 15 words were randomly ordered, and the order of the 5 model voices were randomly ordered as well. Neither conditions had breaks.

2.2. AXB Similarity Task

To quantify convergence we used an AXB similarity task where listeners determined whether shadowed or baseline tokens were more similar to the model talker voices.

2.2.1. Stimuli

Tokens produced by the shadowers in the auditory naming task were used as flanking stimuli (A and B) along with the model talkers tokens used as prompts in the auditory naming task (X).

2.2.2. Participants

Seventy-five listeners were used in assessing the shadowers from the High Variability condition and 78 were used for the Low Variability shadowers. All participants were native speakers of North American English with no reported speech, language, or hearing disorders. They were compensated $10 for their time.

2.2.3. Procedures

Listeners were seated at a computer workstation and presented with auditory stimuli over AKG K240 headphones using E-Prime. The basic procedure was an AXB similarity judgment. Each trial consisted of three sound files separated by a 300 ms ISI. The middle token (X) was always a token from the model talker, and the first (A) and third (B) tokens were baseline and shadowed productions from a single participant. Each trial consisted of a single lexical item. If a participant had misread or mispronounced the baseline production, that word for that participant was not used in the task. The design was blocked by shadower, and within that, blocked by model. Each shadower/model word comparison was played twice to counterbalance the order of the baseline and shadowed tokens. The order of presentation within each shadower was fully randomized. Listeners were presented with the productions from two shadowers, and were offered a break between these voice sets. Each shadower was assigned to at least 15 listeners.

The listeners' task was to determine which participant production sounded more like the model talker. If listeners consistently select the shadowed token as more similar, that is taken as evidence of convergence. If listeners consistently select the baseline token as more similar to the model, that is taken as evidence of divergence; that is, the participant sounded more like the model during baseline productions and sounded less like the model during the auditory naming task. If listeners choose baseline and shadowed tokens with equal probability, it suggests that the shadower did not modify their speech as a result of exposure to the model.

3. ANALYSIS AND RESULTS

3.1. Phonetic Imitation

Accommodation was accessed as the proportion of shadowed tokens judged as more similar-sounding to a model talker's production of the same word. Overall accommodation in this task was low – 53%

for the Low Variability Task and 52% for the High Variability Condition – and varied across model talkers. Table 1 provides some summary statistics.

Table 1: Summary statistics for accommodation by model across the High and Low Variability Conditions

	High Variability	Low Variability
Model 1	52%	54%
Model 2	53%	56%
Model 3	51%	52%
Model 4	49%	50%
Model 5	55%	53%
Overall	52%	53%

The goal of this analysis is to assess whether there is more accommodation in the Low Variability condition compared to the High Variability condition. We included two measures of task experience in our statistical model: Trial Order across the entire experiment (a range from 1-75) and Within Model Order (1-15 for each of the five model talkers). In the Low Variability task, Within Model Order is equivalent to within block. Trial Order, which was centred prior to the analysis, and Within Model Order were entered as main effect predictors and in interactions with Condition in a generalized linear mixed effects model predicting listeners' similarity judgments. Listeners' judgments were coded with 1 indicating listeners selected a shadowed token as more similar to the model and 0 indicating listeners selected a baseline token as more similar. There were random intercepts for Shadowers, Listeners, Models, and Words. Trial Order was a random slope for Shadowers, Listeners, and Words, and Condition was random slope for Model.

The model intercept was not significant [$\beta = 0.08, SE = 0.08, z = 0.928$]. There were however interactions between Condition (High Variability) and Within Model Order [$\beta = 0.02, SE = 0.01, z = 2.09, p < 0.05$] and Condition (High Variability) and Trial [$\beta = -0.24, SE = 0.10, z = -2.387, p < 0.05$]. These results are shown in Figures 1 and 2. Across the course of the experiment shadowers in the High Variability condition spontaneously imitated less, whereas those in the Low Variability condition stayed relatively steady in their accommodation levels. Those in the High Variability condition show the same downward trend when accommodation is examined as a function of trials for a particular model talker (Within Model Talker), but those in the Low Variability condition show a slight rise.

Figure 1: The effect of Trial on phonetic accommodation in the High and Low Variability groups with glm and loess fits.

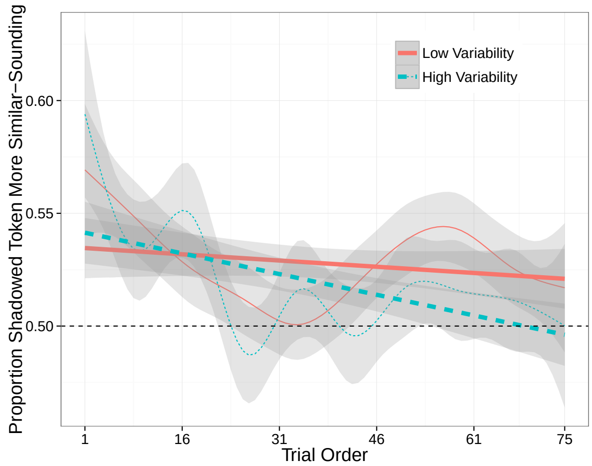
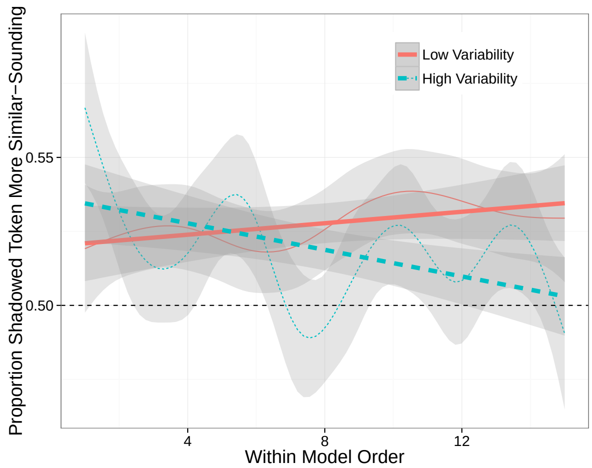


Figure 2: The effect of Within-Model Order on phonetic accommodation in the High and Low Variability groups with glm and loess fits

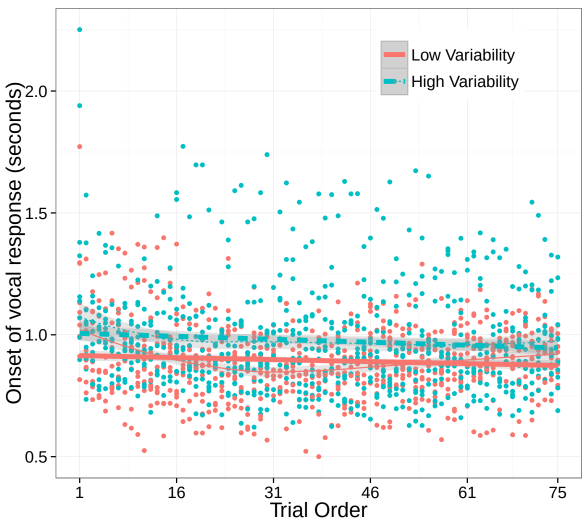


3.2. Shadowing latency

Extensive research exists on the perceptual effects of being exposed to single or mixed talker blocks. This research nearly unanimously shows that listeners in mixed talker blocks exhibit delays in response time and decreases in identification accuracy compared to listeners presented with the same contrasts in single talker blocks. This is generally taken as an indication that processing multiple voices is taxing because of the need to normalize away differences between talkers prior to identifying the linguistic message in the signal.

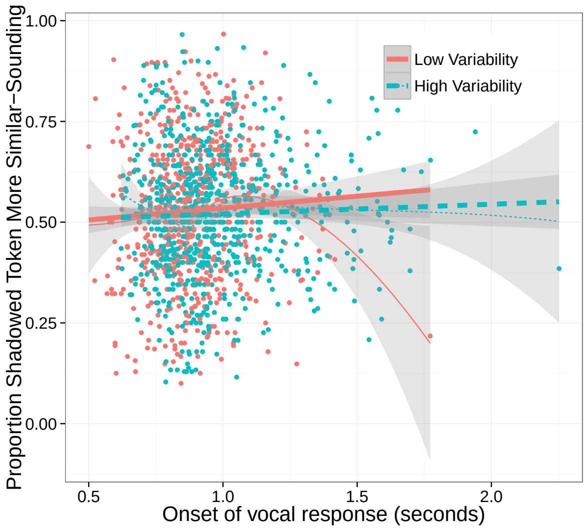
[8] argues that delays in shadowing will naturally spur feedback loops between working memory, which is holding a phonetic trace of the recently presented token, and long term memory, which will be biased “toward the central tendency of all prior traces” (p 256.) In order to explore such phenomena, we conducted a post-hoc examination of shadowing latency. It should be noted, however that unlike [8], shadowers were not deliberately slowed as a condition of the experiment. Overall shadowing latencies in the Low Variability condition were faster ($M = 894, SD = 153$) than those in the High Variability condition ($M = 976, SD = 219$). We ran a linear mixed effects model with random effects for Word, Model, and Shadower and fixed effects of Experiment, Within Model Order, and Trial Order to assess this trend. There was a significant intercept [$\beta = 0.96, SE = 0.06, t = 16.8$] along with an effect of Within Model Order [$\beta = -0.003, SE = 0.0009, t = -2.87$], Trial Order [$\beta = -0.001, SE = 0.0002, t = -5.44$], and an interaction between Experiment (High Variability) and Trial Order [$\beta = 0.001, SE = 0.0006, t = 2.46$]. This interaction can be seen in Figure 3: both shadower groups got faster through the course of the experiment, but this decrease in shadowing latency is greater for the High Variability group.

Figure 3: Shadowing latencies by trial with linear and loess fits.



We also tested whether there was a relationship between perceived imitation and response latency. This relationship was not significant [$t(1450) = 1.32, cor = 0.03$]; these data are shown in Figure 4.

Figure 4: There is no relationship between shadowing latency and perceived accommodation with linear and loess fits.



4. DISCUSSION AND CONCLUSION

Shadowers assigned to the High Variability condition phonetically accommodated less than those in the Low Variability condition, but this effect of condition varied across the task. High Variability shadowers imitated less through the course of the experiment and with increased exposure to a model talker, which entails being further along in the task. Those in the Low Variability condition did not show this overall declination as a function of trial, and there was a slight increase in imitation with more exposure to a model talker. There was no relationship between shadowing latency and accommodation. All shadowers got faster as the experiment progressed, and this increase in speed was greater for those in the High Variability condition.

Our design involved shadowers engaging in a task with a unique lexical item in each trial and conditions of either high talker variability (5 voices) or low talker variability (5 blocked voices), thus requiring participants to engage in word recognition and talker switching. This required continued attention throughout the task, and the consequences which accompany multi-talker processing led to a continuous decrease in accommodation across trials for those in the high variability group. These results support the hypothesis that multi-talker contexts are akin to an increase in cognitive load. Under this interpretation, a decrease in phonetic accommodation in multi-talker contexts is the result of less detailed sensory analyses [14].

5. ACKNOWLEDGMENTS

Thanks to Anna Tinker Valle, Jamie Russell, Jobie Hui, Michelle Chan, Rheanne Brownridge, Veronica Letawsky, and David Kurbis for their assistance with this project.

6. REFERENCES

[1] Babel, M. 2010. Dialect divergence and convergence in new zealand english. *Language in Society* 39(04), 437–456.

[2] Babel, M. 2012. Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics* 40(1), 177–189.

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

[4] Baese-Berk, M. M., Bradlow, A. R., Wright, B. A. 2013. Accent-independent adaptation to foreign accented speech. *The Journal of the Acoustical Society of America* 133(3), EL174–EL180.

[5] Barreda, S. 2012. Vowel normalization and the perception of speaker changes: An exploration of the contextual tuning hypothesis. *The Journal of the Acoustical Society of America* 132(5), 3453–3464.

[6] Bradlow, A. R., Bent, T. 2008. Perceptual adaptation to non-native speech. *Cognition* 106(2), 707–729.

[7] Goldinger, S. D. 1996. Words and voices: episodic traces in spoken word identification and recognition memory. *Journal of experimental psychology: Learning, memory, and cognition* 22(5), 1166.

[8] Goldinger, S. D. 1998. Echoes of echoes? an episodic theory of lexical access. *Psychological review* 105(2), 251.

[9] Hintzman, D. L. 1984. Minerva 2: A simulation model of human memory. *Behavior Research Methods, Instruments, & Computers* 16(2), 96–101.

[10] Kim, M., Horton, W. S., Bradlow, A. R. 2011. Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Laboratory phonology* 2(1), 125–156.

[11] Lively, S. E., Logan, J. S., Pisoni, D. B. 1993. Training japanese listeners to identify english/r/and/l/. ii: The role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the Acoustical Society of America* 94(3), 1242–1255.

[12] Magnuson, J. S., Nusbaum, H. C. 2007. Acoustic differences, listener expectations, and the perceptual accommodation of talker variability. *Journal of Experimental Psychology: Human Perception and Performance* 33(2), 391.

[13] Martin, C. S., Mullennix, J. W., Pisoni, D. B., Summers, W. V. 1989. Effects of talker variability on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 15(4), 676.

[14] Mattys, S. L., Wiget, L. 2011. Effects of cognitive load on speech recognition. *Journal of Memory and Language* 65(2), 145–160.

[15] Mitterer, H., Ernestus, M. 2008. The link between speech perception and production is phonological and abstract: Evidence from the shadowing task. *Cognition* 109(1), 168–173.

[16] Mullennix, J. W., Pisoni, D. B., Martin, C. S. 1989. Some effects of talker variability on spoken word recognition. *The Journal of the Acoustical Society of America* 85(1), 365–378.

[17] Nielsen, K. 2011. Specificity and abstractness of vot imitation. *Journal of Phonetics* 39(2), 132–142.

[18] Nusbaum, H. C., Morin, T. M. 1992. Paying attention to differences among talkers. In: Tohkura, Y., Vatikiotis-Bateson, E., Sagisaka, Y., (eds), *Speech perception, production and linguistic structure*. IOSPress: IOS Press 113–134.

[19] Pardo, J. S. 2006. On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America* 119(4), 2382–2393.

[20] Yu, A. 2010. Perceptual compensation is correlated with individuals'?autistic? traits: implications for models of sound change. *PLoS One* 5(8), e11950.