Individual-level connections between perceptual adaptation and phonetic imitation

Molly Babel, Graham Haber, and Sophia Walters
Department of Linguistics
University of British Columbia

Several researchers have recently claimed that language alignment and accommodation may provide the means by which sound change is spread through a community (Pardo, 2006; Delvaux and Soquet, 2007; Trudgill, 2008). This has recently been termed the change-by-accommodation model by Sonderegger (2012). Under such a model, interacting individuals in a speech community will accommodate to one another, thereby spreading innovations through a speech community. Innovations will also be introduced and spread through a community through imperfect accommodation (à la Ohala 1981; see also Garrett and Johnson, in press).

Phonetic imitation and accommodation has been subjected to increased study in recent years (Babel 2010, 2012; Goldinger 1998, Namy et al. 2002; Nielsen, 2011), providing evidence that speakers shift their productions in response to auditory and visual (Miller et al. 2011) stimuli. Such shifts depend on listeners being perceptually sensitive to within-category differences and updating their speech production representations accordingly. In this paper we examine how listener sensitivity to a novel pronunciation may cause shifts in production by examining how perceptual recalibration to a New Zealand accent predicts phonetic imitation to the same voice.

In the process of acquiring a first language, the input shapes perceptual categories and production "targets". Infants' perceptual abilities are shaped around the contrasts of the input language (e.g., Werker & Tees 1984) and they eventually learn to produce speech that is driven by those meaningful contrasts. In fact, Cristià (2011) showed that infants' perceptual category for /s/ was tuned to the input characteristics of their primary caretaker. This means that early on phonetic categories related to both production and perception are largely affected by ambient language. For adults, the evidence from a range of tasks indicates that individual's production contrasts predict their abilities in perception (Ghosh et al. 2010, Hay et al. 2006, Perkell et al. 2004). Individuals' abilities, however, are not necessarily indicative of their default or spontaneous behaviors.

Of interest here is the flexible and dynamic nature of perception and production categories, and how individual variation in the relationship between perception and production may serve as a catalyst for sound change. Previous research has demonstrated that listeners' perceptual categories are highly adaptable and flexible (Norris et al. 2003, Kraljic & Samuel 2005, 2006, 2007, Kraljic et al. 2008a, 2008b, Maye et al. 2008, Clayards et al. 2008), and work on phonetic accommodation has shown that speech production is labile as well. In this paper we approach the next step in understanding how flexibility in production might be related to flexibility in perception across individuals.

Previous research has touched on these issues. Modifying production through changes in auditory feedback, Shiller et al. (2009) found a relationship between the adjustments made during the manipulation of auditory feedback and perceptual judgments; they demonstrated that recalibration of perceptual categories occurs when participants' retune their productions. This finding suggests there is a connection between an individual's own perceptual categories and production "targets." Kraljic et al. (2008), however, retuned listeners' perceptual categories of /s/

and /ʃ/ and then examined whether their productions had changed as a result: they found no evidence for changes in participants' productions.

Baese-Berk (2010) recently examined the relationship between perception and production through a series of learning experiments which trained listeners to perceive the contrast between pre-voiced [d] and voiceless unaspirated [t] in CV nonwords. She compared performance across groups that received only perceptual training and both perception and production training. The production training instructions were to "repeat the token so that it was as close as possible to the token they heard" (Baese-Berk 2010:60); i.e., to explicitly imitate. Pre- and post-training discrimination and categorization tasks assessed perceptual learning, while a picture naming task was used to assess learning in production. Baese-Berk's results indicate that after two days of training production improved after perception training alone *or* with perception and production training; this indicates a coupled relationship between perception and production with some hedging necessary: participants who experienced perception-only training exhibited a connection between perceptual learning and production, but perception did not improve in conjunction with production learning for participants who received the combined perception and production training.

In summary, the results of Shiller et al. suggest changes in production lead to changes in perception; the results of Kraljic et al. show no change in production as a result of change in perception; and the results of Baese-Berk indicate changes in production as a result of changes in perception, but not changes in perception as a result of changes in production. These results are contradictory, but crucially they use different methods and stimuli. Shiller et al. and Kraljic et al. manipulate real words, thereby examining the flexibility of already instated phonetic categories. Baese-Berk's study involved training listeners to create two new categories (pre-voiced vs voiceless unaspirated categories) from the larger voiced category in English. This study integrates these lines of research to examine the flexibility of vowel categories in perception and production using natural phonetic variation and real words.

The dialect to be used as the source of perceptual learning is New Zealand English (NZE). The vowels of New Zealand English have shifted such that words which belong to the TRAP lexical set are phonetically shifted to [ɛ], as compared to the [æ] used in the same words by speakers of North American English. We predict that the perceptual recalibration due to exposure to a speaker of NZE whose production of /æ/ is lower in F1 than that of the listener population will cause listeners to increase the size of their /æ/ category, labelling more tokens on a bed-bad continuum as *bad*. In addition, we predict that in the process of phonetic accommodation participants should lower the F1 of their /æ/ words during the auditory naming task.

2. Methods

2.1 Participants

Twenty-nine participants completed the task, with participants nearly equally divided across conditions (14 = Production Condition, 15 = No Production Condition). Participants were self-identified native speakers of North American English with no reported speech, language, or hearing disorders. They were compensated $10 for their time.

2.2 Stimuli

Twenty low frequency monosyllabic and bisyllabic words with /æ/ in the stressed syllable were selected as target words. An additional 80 words without /æ/ or /ɛ/ served as fillers.

2.2.1 Pictures

Digital images were used to elicit productions from the participants.

2.2 .2 Model Talker

A female speaker of New Zealand English served as the model talker. Her productions were used to create the bed-bad continuum. All filler words were also produced by the same model talker. Recordings were made directly to a PC through a USBPre device using a head-mounted microphone at 44.1 kHz.

2.2.3 Speech Continuum

Original tokens of the model's production of bed and bad were holistically synthesized into an 11-step continuum using Akustyk (http://bartus.org/akustyk/).

2.3 Procedure

Participants were seated at a workstation running E-prime Experimental Software and outfitted with a 5-point serial response box. Participants wore AKG headphones and a head-mounted microphone connected to the computer workstation through a USBpre Sound Device.

The task consisted of multiple components:

(1) *Perception pre-test.* Participants were randomly presented with tokens from the *bed-bad* continuum. Each of the eleven tokens from the continuum was presented ten times. Listeners were instructed to categorize each token as *bed* or *bad* by pressing buttons 1 and 5 on the button box, respectively.

(2) *Picture Names*. A picture-naming task was used to elicit participants' baseline productions. To ensure that participants used the correct name for the image (identifying an image of a head of cabbage as *cabbage* and not *lettuce*), the picture naming task was preceded by a nearly identical block which differed only in that it also included an orthographic rendering of the target word with each picture. Participants were not instructed to produce the words aloud when learning the word-picture pairings. Images were presented in the middle of the computer screen with the word presented above in 36 point font. Participants were instructed to learn the word for each picture. The order of presentation of the pictures was fully randomized for each participant.

(3) Production baseline. Participants' baseline productions were recorded through a picture-naming task. The same images presented in the preceding block were presented again in a random order but without the supporting written word. Participants were instructed to identify the picture aloud.

(4) *Model talker exposure.* Participants were presented with the model talker's productions of the same 100 words that appeared in the picture naming block over headphones. One group of

participants was assigned to a Production Group; their exposure to the model talker took the form of an auditory naming task. This group's instructions were to identify (name) the auditory object aloud. Productions were recorded through E-Prime's interface. Upon presentation of the model's production of each word, a recording window of 4 seconds was initiated. The other group of participants were assigned to a Listen Condition. Those in this condition were instructed to listen quietly to the model talker's productions.

(5) *Perception post-test*. Participants will again complete the perceptual identification task presented to them in (1).

Figures 1a and 1b. The probability of classifying each token as *bad* across the baseline and post-task for the No Production group and the Production groups. Listeners' responses are categorical binary values, but the points in the plots have been jittered to show the change in responses across continuum steps.

3. Analysis

3.1 Perceptual recalibration

To eliminate trials where listeners were inattentive or distracted, those trials with response times more than two standard deviations away from the mean were discarded. We then went on to explore perceptual recalibration in two ways. First, the perceptual identification data were scored as *bed* or *bad* responses. A mixed effects regression model predicting the proportion of bad responses was fit with Block, continuum step, and Condition as predictor variables. Random slopes were fit by Subject, with random intercepts for Block. There were main effects of Block [B= 1.76, SE = 0.46, $p < 0.001$] and Step [B = 1.25, SE = 18.92, $p < 0.001$]. The interaction between Block and Step was also significant [B = -0.32, SE = -4.05, $p < 0.001$]. Figures 1a and 1b shows sigmoid functions of the probability of a *bad* response for pre- and post-task tests in the No Production and Production conditions. Listeners' actual responses are shown in the dots which have been jittered to better show the distributions across the continuum steps. While the analysis did not report any interactions with Condition, this figure suggests there is at least a trend towards a larger separation between the pre and post-task responses for the Production condition. This suggests a tendency towards an increase in *bad* responses in the post-task for the Production group who shadowed the model's productions.

A second perceptual analysis examined the crossover point in the continuum where listeners shifted their responses from *bed* to *bad*. A lower crossover point would indicate that more of the continuum was devoted to *bad* responses. The estimated crossover point was linearly interpolated for each listener in each block, and those values were used as the dependent measure in a repeated measures ANOVA. Two listeners' categorization functions had two crossover points such that in the middle of the continuum steps 4 through 6 zigzagged between being below 0.5 and above 0.5, so their data were excluded from this analysis. Block and Condition were used as independent variables with Block repeated across listeners. The interaction between Block and Condition was just beyond significant [$F(1, 25) = 3.6$, $p = 0.069$], and this is shown in Figure 2. Listeners in the Production condition tended to have a lower crossover point in the Post-task identification block.

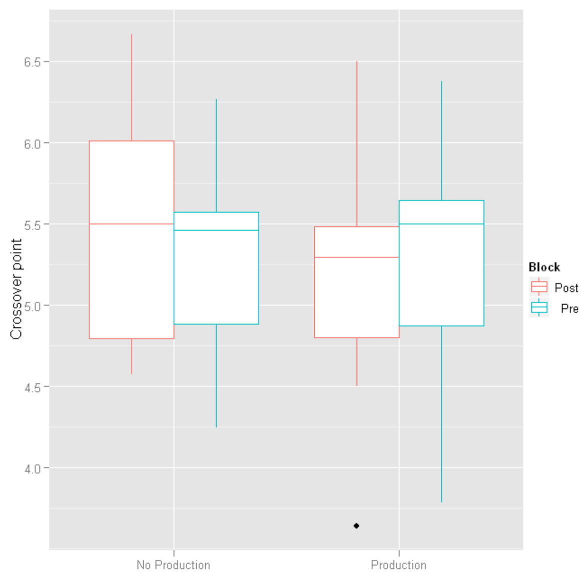


Figure 2. A box-and-whisker plot illustrating point along the bed-bad continuum where listeners' responses crossed over from *bed* to *bad* shown by Condition and Block.

These two perceptual measures illustrate different ways of describing listener behavior, and these measures were highly correlated ($t(52)=-10.29$, $p < 0.001$, Pearson's $r = -0.82$). This means that listeners who has a higher proportion of *bad* responses across the continuum also had a lower crossover point.

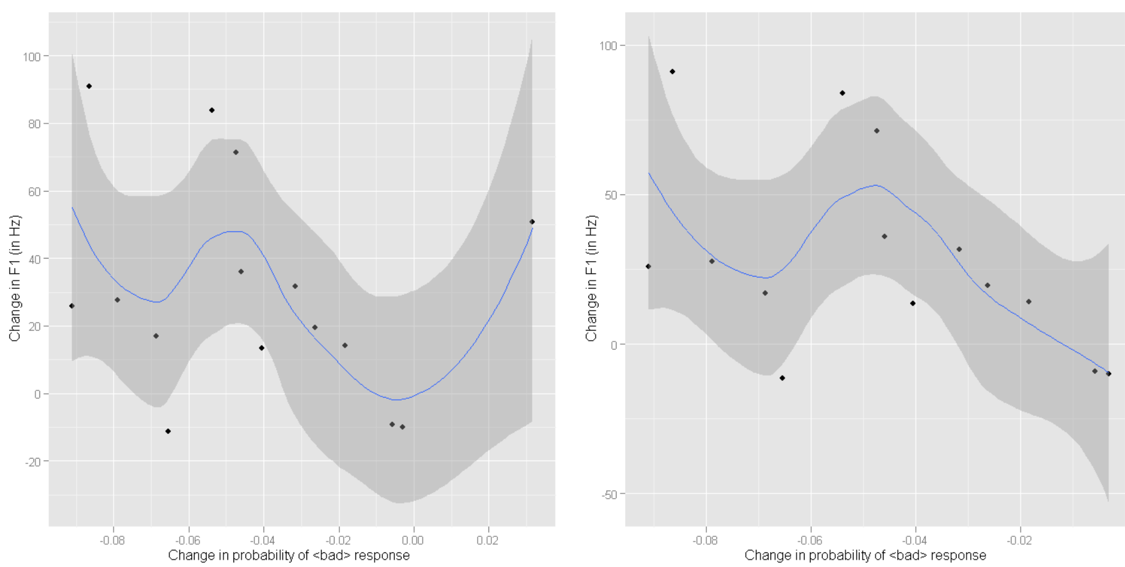
3.2 Shifts in production

Given the known primary difference between /æ/ in the dialect of the model talker, and that of the participants, the production data were simply analyzed in terms of changes in F1. A decrease in F1 from baseline to shadowed productions indicates a shift towards the model talker.

A repeated-measures ANOVA with F1 measures as the dependent measure and Block (baseline vs shadowed) repeated across participants found a main effect of Block [F(1, 14) = 12.04, p = 0.003]. Participants produced /æ/ with a lower F1 when shadowing the model talker (M = 837, SD = 135) compared to their baseline productions (M = 871, SD = 135).

3.3 Perception and production

To assess how speech accommodation relates to perceptual recalibration, the amount of F1 shift was calculated for each individual and their shift in F1 was compared to each of the perceptual recalibration measures. First we compared change in F1 to overall change in to listeners' probability of responding *bad*. Across all participants this was not significant [t(13)=-0.97, p = 0.35, Pearson's r = -0.26]; these data are shown on the left panel of Figure 3 with a best-fit line cutting through the distribution. In this figure, the change in F1 illustrates a negative change in F1, lowering the formant value towards that of the model's dialect. A negative change in probability of bad response is indicative of more *bad* responses in the post-task identification block compared to the pre-task block. If the single participant who showed a shift towards more *bed* responses in the post-task block is removed, the relationship improves [t(12)=1.8, p = 0.096, Pearson's r = 0.46]. There seems to be a tendency for individuals who accommodated more in production to accommodate more in perception, although this relationship was not statistically significant.



Figures 3a and 3b. The best-fit relationship between speech accommodation as change in F1 by the overall change in the probability of a bad response. The figure on the left shows the relationship across all participants in the Production group. The figure on the right presents the same data with one participant removed.

In the second set of analyses we compared overall change in F1to listeners' shifts in the crossover point along the bed-bad continuum. These data for this analysis are shown in Figure 4 where a negative shift in the crossover point indicates a shift towards listeners hearing more of the continuum as *bad*. The relationship between the shift in production and this measure of perceptual recalibration was not significant [t(11) = 1.06, p < 0.31, Pearson's r = 0.30].

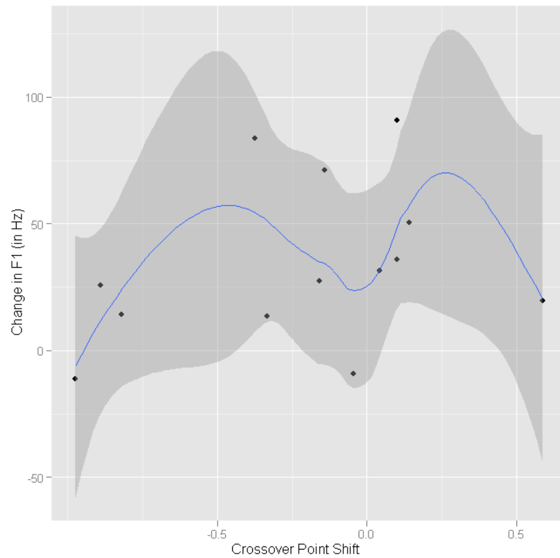


Figure 4. Participants' change in F1 in production and their shift in crossover point along the bed-bad continuum. A negative crossover point means shift to a lower crossover point for bad, indicating more of the continuum was devoted to bad.

4. Discussion and Conclusion

This study examined whether shadowing an unfamiliar dialect affected perceptual recalibration and whether phonetic accommodation is linked to perceptual recalibration. There were two groups in this experiment: a No Production group that simply listened to the model talker's single word productions and a Production group that shadowed the model tokens. The shadowed tokens were compared to baseline productions for the Production group to quantify spontaneous accommodation. Categorization functions for an 11-step bed-bad continuum before and after exposure to the model talker were also compared for both groups. Given previous research, it was predicted that participants in the Production group would shift their /æ/ productions towards that of the model. A long tradition in perceptual recalibration would predict that listeners would modify their perceptual categories; in this case, we predicted an increase in *bad* responses along the bed-bad continuum in response to a model talker whose dialect produces /æ/ words with a lower F1. The goal of this paper was to examine the relationship at the individual level between these shifts in production and perception.

At best, this paper is full of trends and tendencies. Figure 1 seems to suggest that those in the Production group shifted their perceptual categories more from pre- to post-task, but the analysis does not support this interpretation. The analysis of crossover points, depicted in Figure 2, hints at a lower crossover point (more of the continuum devoted to *bad*) in the post-task in the Production group, but, again, this was not significant in the analysis. In line with previous work, those in the Production group did shift their productions of the /æ/ words in the direction of the dialect of the model; their F1s were lower in the shadowed productions compared to their baseline productions. In assessing individual-level relationships between shifts in production and shifts in perception, the patterns were statistically insignificant. Removing a single listener who did not increase her proportion of *bad* responses in the post-task made the relationship between accommodation in production and perceptual recalibration a bit tighter.

These data do not provide definitive evidence for any approach to the relationship between perception and production. They do, however, suggest that 1) individual levels of perceptual adaptation *might* be slightly predictive of phonetic imitation, and 2) producing speech *might* have some subtle effects on perceptual recalibration. These highly tentative results suggest that the connection between perception and production may be more streamlined for established phonetic categories, as opposed to the connection forged when learning new phonetic categories (e.g., Baese-Berk, 2010). Moreover, with respect to sound change, these results indicate that variation in production may be constrained by an individual's level of variation in perception. The short-term nature of the experiment precludes any conclusion about long-term change, yet indicates that short-term shifts can be involuntary.

References

Adank, P., Hagoort, P., & Bekkering, H. (2010). Imitation improves language comprehension. *Psychological Science,* 21 (12), 1903-1909.

Babel, Molly. (2010). Dialect convergence and divergence in New Zealand English. *Language in Society*, 39 (4), 437-456.

Babel, M. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics* 40, pp. 177-189.

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

Baese-Berk, Melissa. (2010). *An Examination of the Relationship Between Perception and Production,* Ph.D. Dissertation, Northwestern University.

Cristià, Alejandrina. (2011). Fine-grained variation in caregivers' /s/ predicts their infants' /s/ category. *Journal of the Acoustical Society of America*, 129, 5, 3271-3280.

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

Eisner, F. & McQueen, J.M. (2005) The specificity of perceptual learning in speech processing. *Perception & Psychophysics*, 67, 224-238.

Evans, Bronwen G. & Paul Iverson. (2007.) Plasticity in vowel perception and production: A study of accent change in young adults. *Journal of the Acoustical Society of America*, 121:3814–3826.

Giles, Howard; Nikolas Coupland; and Justine Coupland. 1991. Accommodation theory: Communication, context, and consequence. In *Contexts of Accommodation: Developments in applied sociolinguistics*, Howard Giles, Justine Coupland, and Nikolas Coupland (eds.), 1-68. Cambridge: Cambridge University Press.

Goldinger, Stephen D. (1997.) Perception and production in an episodic lexicon. In *Talker Variability in Speech Processing*, ed. Keith Johnson and John W. Mullennix, 33–66. Academic Press: San Diego.

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

Goldstein, L. and Fowler, C.A. (2003.) Articulatory phonology: A phonology for public language use. In eds. N. Schiller and A. Meyer *Phonetics and phonology in language comprehension and production: Differences and similarities*. 159-207. Berlin: Mouton de Gruyter.

Ghosh, Satrajit, Melanie Matthies, Edwin Maas, Alexandra Hanson, Mark Tiede, Lucie Ménard, Frank Guenther, Harlan Lane, and Joseph Perkell. (2010). An investigation of the relation between sibilant production and somatosensory and auditory acuity. *Journal of the Acoustical Society of America*, 128, 5, 3079-3087.

Kim, M., Horton, W.S., & Bradlow, A. R. (2011). Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Journal of Laboratory Phonology*, 2, 125-156.

Kraljic, T., Brennan, S.E., and Samuel, A.G.(2008.) Accommodating variation: Dialects, idiolects, and speech processing. *Cognition* 107:54–81.

Kraljic, T. and Samuel, A.G. (2005.) Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology*. 51:141–178.

Kraljic, T. and Samuel, A.G. (2006.) Generalization in perceptual learning for speech.

Psychonomic Bulletin Review 13:262–268.
Kraljic, T. and Samuel, A.G. (2007.) Perceptual adjustments to multiple speakers. *Journal of Memory and Language* 56:1–15.
Kraljic, T., Samuel, A.G., and Brennan, S.E.(2008.) First impressions and last resorts: How listeners adjust to speaker variability. *Psychological Science*. 19:332–338.
Maye, J., Aslin, R.N., and Tanenhaus, M.K. (2008.) The weckud wetch of the wast: Lexical adaptation to a novel accent. *Cognitive Science*. 32:543–562.
Miller, R., Sanchez, K., & Rosenblum, L.(2010.) Alignment to visual speech information. *Attention, Perception, & Psychophysics*, 72 (6), 1614-1625.
Mitterer, H. & Ernestus, M. (2008). The link between speech perception and production is phonological and abstract: Evidence from a shadowing task. *Cognition*, 109, 168–173.
Namy, Laura L., Lynne C. Nygaard, & Denise Sauerteig.(2002.) Gender differences in vocal accommodation: the role of perception. *Journal of Language and Social Psychology*, 21:422–432.
Nielsen, K. (2011). Specificity and abstractness of VOT imitation. *Journal of Phonetics*, 39, 132–142.
Norris, D., McQueen, J.M., and Cutler, A. 2003. Perceptual learning in speech. *Cognitive Psychology* 47:204–238.
Pardo, Jennifer S. (2006). On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America*, 119:2382–2393.
Pardo, J. S. 2009. Acoustic-phonetic variation and conversational interaction. In *Expressing Oneself/Expressing One's self: A Festschrift in Honor of Robert M. Krauss*, ed. E. Morsella.
Pardo, Jennifer S.; Isabel Cajori Jay; and Robert M. Krauss. 2010. Conversational role influences speech imitation. *Attention, Perception, & Psychophysics* 72(8).2254-2264.
Perkell, Joseph, Frank Guenther, Harlan Lane, Melanie Matthies, Ellen Stockmann, Mark Tiede, & Majid Zandipour. (2004). The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts. *Journal of the Acoustical Society of America*, 116, 4, 2338-2344.
Pickering, M. J., & Ferreira, V.S. (2008). Structural priming: A critical review. *Psychological Bulletin*, 134, 427–459.
Pickering, M.J. & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioural & Brain Sciences*, 27, 169–226.
Schneider, Walter, Amy Eschman, & Anthony Zuccolotto. (2007). *E-Prime: User's Guide, version 2.0.* Psychology Software Tools.
Shiller, D., M. Sato, V.L. Gracco, S.R. Baum. (2009). Perceptual recalibration of speech sounds following speech motor learning. *Journal of the Acoustical Society of America*, 125 (2), 1103-1113.
Shockley, Kevin, Laura Sabadini, & Carol A. Fowler. (2004). Imitation in shadowing words. *Perception & Psychophysics*, 66:422–429.
Sonderegger, M. (2012). Phonetic and Phonological Dynamics on Reality Television. PhD Dissertation, University of Chicago.
Trudgill, P. (1986). *Dialects in contact.* New York: Blackwell Publishing.
Trudgill, P. (2008). Colonial dialect contact in the history of European Languages: On the irrelevance of identity in new-dialect formation. *Language in Society*, 37, 241–280.