

Lexically guided perceptual tuning of internal phonetic category structure

Julia R. Drouin, Rachel M. Theodore, and Emily B. MyersDDO

Citation: *The Journal of the Acoustical Society of America* **140**, EL307 (2016); doi: 10.1121/1.4964468

View online: <http://dx.doi.org/10.1121/1.4964468>

View Table of Contents: <http://asa.scitation.org/toc/jas/140/4>

Published by the [Acoustical Society of America](#)

Articles you may be interested in

[Comparison of perception-production vowel spaces for speakers of Standard Modern Greek and two regional dialects](#)

The Journal of the Acoustical Society of America **140**, EL314 (2016); 10.1121/1.4964397

[The bursts of stops can convey dialectal information](#)

The Journal of the Acoustical Society of America **140**, EL334 (2016); 10.1121/1.4964818

[Evaluation of smartphone sound measurement applications \(apps\) using external microphones—A follow-up study](#)

The Journal of the Acoustical Society of America **140**, EL327 (2016); 10.1121/1.4964639

[Do talkers produce less dispersed phoneme categories in a clear speaking style?](#)

The Journal of the Acoustical Society of America **140**, EL320 (2016); 10.1121/1.4964815

[An evaluation of the perceptual quality of phase-aware single-channel speech enhancement](#)

The Journal of the Acoustical Society of America **140**, EL364 (2016); 10.1121/1.4965288

[Wave propagation analysis in spiral geometries](#)

The Journal of the Acoustical Society of America **140**, EL291 (2016); 10.1121/1.4964372

Lexically guided perceptual tuning of internal phonetic category structure

Julia R. Drouin,^{a)} Rachel M. Theodore,^{a)} and Emily B. Myers^{b)}

Department of Speech, Language, and Hearing Sciences, University of Connecticut,
850 Bolton Road, Storrs, Connecticut 06269, USA
julia.drouin@uconn.edu, rachel.theodore@uconn.edu, emily.myers@uconn.edu

Abstract: Listeners use lexical information to retune the mapping between the acoustic signal and speech sound representations, resulting in changes to phonetic category boundaries. Other research shows that phonetic categories have a rich internal structure; within-category variation is represented in a graded fashion. The current work examined whether lexically informed perceptual learning promotes a comprehensive reorganization of internal category structure. The results showed a reorganization of internal structure for one but not both of the examined categories, which may reflect an attenuation of learning for distributions with extensive category overlap. This finding points towards potential input-driven constraints on lexically guided phonetic retuning.

© 2016 Acoustical Society of America

[DDO]

Date Received: April 26, 2016 Date Accepted: September 23, 2016

1. Introduction

Listeners map the acoustic signal to speech sound representations despite wide variability in their acoustic instantiation. Lexical information is one factor that facilitates this mapping process. Local lexical context can be used to resolve ambiguity such that a phonetic token that is ambiguous between /g/ and /k/ will be perceived as /g/ in the context of the frame /ift/ but /k/ in the context of the frame /is/ (e.g., Ganong, 1980). Strikingly, lexical influences on speech sound categorization are not limited to online processing; rather, lexical influences lead to long-term perceptual learning for speech sounds (e.g., Norris *et al.*, 2003). Kraljic and Samuel (2005) exposed listeners to an ambiguous phoneme midway between /s/ and /ʃ/ during a lexical decision task. For some listeners, the ambiguous phoneme replaced medial /s/ in words like *dinosaur*; for others, the same ambiguous phoneme replaced medial /ʃ/ in words like *publisher*. Following exposure, all listeners categorized items along an /ɑʃi/-/asi/ continuum. The results showed that listeners who were biased to perceive the ambiguity as /s/ showed more /asi/ responses compared to listeners who were biased to perceive the ambiguity as /ʃ/, suggesting that the ambiguous sound was incorporated into the category consistent with prior lexical context. Many studies have examined the outcome of this type of perceptual learning and have found that learning is long-lasting (Kraljic and Samuel, 2005; Eisner and McQueen, 2006), is sometimes talker-specific (Eisner and McQueen, 2005; Kraljic and Samuel, 2005; Kraljic and Samuel, 2007), and generalizes to novel utterances (Kraljic and Samuel, 2006).

Lexically informed perceptual learning has been measured primarily in terms of changes to phonetic category membership. However, it has long been known that speech sound categories have a graded internal structure, with some members of the category represented as better exemplars than others. Even though many tokens may be considered part of the same phonetic category, some may be weighted as more prototypical than others as reflected by ratings in a goodness judgment task (Miller, 1994) or by reaction times in identification tasks (Samuel and Kat, 1996). Moreover, internal category structure is functionally plastic; which members are considered most prototypical robustly shifts as a function of many factors including speaking rate (Volaitis and Miller, 1992), place of articulation (Volaitis and Miller, 1992), and a talker's idiolect (Theodore *et al.*, 2015). Importantly, internal category structure shifts do not necessarily parallel changes in category boundary; although lexical status expands the range of acoustic-phonetic space that is mapped to a phonetic category (e.g., Ganong, 1980), it does *not* result in a concomitant change in perceived category goodness (Allen and

^{a)}Also at: Connecticut Institute for the Brain and Cognitive Sciences, University of Connecticut, 337 Mansfield Road, Storrs, CT 06269, USA.

^{b)}Also at: Department of Psychological Sciences, University of Connecticut, 406 Babbidge Road, Storrs, CT, 06269, USA.

Miller, 2001). Of critical interest is why some factors induce changes in category goodness while others do not. One explanation hinges on the observation that lexical status does not *de facto* lead to contextual variation in speech production, whereas factors such as speaking rate and talker idiolect do (e.g., Volaitis and Miller, 1992; cf. Baese-Berk and Goldrick, 2009). This production-based account makes the prediction that in situations where lexical status is linked to systematic changes to the speech signal—as in the case of lexically informed perceptual learning for speech—listeners will show sensitivity to this contextual influence with respect to both the category boundary and internal category structure. In the context of previous research, the novelty of the present study is to examine outcomes of lexically informed perceptual learning in terms of structure within a phonetic category in addition to boundary regions between phonetic categories.

Two experiments were conducted. In each, two groups of listeners completed a lexical decision exposure task; during this exposure, one group was biased to perceive an ambiguous fricative as /ʃ/ and the other group was biased to perceive the same ambiguity as /s/. After exposure, both groups completed a category goodness test to assess changes to internal category structure (/ʃ/ in experiment 1 and /s/ in experiment 2) and a category identification test to assess changes to the /ʃ-/s/ boundary. If lexically informed perceptual learning promotes a comprehensive reorganization of phonetic category structure, then both the identification responses and goodness judgments will reflect previous lexical exposure. A failure to observe changes to internal category structure would suggest that the representational influence of this learning mechanism is constrained to the boundary region.

2. Experiment 1

2.1 Methods

Participants included twenty-four listeners between the ages of 18 and 31 years ($M = 21.2$, $SD = 3.58$). All were monolingual speakers of American English, reported no history of language or hearing disorders, and passed a pure tone hearing screen (administered at 20 dB for octave frequencies between 500 and 4000 Hz). Participants were randomly assigned to either the /ʃ/-biasing condition ($n = 12$) or the /s/-biasing condition ($n = 12$).

The stimulus set is the same as that used in Myers and Mesite (2014), to which the reader is referred for a complete description of the methods for stimulus construction. In brief, the exposure stimuli consisted of 100 auditory words and 100 auditory nonwords produced by a native female speaker of American English. The 100 auditory words were divided into three classes: 20 /s/ words (e.g., *pencil*), 20 /ʃ/ words (e.g., *ambition*), and 60 filler words that contained no instance of an /s/ or /ʃ/ (e.g., *napkin*). Two versions of the /s/ and /ʃ/ items were used—one that was the natural production and one that was modified to replace the fricative in the natural production with an ambiguous fricative. Ambiguous tokens were created by averaging amplitude-normalized fricative segments drawn from the same phonological environment (e.g., *pencil* and *penshil*), normalizing duration to the shorter of the two fricatives, and then re-inserting the ambiguous fricative into the /s/-consistent frame (e.g., *pencil*). The exposure stimuli were arranged into two sets, one for the /ʃ/-biasing group (20 natural /s/ words, 20 modified /ʃ/ words) and one for the /s/-biasing group (20 modified /s/ words, the 20 natural /ʃ/ words). Stimulus sets for both groups were completed with 60 filler words and 100 nonwords, for a total of 200 items in each set. The test stimuli consisted of a six-step continuum ranging from a clear /ɑʃi/ token to a clear /asi/ token spoken by the talker used for the training stimuli. The test continuum was created following the methods outlined for the ambiguous training stimuli. Waveform averaging was used to create six unique fricative blends; in terms of percent /s/ energy, these blends were 30%, 40%, 50%, 60%, 70%, and 80%. These six test tokens were used for both training groups in both the goodness judgment and category identification test tasks.

All participants completed an exposure phase followed by two test phases, a goodness judgment task to assess changes to the internal category structure and an identification test to assess changes to the category boundary. Order of the test tasks was fixed, with the goodness judgment task preceding the identification task. Testing took place individually in a sound-attenuated booth. Stimuli were presented via headphones at a comfortable listening level that was constant across participants. Responses were made using a button box. Button assignments to YES/NO responses (during exposure) and the ASI/ASHI responses (during the identification task) were counterbalanced for dominant hand.

During exposure, participants heard one randomization of the 200 word and nonword items appropriate for their biasing group and indicated whether each item was a word. For the goodness judgment task, six randomizations of the six test stimuli were presented and participants were directed to listen to the middle sound and rate how good of an exemplar that sound was as a member of the /*f*/ category. Responses were made using a 1–7 Likert scale, with 7 indicating the best exemplar of the /*f*/ category. Participants were encouraged to use the full range of the scale. For the category identification test, participants heard six randomizations of the six steps of the test continuum and identified each token as either /*asi*/ or /*afi*/. No feedback was provided either during exposure or during the two test tasks, and ISI was constant at 2000 ms (timed from the participant's response). The procedure lasted approximately 30 min.

2.2 Results

Performance during training was measured in terms of percent correct lexical decision responses and was calculated separately for each training group and for each item type. As shown in Table 1, performance for both groups was near ceiling across the four item types, as expected based on previous research (e.g., Norris et al., 2003).

Test performance was measured separately for the identification and goodness rating test tasks. Consider first the identification task. For each participant, mean percent /*s*/ responses were calculated for each step of the test continuum. Figure 1(b) shows mean /*s*/ responses for the two training groups. As expected, the categorization functions are displaced; listeners who were biased to perceive the ambiguous fricative as /*s*/ show more /*s*/ responses compared to listeners who were biased to perceive the ambiguity as /*f*/. Mean percent /*s*/ responses was submitted to analysis of variance (ANOVA) with the between-subjects factor of training bias and the within-subjects factor of continuum step. The results showed a main effect of continuum step ($F_{5,110} = 98.430$, $p < 0.001$, $\eta_p^2 = 0.817$), reflecting more /*s*/ responses for both groups of listeners towards the /*s*/ end of the test continuum compared to the /*f*/ end of the continuum. Critically, the ANOVA confirmed a main effect of training bias ($F_{1,22} = 6.229$, $p = 0.021$, $\eta_p^2 = 0.221$), indicating that there were more /*s*/ responses in the /*s*-biasing group compared to the /*f*-biasing group. The ANOVA also revealed an interaction between continuum step and training bias ($F_{5,110} = 2.755$, $p = 0.022$, $\eta_p^2 = 0.111$), indicating that the magnitude of the learning effect differed along the test continuum.

For the goodness rating task, we calculated mean goodness as /*f*/ for each step of the continuum; mean performance between the two biasing groups is shown in Fig. 1(a). Visual inspection suggests that exposure during training did indeed influence perceived prototypicality within the /*f*/ category; listeners in the /*f*-biasing condition rated the test tokens as better exemplars compared to listeners in the /*s*-biasing condition, particularly for the continuum steps near the /*afi*/ endpoint. Mean goodness ratings were submitted to ANOVA with the factors of training bias and continuum step. The ANOVA revealed a main effect of continuum step ($F_{5,110} = 46.427$, $p < 0.001$, $\eta_p^2 = 0.678$), indicating that goodness ratings decreased towards the /*s*/ end of the test continuum for both training groups. The ANOVA showed no main effect of training bias ($F_{1,22} = 2.706$, $p = 0.114$, $\eta_p^2 = 0.110$), but did reveal a significant interaction between training bias and continuum step ($F_{5,110} = 2.568$, $p = 0.031$, $\eta_p^2 = 0.105$). Though ANOVA is robust to normality assumptions, we used the Mann-Whitney test to confirm that group differences in Likert ratings were also observed when using non-parametric statistics at the 30%, 40%, and 50% /*s*/ steps ($U < 37$, $p < 0.05$ in all cases).

Collectively, the results from experiment 1 indicate that the lexical information provided during the exposure phase resulted in comprehensive perceptual learning; learning was reflected not only with a change in the mapping of acoustic variants to

Table 1. Mean percent correct and standard deviation (in parentheses) for performance during the lexical decision training task for each item type.

Experiment	Item type				
	Training group	Ambiguous	Clear	Filler	Nonwords
1	/i/-bias	95.58 (5.50)	99.17 (2.89)	94.16 (3.35)	94.35 (3.81)
	/s/-bias	96.94 (4.66)	98.33 (3.26)	93.69 (3.12)	94.67 (2.99)
2	/i/-bias	94.16 (5.21)	98.32 (4.06)	94.07 (3.70)	92.73 (5.89)
	/s/-bias	95.10 (7.07)	96.46 (4.67)	92.33 (6.29)	93.84 (3.95)

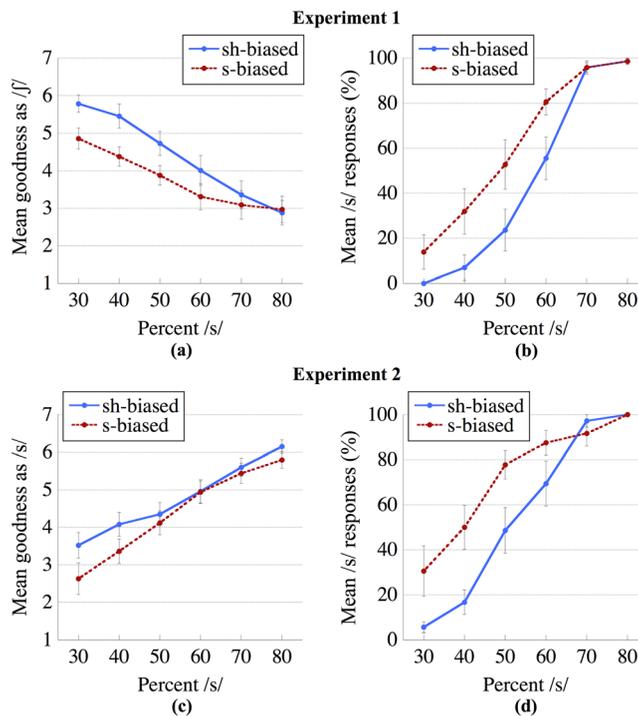


Fig. 1. (Color online) Mean goodness as /f/ ratings are shown in (a) and mean /s/ responses are shown in (b) for experiment 1. (c) Mean goodness as /s/ ratings and (d) mean /s/ responses for experiment 2. Error bars indicate standard error of the mean.

the /f/ and /s/ categories, in line with previous studies (i.e., Myers and Mesite, 2014), but also by a reorganization of exemplar space within the /f/ category.

3. Experiment 2

3.1 Methods

Twenty-four additional listeners were recruited for experiment 2. All participants met the demographic criteria outlined above and were between the ages of 18 and 30 years ($M=20.5$, $SD=2.75$). Half were assigned to the /f/-biasing condition and the other half were assigned to the /s/-biasing condition. The stimuli and procedures described in experiment 1 were used here with one exception; during the goodness judgment task, participants were directed to rate the middle sound in each test token for goodness as a member of the /s/ category.

3.2 Results

Performance during training and test was measured as outlined for experiment 1. As shown in Table 1, both groups showed near ceiling performance during training for all item types. Inspection of Fig. 1(d) suggests that exposure during the training phase did indeed yield perceptual learning such that in the identification test, more tokens were identified as /s/ for listeners in the /s/-biasing group compared to the /f/-biasing group. However, it appears that learning was limited to changes in the category boundary; no systematic differences between the two training groups are seen for the goodness judgments in Fig. 1(c).

Mean percent /s/ responses and mean goodness as /s/ ratings were analyzed in separate ANOVAs. For the identification responses, the ANOVA showed a main effect of continuum step ($F_{5,110} = 64.029$, $p < 0.001$, $\eta_p^2 = 0.744$), a main effect of training bias ($F_{1,22} = 7.350$, $p = 0.013$, $\eta_p^2 = 0.250$), and a significant interaction between continuum step and training bias ($F_{5,110} = 3.729$, $p = 0.004$, $\eta_p^2 = 0.145$). For the goodness judgment responses, we again observed a main effect of continuum step ($F_{5,110} = 55.382$, $p < 0.001$, $\eta_p^2 = 0.716$), indicating that goodness ratings increased as the continuum became more /s/-like. However, the ANOVA showed no main effect of training bias ($F_{1,22} = 1.422$, $p = 0.246$, $\eta_p^2 = 0.061$) nor an interaction between training bias and continuum step ($F_{5,110} = 1.330$, $p = 0.257$, $\eta_p^2 = 0.057$), indicating that mean goodness as /s/ judgments did not differ between the two training groups. When taken together, the results from the identification and goodness rating tasks suggest that perceptual learning for the /s/ category was constrained compared to the /f/ category examined in

experiment 1; specifically, exposure during training led to a robust change in the mapping to the /ʃ/ and /s/ categories, but there was no evidence that learning resulted in a concomitant change to the internal category structure of the /s/ category.

4. Discussion

Here we examined whether lexically informed perceptual learning leads to changes to the internal structure of native speech sound categories or whether learning was limited to the category boundary region. Robust changes to the boundary between /s/ and /ʃ/ were observed in both experiments; however, we only observed a change to internal category structure for the /ʃ/ category. We consider three explanations for this difference, noting that the current sample size may be a potential limitation. One explanation is that coarticulatory cues influenced the goodness as /s/ judgments to a greater degree than did perceptual learning. Recall that the test stimuli were created by placing the modified fricatives into the /s/-frame; thus, the coarticulatory information of the test continuum was consistent with a medial /s/, which may have led to equivalent goodness as /s/ judgments for both biasing groups, consistent with findings demonstrating that this type of learning mechanism can be constrained by coarticulatory cues (Stevens *et al.*, 2007). Another explanation concerns the possibility of asymmetrical learning effects. In the current study, perceptual learning is measured by comparing performance between the two biasing groups following exposure; however, an alternative is to consider the performance of each biasing group compared to a no-exposure baseline condition. Previous examinations using this latter approach have found evidence of asymmetric learning such that one exposure condition differs from a no-exposure baseline but the other does not (e.g., Zhang and Samuel, 2014). The differential influences of learning on the categories examined here would be expected if the current stimulus set promotes recalibration for /ʃ/ but not /s/. We also considered a third explanation, that being differences in the precise acoustic input presented during the exposure phase. Consistency in speech production influences listeners' ability to categorize speech sounds (e.g., Newman *et al.*, 2001). Listeners show less stable identification responses for talkers with highly variable productions compared to talkers who are more consistent and who have minimal category overlap in their productions (Newman *et al.*, 2001; Clayards *et al.*, 2008). Given the methods used to create the current stimulus set, it may have been the case that the distributional information presented for the /s/ and /ʃ/ categories between the two training groups differed in ways that influenced perceptual learning.

We examined this possibility through acoustic analyses on the fricatives from the critical training items using a script to extract center of gravity (Elvira-Garcia, 2015), which provides a weighted average of spectral energy in each fricative and systematically differs between the /s/ and /ʃ/ categories (Jongman *et al.*, 2000). Figure 2(a) shows histograms for the /s/ and /ʃ/ distributions presented during the /s/-biasing (top panel) and /ʃ/-biasing (bottom panel) exposure phase. The center of gravity for the six test tokens is plotted in black at the top of each histogram. In the /ʃ/-biasing condition, productions for /ʃ/ and /s/ fall into two distinct distributions that are far apart in acoustic-phonetic space and collectively span the range of frequency present in the test tokens. However, for the /s/-biasing condition, the two distributions are relatively closer together, show more overlap, and are clustered near the /aʃi/ end of the test continuum. Figure 2(b) shows the same distributions, but plotted with respect to the difference between the natural (i.e., unambiguous) productions and the modified versions that were used during the learning phase. That is, Fig. 2(a) shows the distributions as provided as input to each training group, but Fig. 2(b) shows the distributions with respect to how deviant the modified (i.e., ambiguous) tokens were relative to natural productions. When compared to the /s/ tokens, the difference between the natural and modified /ʃ/ tokens is quite slight, suggesting that the ambiguous tokens presented during training to the /ʃ/-biasing group were *a priori* better exemplars. This is in stark contrast to the modified /s/ tokens, which show values closer to those typical of the unmodified /ʃ/ category. Collectively, these analyses show that compared to the /ʃ/-biasing group, the stimuli presented to the /s/-biasing group contained distributions with increased category overlap and poorer exemplars of /s/, either of which may have attenuated listeners' ability to incorporate the modified tokens as good exemplars of the /s/ category. Additional examination is needed to experimentally confirm this account.

The current experiments demonstrate that lexically guided perceptual tuning can lead to a comprehensive restructuring of phonetic category space, including changes to perceived category typicality, but point to potential constraints on the degree to which internal category structure is modified. Namely, these findings suggest

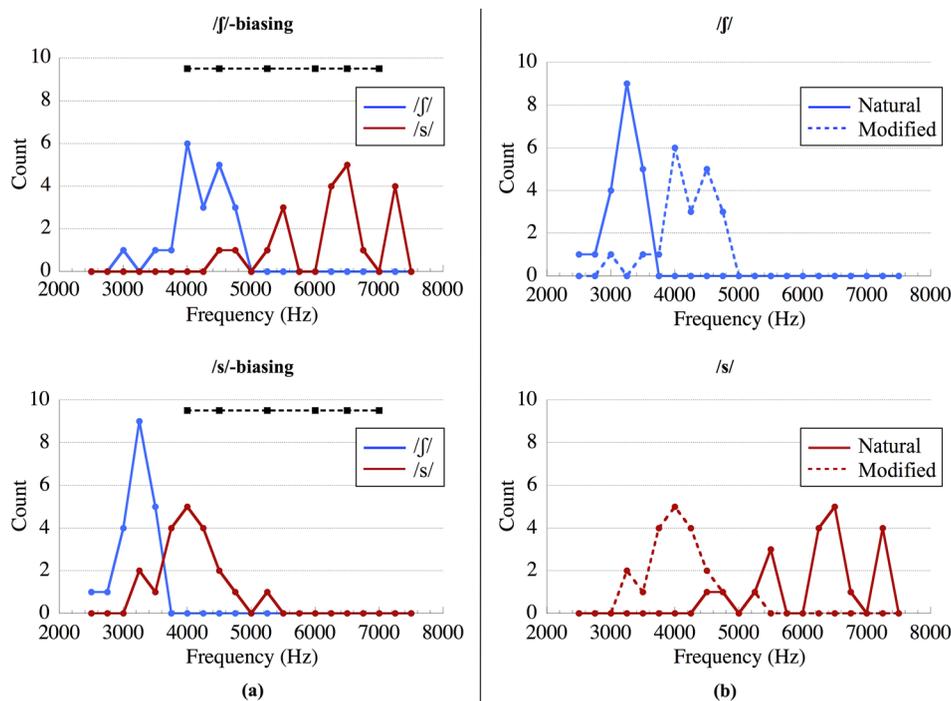


Fig. 2. (Color online) (a) Histograms for the center of gravity (in Hz) of the /f/ and /s/ tokens presented to the /f/-biasing and /s/-biasing training groups. The square black points at the top of each histogram indicate center of gravity for the six test tokens. (b) The same distributions plotted in terms of the natural versus modified versions of the tokens in each fricative category.

that reorganization of internal phonetic category space is limited by the available information in the bottom-up signal—that is, the degree to which distributions of acoustic tokens are separable (McMurray *et al.*, 2009). Future work is directed at examining how, and with what weight, bottom-up acoustic information and top-down lexical information combine to influence perceptual learning for speech. Such an approach has potential to provide an explanation for previously reported asymmetries in the perceptual learning literature (e.g., Zhang and Samuel, 2014; Eisner and McQueen, 2005) and will advance a theoretical understanding of how listeners adapt to systematic variation in the speech signal while maintaining representational stability for spoken language processing.

Acknowledgments

Research was supported by start-up funds to R.M.T. (College of Liberal Arts and Sciences, University of Connecticut) and by the National Science Foundation (NSF Grant No. DGE 1144399; Dr. James Magnuson, PI). We extend gratitude to Ashley Bean, Maria Marotti, Myles Mocarski, and Kayla Marcinczyk for their assistance with data collection.

References and links

- Allen, J. S., and Miller, J. L. (2001). "Contextual influences on the internal structure of phonetic categories: A distinction between lexical status and speaking rate," *Percept. Psychophys.* **63**(5), 798–810.
- Baese-Berk, M., and Goldrick, M. (2009). "Mechanisms of interaction in speech production," *Lang. Cogn. Process.* **24**(4), 527–554.
- Clayards, M., Tanenhaus, M. K., Aslin, R. N., and Jacobs, R. A. (2008). "Perception of speech reflects optimal use of probabilistic speech cues," *Cognition* **108**(3), 804–809.
- Eisner, F., and McQueen, J. M. (2005). "The specificity of perceptual learning in speech processing," *Percept. Psychophys.* **67**(2), 224–238.
- Eisner, F., and McQueen, J. M. (2006). "Perceptual learning in speech: Stability over time," *J. Acoust. Soc. Am.* **119**(4), 1950–1953.
- Elvira-García, W. (2015). "Zero crossings and spectral moments," retrieved from <http://stel.uconn.edu/labfon/sites/default/files/zero-crossing-and-spectral-moments13.praat> (Last viewed October 7, 2016).
- Ganong, W. F. (1980). "Phonetic categorization in auditory word perception," *J. Exp. Psych.: Hum. Percept. Perf.* **6**(1), 110–115.
- Jongman, A., Wayland, R., and Wong, S. (2000). "Acoustic characteristics of English fricatives," *J. Acoust. Soc. Am.* **108**(3), 1252–1263.
- Kraljic, T., and Samuel, A. G. (2005). "Perceptual learning for speech: Is there a return to normal?," *Cogn. Psych.* **51**(2), 141–178.

- Kraljic, T., and Samuel, A. G. (2006). "Generalization in perceptual learning for speech," *Psych. Bull. Rev.* **13**(2), 262–268.
- Kraljic, T., and Samuel, A. G. (2007). "Perceptual adjustments to multiple speakers," *J. Mem. Lang.* **56**(1), 1–15.
- McMurray, B., Aslin, R. N., and Toscano, J. C. (2009). "Statistical learning of phonetic categories: Insights from a computational approach," *Dev. Sci.* **12**(3), 369–378.
- Miller, J. L. (1994). "On the internal structure of phonetic categories: A progress report," *Cognition* **50**(1–3), 271–285.
- Myers, E. B., and Mesite, L. M. (2014). "Neural systems underlying perceptual adjustments to non-standard speech tokens," *J. Mem. Lang.* **76**, 80–93.
- Newman, R. S., Clouse, S. A., and Burnham, J. L. (2001). "The perceptual consequences of within-talker variability in fricative production," *J. Acoust. Soc. Am.* **109**(3), 1181–1196.
- Norris, D., McQueen, J. M., and Cutler, A. (2003). "Perceptual learning in speech," *Cogn. Psych.* **47**(2), 204–238.
- Samuel, A. G., and Kat, D. (1996). "Early levels of analysis of speech," *J. Exp. Psych.: Hum. Perc. and Perf.* **22**(3), 676–694.
- Stevens, M. A., McQueen, J. M., and Hartsuiker, R. J. (2007). "No lexically-driven perceptual adjustments of the [x]-[h] boundary," in *16th International Congress of Phonetics Sciences (ICPhS 2007)*, pp. 1897–1900.
- Theodore, R. M., Myers, E. B., and Lomibao, J. A. (2015). "Talker-specific influences on phonetic category structure," *J. Acoust. Soc. Am.* **138**, 1068–1078.
- Volaitis, L. E., and Miller, J. L. (1992). "Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories," *J. Acoust. Soc. Am.* **92**(2), 723–735.
- Zhang, X., and Samuel, A. G. (2014). "Perceptual learning of speech under optimal and adverse conditions," *J. Exp. Psych.: Hum. Perc. and Perf.* **40**(1), 200–217.