

7-1-2017

Comparing malleability of phonetic category between [i] and [u]

Reiko Kataoka
San Jose State University

Hahn Koo
San Jose State University, hahn.koo@sjsu.edu

Follow this and additional works at: https://scholarworks.sjsu.edu/linguistics_pub



Part of the [Phonetics and Phonology Commons](#)

Recommended Citation

Reiko Kataoka and Hahn Koo. "Comparing malleability of phonetic category between [i] and [u]" *Journal of the Acoustical Society of America* (2017): EL42-EL48. <https://doi.org/10.1121/1.4986422>

This Article is brought to you for free and open access by the Linguistics and Language Development at SJSU ScholarWorks. It has been accepted for inclusion in Faculty Publications by an authorized administrator of SJSU ScholarWorks. For more information, please contact scholarworks@sjsu.edu.

Comparing malleability of phonetic category between [i] and [u]

Reiko Kataoka^{a)} and Hahn Koo

Department of Linguistics and Language Development, San José State University,
One Washington Square, San José, California 95192-0093, USA
reiko.kataoka@sjsu.edu, hahn.koo@sjsu.edu

Abstract: This study reports differential category retuning effect between [i] and [u]. Two groups of American listeners were exposed to ambiguous vowels ([i/u]) within words that index a phoneme /i/ (e.g., *athl[ilu]t*) (i-group) or /u/ (e.g., *aftern[ilu]n*) (u-group). Before and after the exposure these listeners categorized sounds from a [bip]-[bup] continuum. The i-group significantly increased /bip/ responses after exposure, but the u-group did not change their responses significantly. These results suggest that the way mental representation handles phonetic variation may influence malleability of each category, highlighting the complex relationship among distribution of sounds, their mental representation, and speech perception.

© 2017 Acoustical Society of America

[DDO]

Date Received: March 14, 2017 Date Accepted: June 3, 2017

1. Introduction

Natural speech sounds exhibit a wide range of variation, thus understanding how listeners adeptly interpret them into stable linguistic codes has been a central issue in speech perception research. One approach to this question has been to examine the role of phonetic categories in speech perception. Previous studies (e.g., [Lisker and Abramson, 1970](#)) have shown that category boundaries found in categorization tasks along a sound continuum (e.g., a VOT continuum from voiceless to voiced stops) reflect the range of acoustic variation found in listener's native language. Such findings suggest that phonetic categories reflect distributional property of ambient speech sounds, and these categories support stable speech perception. Previous studies also show that these phonetic categories remain malleable throughout adulthood so that listeners can adapt to on-going sound changes (e.g., [Harrington et al., 2000](#)) and newly encountered pronunciation patterns (e.g., [Norris et al., 2003](#)).

An open question is whether all phonetic categories are equally malleable or if there is a systematic difference in their malleability that is attributable to constraints of the speech perception system. Here, we tested a hypothesis that the more variable speech sounds are, the less malleable their categories would be. Our hypothesis is motivated by recent studies suggesting that adjustability of phonetic categories may vary depending on structural properties of categories such as density ([Scharenborg and Janse, 2013](#)) and variability of categories ([Stevens et al., 2007](#)). For example, [Stevens et al. \(2007\)](#) reported lack of adjustability of category boundary between [x] and [h] for Dutch listeners, and attributed this result to susceptibility of [h] to coarticulatory variation, which would make the contrast between [x] and [h] rather unclear.

While the above result may indicate lack of boundary adjustment *between* [x] and [h], another possible interpretation is that the listeners resisted retuning *both* [x] and [h] categories independently. Similar to /h/, phonetic forms of velar consonants such as /x/ have been suggested to vary considerably depending on vowel contexts at the articulatory as well as acoustic levels. For example, both in Swedish ([Öhman, 1966](#)) and American English ([Kent and Moll, 1972](#)), [g] is articulated with varying tongue positions along the horizontal axis in symmetrical VCV sequences (e.g., [igi]). Its place of constriction is more front adjacent to [i] and more back adjacent to [u]. As an acoustic consequence, the F2 transition terminal frequency, which characterizes the consonant's place of articulation, also varies from high to low values depending on the vowel context ([Öhman, 1966](#)).

^{a)} Author to whom correspondence should be addressed.

We further examined the potential relationship between category variability and malleability by comparing American English vowels [i] and [u], which previous studies have shown to have different degrees of variability. First, the extent of consonant-to-vowel coarticulatory influence in terms of F2 variability is much greater in [u] than [i] (Stevens *et al.*, 1966) because the coarticulation mainly affects the front part of the tongue (MacNeilage and DeClerk, 1969), changing F2 more in back vowels than front vowels. Second, incomplete lip rounding with a nearby consonant would shorten front cavity and raise F2 (Stevens *et al.*, 1966), affecting [u] but not [i]. Third, there is an on-going sound change of back vowel fronting in many parts of the US including California (e.g., Labov *et al.*, 2006). Finally, previous studies have also shown that there is more individual variation in F2 in [u] than in [i] when produced in [hVd] syllables (e.g., Hillenbrand *et al.*, 2001). If category malleability is related to variability, then we expect [i] to be more malleable than [u].

We tested our prediction using a lexically guided perceptual learning paradigm¹ following Norris *et al.* (2003). We put listeners into two groups (i-group and u-group) and let them learn to map a sound that is ambiguous between [i] and [u] (henceforth [i/u]) to either vowel by presenting it within English words, which reveal the phonemic identity of the sound (/i/ for the i-group or /u/ for the u-group), allowing the listeners to learn to accept the new sound [i/u] as an instance of [i] or [u] depending on their group. The effect of learning was assessed by measuring changes in how listeners categorize different sounds along the [i]-[u] continuum before and after the exposure. If [u] is indeed less malleable than [i], we should see a weaker change in categorization responses in the u-group than the i-group.

2. Methods

2.1 Participants

Sixty-eight students, who were between 19 and 49 years old, native speakers of English, and with normal hearing, from San José State University participated for course credit. Each participant was given a subject number by the order of attendance. Even numbered participants were assigned to the i-group and the odd numbered participants were assigned to the u-group.

2.2 Procedure

Each participant attended two experiment sessions which were at least a week apart from each other. The first session consisted of pre-exposure (henceforth *pre-test*) trials with categorization tasks: In each trial, the participants first listened to an auditory stimulus from a nine-step [bip]-to-[bup] continuum in which vowels ranged from natural [i] to natural [u]. They were then asked to indicate whether they had heard “beep” (/i/-response) or “boop” (/u/-response).

The second session began with an exposure followed by a post-exposure (henceforth *post-test*) trials, which were the same as the pre-test trials. During the exposure phase, the participants performed auditory lexical decision tasks. In each of 200 trials, they listened to a different auditory stimulus and were asked to indicate whether they had heard an actual word in English or not. Feedback was displayed on a monitor after each response—a green dot for a correct response or a red dot for an incorrect response. Of the 200 stimuli, 100 were actual words, and the other 100 were non-words. Critically, the set of 100 word stimuli included twenty words that contained an ambiguous vowel [i/u] to induce perceptual learning.

For all tasks, stimuli were presented from headphones in a different random order for each participant, and the participants were asked to respond as quickly as possible by pressing a key on the keyboard. When a response was made within three seconds after stimulus presentation, the response was logged and a new stimulus was presented one second later. If no response was made within three seconds, the next stimulus was presented without the response logged.

2.3 Materials

Auditory stimuli for both categorization and lexical decision tasks were created by source-filter re-synthesis using PRAAT (Boersma and Weenink, 2013). For both, the process involved creating [i]-to-[u] continua with different consonantal contexts. Care was taken to ensure that the vowels reflect natural coarticulatory effects. F2 in [u] varies considerably across contexts, and due to the sluggishness of the lip rounding gesture (Stevens *et al.*, 1966, p. 131), [u] is expected to have longer transition to and from an adjacent consonant than [i]. These two characteristics—formant frequencies and transition durations—varied in all re-synthesized continua.

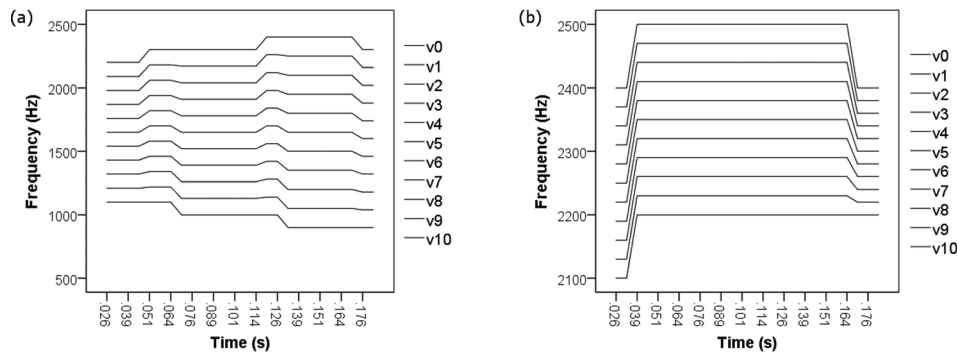


Fig. 1. (a) F2 trajectories and (b) F3 trajectories in the eleven re-synthesized vowels.

Materials for the categorization tasks were syllables from a [bip]-[bup] continuum. A male speaker recorded two syllables [bip] and [bup], which were then adjusted to have the same pitch contours and durations. We then extracted the two vowels and (1) obtained a time-varying LPC filter for each, (2) derived nine intermediate filter functions by interpolating the two, (3) generated a source sound by applying the inverse LPC filter to the original [i], and (4) created an eleven-step vowel continuum by applying all filters to the source. Finally, we inserted the first nine vowels between the original [b] and [p] to create a nine-step syllable continuum ranging from [bip] to [bup].

Each filter consisted of stylized F1 to F5 and associated bandwidths. The F2 and F3 varied over the duration of the vowel to capture formant trajectories while the other three formants remained stable. As shown in Fig. 1, the trajectories for the intermediate vowels shifted with equal frequency intervals so that the resulting stimuli gradually changed from [bip]-like to [bup]-like both in formant frequencies and transition durations. Table 1 presents F2 and F3 values in the final set of nine vowels (v0-v8) near the terminus.

Materials for the lexical decision task were 100 words and 100 non-words selected following the criteria in Kraljic and Samuel (2005). The 100 words consisted of forty *critical* words and sixty *filler* words. Half of the critical words were i-words, each containing one /i/ but no /u/ (e.g., *athlete*). The other half were u-words, containing one /u/ but no /i/ (e.g., *bamboo*). The critical words were two to four syllables long, with the critical vowels (/i/ or /u/) appearing in a stressed syllable in a later part of the word, and without /t/ or /ʊ/. The i-words and u-words were matched in syllable length and mean log-frequency. Mean log-frequency in the Corpus of Contemporary American English (Davies, 2008) was 7.101 for the i-words and 7.076 for the u-words, which were not significantly different ($t(38) = 0.051$, $p = 0.959$). The sixty filler words met the following criteria: (1) no instance of /i/, /t/, /u/, or /ʊ/; and (2) each word matching one pair of critical words in terms of stress pattern and number of syllables. Finally, the 100 non-words were created by changing a few phonemes (without using /i/, /t/, /u/ or /ʊ/) of each of the 100 real words.

These 200 exposure words were then recorded from the same male speaker. Forty critical words were read twice, once with a natural vowel, and once with the critical vowel replaced by the other vowel (e.g., *athlete* as [æθlit] and [æθlut]). We then synthesized an eleven-step continuum separately for each of the 40 pairs in the same way as the [bip]-[bup] continuum. Extraction of the critical vowels (/i/ and /u/) was done by manually comparing the spectrograms and isolating the intervals where the spectral characteristics before and after them were nearly identical between the two words. Thus, when the vowel is adjacent to a sonorant, the extracted interval includes a part of the sonorant to ensure natural transition. After synthesizing the forty [i]-[u] word continua, one of the authors selected the most ambiguous token from each continuum to generate a set of forty *ambiguous* critical words. Figure 2 shows, using *athlete* as an example, how the resulting ambiguous word differs from the original

Table 1. F2 and F3 near the terminus (0.139 s point) of each of the nine vowels used in the experiment.

Step #	1	2	3	4	5	6	7	8	9
F3 (Hz)	2500	2470	2440	2410	2380	2350	2320	2290	2260
F2 (Hz)	2400	2250	2100	1950	1800	1650	1500	1350	1200

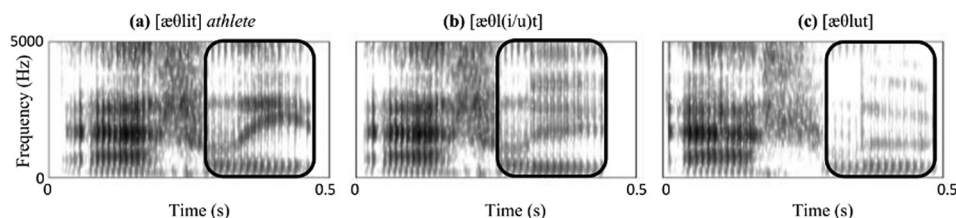


Fig. 2. Spectrograms illustrating stimulus construction: (a) and (c) [i] and [u] in the original recordings (*athlete* as [æθlit] and [æθlut]), respectively. (b) An ambiguous vowel [i/u] embedded in the [i]-frame ([æθl(i)u)t]).

utterances. Finally, since the synthesis involved down-sampling of the original recordings to 10 000 Hz, all natural stimuli were down-sampled to 10 000 Hz.

The resulting auditory stimuli consisted of 200 sound files: twenty ambiguous critical words, twenty natural critical words, sixty word fillers, and 100 non-word fillers. The critical stimuli differed between the two groups. Listeners in the i-group heard the i-words as ambiguous words (e.g., *athlete* as [æθl(i)u)t]) and the u-words as natural words (e.g., *bamboo* as [bæmbu]), and listeners in the u-group heard the u-words as ambiguous words (e.g., *bamboo* as [bæmb(i)u]) and the i-words as natural words (e.g., *athlete* as [æθlit]).

3. Analyses and results

Data from twelve participants were excluded from analyses: six did not complete and six accepted the critical words as real words less than half the time. Below we discuss results from the remaining 56 participants.

3.1 Lexical decision

Table 2 lists mean rates at which stimuli were accepted as real words. Our listeners did not accept the ambiguous words as often as the natural critical words or filler words. But they clearly differentiated the ambiguous words from the non-word fillers. It is worth noting that listeners in the i-group accepted more ambiguous words than the u-group [$t(51) = 3.664$, $p < 0.001$]. Although explicit lexical decision is not a necessary condition for perceptual learning (McQueen *et al.*, 2006), this could mean they had different degrees of exposure. However, this is unlikely to be an issue. First, the between-group difference in acceptance rate progressively reduced over the course of trials [Fig. 3(a)], and for the last ten trials the mean acceptance rates were no longer significantly different [0.779 for the i-group vs 0.737 for the u-group: $t(49) = 1.055$, $p = 0.148$]. Second, regression analyses (see below) revealed that the acceptance rate was not a significant factor in explaining the variance in pretest-posttest responses.

3.2 Categorization

We conducted mixed effects logistic regression analyses using R version 3.3.0 (R Core Team, 2016). The choice was made for two reasons: We removed responses that took subjects more than three seconds. We also wanted to examine the effect of exposure on the listeners categorization behavior. Mixed effects models are robust to missing data (Baayen *et al.*, 2008) and regression analyses allow one to test effects easily via model comparison. The dependent variable was a binary variable indicating whether the response is /i/ or not ($0 = /u/$, $1 = /i/$). Fixed effects were phase ($0 = \text{pre-test}$, $1 = \text{post-test}$), group ($-1 = \text{u-group}$, $+1 = \text{i-group}$), linear and quadratic terms for step² (coded as a continuous predictor centered on 0), as well as all possible interactions among phase, group, and the quadratic term for step. For random effects, we included a random intercept for subjects and by-subject random slopes for phase, the linear and quadratic terms for step, and the interaction between phase and the quadratic term.

Table 3 shows the summary of the best model we found. There were significant main effects of step (both linear and quadratic terms) as well as a significant interaction between step (quadratic) and phase, reflecting that the effect of exposure is particularly

Table 2. Mean acceptance rate during exposure per subject group and stimulus type.

	Natural critical	Ambiguous critical	Real word filler	Non-word filler
/i/-group	0.801	0.776	0.900	0.133
/u/-group	0.822	0.680	0.944	0.120
both groups	0.811	0.729	0.921	0.127

Table 3. Summary of mixed-effects logistic regression analyses on the /i/-response data.

Parameters	Fixed effects				Random effects (by subject) Variance
	Estimate β	SE	z	p(> z)	
Intercept	-3.487	0.277	-12.583	<0.001	3.524
Step	1.785	0.070	25.605	<0.001	0.169
Step ²	0.341	0.028	12.317	<0.001	0.014
Phase _{POST}	0.239	0.280	0.852	0.394	3.153
Group	0.284	0.273	1.041	0.298	
Step ² × Phase _{POST}	-0.117	0.047	-2.467	0.014	0.058
Step ² × Group	-0.038	0.027	-1.438	0.150	
Phase _{POST} × Group	0.860	0.279	3.079	0.002	
Step ² × Phase _{POST} × Group	-0.050	0.045	-1.102	0.270	

strong in the middle range. More importantly, the interaction between phase and group was significant. After exposure, listeners in the i-group responded with more /i/-responses than before, while listeners in the u-group responded with less /i/-responses than before [Fig. 3(b)]. So their response patterns diverged although they were very similar before exposure.

Given the significant interaction between phase and group, we proceeded to a planned comparison by fitting a model to each group's data separately, using the same set of predictors except for the group variable. Phase was a significant factor in the i-group ($\beta = 1.128$, $SE = 0.360$, $p = 0.002$) but not in the u-group ($\beta = -0.813$, $SE = 0.460$, $p = 0.077$). That is, there is evidence for category retuning through lexically guided perceptual learning in the i-group but not in the u-group. However, one could argue that the main effect of phase in the u-group is at least marginally significant given $p = 0.077$. Therefore, we also directly compared the degree of category boundary shift between the two groups. Category boundary was obtained separately per subject and phase by fitting the logistic function on the proportion of /i/-responses as a function of the step number on the [bip]-[bup] continuum and identifying its midpoint on the x axis. The resulting boundary locations are summarized in Fig. 3(c). On average, the boundary shifted towards the [u]-end of the continuum by 0.401 steps for the i-group, while it shifted towards the [i]-end by 0.219 steps for the u-group. The extent of shift is marginally larger for the i-group than the u-group [$t(110) = 1.430$, $p = 0.078$]. Taken together with the results from the regression analyses, we argue that there is evidence that the i-group shifted their responses more than the u-group.

We attribute the observed group difference to difference in inherent malleability between [i] and [u]. As mentioned earlier (Sec. 3.1), another possibility is to attribute it to difference in acceptance rate of the ambiguous words during the exposure phase. But this alternative interpretation does not seem to explain our data any better: We compared the model in Table 3 with two other models that included acceptance rate as a continuous factor centered on its mean. The first model, in which acceptance rate replaced the group factor of the original model, did not explain the data better; it had a higher AIC (3705.1 vs 3691.6) and a lower log-likelihood (-1828.5 vs -1821.8). The second model, which had acceptance rate in addition to all predictors in the original model, had a slightly lower log-likelihood (-1819.8 vs -1821.8), but a likelihood-ratio test revealed that the two models were not significantly different [$\chi^2(8) = 4.003$, $p = 0.857$]. Moreover, in both models, neither the main effect of acceptance rate nor its interactions with the other predictors were significant, and in the second model, the phase-by-group interaction remained significant ($\beta = 0.860$, $SE = 0.279$, $p = 0.002$).

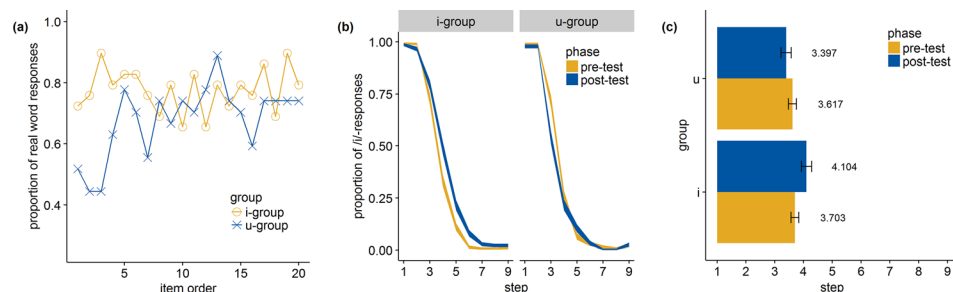


Fig. 3. (Color online) (a) Acceptance rates in response to ambiguous words changing over exposure trials. (b) Proportion of /i/-responses per step and subject group. Width of curve represents 2 standard errors. (c) Mean boundary locations on the [bip]-[bup] continuum.

4. Discussion and conclusion

We observed a larger exposure effect in the i-group than the u-group, which is consistent with our hypothesis that [u] would be less malleable than [i] due to its greater category variability. Relevance of the result to our hypothesis, however, is contingent on whether the two groups had comparable exposure. Their responses during exposure were different: the i-group initially accepted the ambiguous words more than the u-group. Though our mixed effects analyses suggest that this is unlikely to have been a significant factor, there may have been stimuli artifacts, which we plan to reduce in future studies. First, in our [i]-[u] continua, F2 and F3 varied linearly rather than logarithmically. Considering that frequency perception is better modeled using the log scale, two stimuli equidistant from the two ends of our continua may not have sounded equally aberrant: those near the [i]-end may have sounded less aberrant than those near the [u]-end. Second, our synthetic stimuli may not have fully captured dynamic properties such as transition duration and range of frequency change that characterize coarticulation on [u]. We partly addressed these issues by creating an ambiguous vowel for each critical word rather than using the same vowel for all 40 words. In future studies, we plan to control the naturalness of our stimuli better by conducting a perceptual survey at a larger scale.

Assuming that the observed difference is not a mere stimulus artifact, findings in the present study add to the literature possible sources of constraints on category retuning. Previous studies have found that retuning can be modulated by the strength of contextual evidence that doing so would benefit subsequent processing (e.g., [Kraljic et al., 2008](#)), suggesting an evaluation process as one source of constraints. Findings from our study, in line with others ([Scharenborg and Janse, 2013](#); [Stevens et al., 2007](#)), further suggest that the way phonetic categories are represented within the perception system may also constrain the malleability of each category. Perhaps the mental representation of a category reflects its normative range of phonetic variation and the represented category variability in turn constrains the degree of category malleability.

A remaining question is why highly variable categories would resist boundary adjustment. One possibility is that such categories are mentally represented in a way to tolerate deviation more, obviating boundary adjustment. For example, following exemplar-based approaches (e.g., [Johnson, 1997](#)), one might argue that our subjects, mostly from California, had more exemplars of [u] with high F2 values in their mental lexicon than exemplars of [i] with low F2 values. To them, the ambiguous u-words might have sounded less atypical and exerted less need for category retuning. Or, based on the perceptual magnet model ([Kuhl et al., 1992](#)), which assesses the category goodness of a sound based on its auditory distance from the category prototype, one might argue for the opposite, that a deviant u-word would be more likely to be rejected as a possible category member and therefore less effective in inducing retuning than a deviant i-word. These scenarios imply that learning, in a sense of improved sound-to-phoneme mapping, can occur with or without changing existing phonetic category structures depending on whether listeners perceive new tokens as reasonable pronunciation variants or extreme aberrations. A similar argument can be made based on the Featurally Underspecified Lexicon model (e.g., [Eulitz and Lahiri, 2004](#)), which states that some phonemic features of a sound are not stored in the mental lexicon and that deviation in underspecified features is less detectable. Perhaps our ambiguous u-words deviated in a fully specified feature (e.g., [back]), being perceived as more deviant, while our ambiguous i-words deviated in an underspecified feature (e.g., [front]), being perceived as more acceptable. Either way, whether this assumption of asymmetric sensitivity and/or acceptability holds for [i] vs [u] must be tested in future studies.

Another possibility, assuming comparable detection and assessment of deviation between [i] and [u], is that the acoustic space of a category with higher variability has a greater chance of overlap with other categories and this discourages the perceptual system from broadening the category space further. Such differential degree of category overlap is suggested by different confusability between the two vowels. According to confusion matrices compiled by [Luce \(1986\)](#), [u] is indeed more confusable than [i]. At signal-to-noise ratios of +15, +5, and -5 decibels, [u] was misrecognized as other vowels 28.22%, 50.22%, and 83.56% of the time, while [i] was misrecognized 9.33%, 7.78%, and 35.56% of the time. Our conjecture is also in line with an experimental study by [Wade et al. \(2007\)](#), who show that if non-native speakers pronounce a vowel with much variability thereby raising its confusability with other vowels, native listeners have trouble adapting to the foreign accent. We will further explore this idea in future studies by examining the roles of auditory confusability and articulatory and acoustic variabilities as constraints on perceptual learning.

Acknowledgments

We thank the anonymous reviewers for their valuable comments and suggestions. We also thank our research assistants Leslie Bank, Netta Ben-Meir, Veronica Espinoza, Yu-Han Kuo, Janine Robinson, Mary Ryan, Juliet Solheim, Frank Yeh, Matthew Cortez, and Oi Lam Ng for help with data collection in the pilot and final studies.

References and links

¹The term “perceptual learning” has been used to mean different phenomena across studies (Samuel and Kraljic, 2009). In this paper, the term is used to denote phonetic category being “retuned to become more aligned with the input” (Samuel and Kraljic, 2009, p. 1212).

²We included the quadratic term because the log-odds of /i/-response did not decrease linearly for each step. Accordingly, the model with the quadratic term explained our data better, with a lower AIC (3691.6 vs 3901.9) and a higher log-likelihood (−1821.8 vs −1932.9) than the model with only the linear term.

- Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). “Mixed-effects modeling with crossed random effects for subjects and items,” *J. Mem. Lang.* **59**, 390–412.
- Boersma, P., and Weenink, D. (2013). “Praat: Doing phonetics by computer” [computer program], version 5.3, <http://www.praat.org> (Last viewed 4/11/2014).
- Davies, M. (2008). The Corpus of Contemporary American English (COCA): 520 Million Words, 1990–present, <http://corpus.byu.edu/coca/> (Last viewed 3/2/2014).
- Eulitz, C., and Lahiri, A. (2004). “Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition,” *J. Cogn. Neurosci.* **16**, 577–583.
- Harrington, J., Palethorpe, S., and Watson, C. I. (2000). “Does the Queen speak the Queen’s English?,” *Nature* **408**, 927–928.
- Hillenbrand, J. M., Clark, M. J., and Nearey, T. M. (2001). “Effects of consonant environment on vowel formant patterns,” *J. Acoust. Soc. Am.* **109**, 748–763.
- Johnson, K. (1997). “Speech perception without speaker normalization: An exemplar model,” in *Talker Variability in Speech Processing*, edited by K. Johnson and J. W. Mullennix (Academic Press, San Diego, CA).
- Kent, R. D., and Moll, K. L. (1972). “Cinefluorographic analyses of selected lingual consonants,” *J. Speech Hear. Res.* **15**, 453–473.
- Kraljic, T., and Samuel, A. G. (2005). “Perceptual learning for speech: Is there a return to normal?,” *Cogn. Psychol.* **51**, 141–178.
- Kraljic, T., Samuel, A. G., and Brennan, S. E. (2008). “First impressions and last resorts: How listeners adjust to speaker variability,” *Psychol. Sci.* **19**, 332–338.
- Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., and Lindblom, B. (1992). “Linguistic experience alters phonetic perception in infants by 6 months of age,” *Science* **255**, 606–608.
- Labov, W., Ash, S., and Boberg, C. (2006). *The Atlas of North American English: Phonetics, Phonology and Sound Change* (Mouton de Gruyter, New York).
- Lisker, L., and Abramson, A. S. (1970). “The voicing dimension: Some experiments in comparative phonetics,” in *Proceedings of the 6th International Congress of Phonetic Sciences*, Prague, 1967 (Academia, Prague), pp. 563–567.
- Luce, P. A. (1986). “Neighborhoods of words in the mental lexicon,” Ph.D. dissertation, Indiana University, Bloomington, IN.
- MacNeilage, P. F., and DeClerk, J. L. (1969). “On the motor control of coarticulation in CVC monosyllables,” *J. Acoust. Soc. Am.* **45**, 1217–1233.
- McQueen, J. M., Norris, D., and Cutler, A. (2006). “The dynamic nature of speech perception,” *Lang. Speech.* **49**, 101–112.
- Norris, D., McQueen, J. M., and Cutler, A. (2003). “Perceptual learning in speech,” *Cogn. Psychol.* **47**, 204–238.
- Öhman, S. E. G. (1966). “Coarticulation in VCV utterances: Spectrographic measurements,” *J. Acoust. Soc. Am.* **39**, 151–168.
- R Core Team (2016). “R: A language and environment for statistical computing,” R Foundation for Statistical Computing, Vienna, Austria, <http://www.R-project.org/> (Last viewed 6/11/2016).
- Samuel, A. G., and Kraljic, T. (2009). “Perceptual learning for speech,” *Atten. Percept. Psychophys.* **71**, 1207–1218.
- Scharenborg, O., and Janse, E. (2013). “Comparing lexically guided perceptual learning in younger and older listeners,” *Atten. Percept. Psychophys.* **75**, 525–536.
- Stevens, K. N., House, A., and Paul, A. P. (1966). “Acoustical description of syllabic nuclei: An interpretation in terms of a dynamic model of articulation,” *J. Acoust. Soc. Am.* **40**, 123–132.
- Stevens, M. A., McQueen, J. M., and Hartsuiker, R. J. (2007). “No lexically-driven perceptual adjustments of the [x]-[h] boundary,” in *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken, 2007 (Pirrot, Dudweiler), pp. 1897–1900.
- Wade, T., Jongman, A., and Sereno, J. (2007). “Effects of acoustic variability in the perceptual learning of non-native-accented speech sounds,” *Phonetica* **64**, 122–144.