The distribution of coarticulatory variation influences perceptual adaptation

Georgia Zellou and Bruno Ferenc Segedin

University of California, Davis
gzellou@ucdavis.edu, bferencsegedin@ucdavis.edu

ABSTRACT

This study explores how variations in the distribution of a novel phonetic pattern across talkers and vowel categories affect perceptual adaptation to speech. Listeners were exposed to a shifted phoneme category realization either in all words produced by one of two speakers (Expt 1) or within a subset of the words (comprising a natural class: i.e., containing mid vowels) produced by one of two speakers (Expt 2). We find that when listeners are exposed to the shifted pattern across all words in a single voice, they adapt to that speaker only. However, when the shifted pattern is present only in mid vowels from one speaker, listeners show adaptation to words with that vowel category and generalize across speakers. Our results indicate that the *distribution* of a phonetic shift across linguistic (vowel-class) and social (talker-specific) categories affects the target for adaptation that listeners generate.

Keywords: Perceptual Adaptation, Coarticulation, Distribution of Phonetic Variation.

1. INTRODUCTION

Using lexical knowledge, L1 sound-to-meaning mappings can be retuned following exposure to a novel accent [1, 2]. Since talkers' voices differ in ways that can be idiosyncratic ("Sarah sounds rather nasal") or social ("Minnesotans talk nasally"), successful comprehension depends on adaptation to variable speech patterns. Exploring how listeners adapt to variations in the realization of lexical categories can inform speech perception theories. Yet, the nature of the representations that listeners generate during perceptual adaptation is unresolved. For one, there is conflicting evidence as to whether perceptual adaptation is talker-specific or not. In some work, perceptual learning in one voice generalizes to new talkers, e.g., [2]; However, talker-specific adaptation has also been observed, e.g., [4]. Thus, it is unclear under what conditions listeners generate veridical, talker-specific or generalizable, cross-talker category shifts.

How patterns of exposure influence the *specificity* of categories over which perceptual shifts apply is the focus of the current study. Specifically, if listeners hear particular words with similar phonological structures, like "den" and "ben", from a single talker with a novel phonetic realization, which category will listeners employ as the target of adaptation? A phonetic shift might be encoded with: 1) just the social information (for that talker, generalizing across words), 2) just linguistic information (for those words, generalizing across talkers), or 3) both (for just those shifted words spoken by that talker). The *distribution* of phonetic variation over productions by talkers during exposure influences specificity of category learning in infants [9]: talker-word correlated variation led babies to generate highly-specific shifted categories (i.e., our option 3). We test these possibilities for how talker-item correlation influences adult perceptual adaptation by varying the distribution of a novel phonetic pattern present across words and talkers.

2. EXPERIMENT 1

Experiment 1 tests whether listeners adapt to distinct talker-specific coarticulatory patterns heard across all word productions by each talker.

2.1. Methods

2.1.1. Stimuli

A female native English speaker produced repetitions of words used to create the stimuli for Experiments 1 & 2. Words used for stimuli creation included ten sets of CVC-CVN-NVN minimal triplets with matching onset and coda place of articulation (/ɛ, ɑ, ʌ, e, æ/ in bVd-bVn-mVn and dVd-dVn-nVn frames).

These items were modified in three ways to create the stimulus items (after amplitude-normalization): First, all items were modified to create an additional apparent talker, i.e., apparent male voice. The voice modification was intended to create two apparent talker versions of identical items, so that idiosyncratic aspects of two talkers remained identical. Items were manipulated by lowering both f0 and formant frequencies to a value roughly appropriate for an adult male. The result was two versions of each item: one with characteristics of an adult female (high f0 and FFs) and another with characteristics appropriate for an adult male (low f0 and FFs). Next, stimuli for the exposure phase (procedure details below) consisted of items with vowels same- or cross-spliced from different tokens to create distinct phonetic variation patterns for each talker across words. For the "unshifted" pattern, nasal vowels (from CVN contexts) were same-spliced into C_N frames; also, oral vowels (from CVCs) same-spliced into C_C frames. For the "shifted" pattern, hypernasal vowels (from NVNs) were cross-spliced into C_N frames; and, nasal vowels (from CVNs) were cross-spliced into C_C frames. Finally, stimuli for the test phase (procedure details below) consisted of CV syllables containing nasal vowels spliced onto initial consonants of each item for each speaker. Syllables containing were also generated as control items. Syllables were gated into wide-band noise, 5dB less than the vowel's peak intensity (following methods in [8]) to reduce perceptual biases toward a final stop coda.

2.1.2. Participants & Procedure

58 native English-speaking UC Davis undergraduates participated in Experiment 1, consisting of an exposure phase, followed by a word-completion task (test phase). None reported any visual or hearing impairments.

The goal of the **exposure phase** was to provide lexically-guided experience with the speaker-specific phonetic patterns. During exposure, listeners heard the two apparent talkers with systematically different patterns of vowel nasality in CVC and CVN words productions. One voice produced appropriate vowels in CVC and CVN words. This is the "unshifted" voice, whose productions in exposure reflect their natural coarticulatory patterns. The other voice, referred to as "shifted", produced nasalized vowels in CVC words and hypernasalized vowels in CVN words; their productions reflect nasality patterns of shifted structural categories, with enhanced vowel nasality (see Table 1 for a examples). During exposure, we also presented listeners with an *unshifted* voice, of the other gender (e.g., if the listener heard a female shifted voice, they also heard a male unshifted voice, and vice versa). In so doing, we aimed emphasize differences in the structured variation present in both voices.

Apparent talker assignments were counterbalanced across two groups: 28 participants heard the female voice "shifted" in exposure; 30 participants heard the male voice "shifted" in exposure. Participants heard each of the 20 words (comprised of 10 CVC-CVN minimal pair items) two times in each voice (=80 exposure trials). On each exposure trial, a word was presented auditorily, and the corresponding lexical item was presented on the screen.

Table 1: Exposure Phase example trials.

Talker	**CVC category** Hear:	See:	**CVN category** Hear:	See:
"Unshifted" Voice	[ded]	*"dead"*	[dẽn]	*"den"*
"Shifted" Voice	[dẽd]	*"dead"*	[dẽ̃n]	*"den"*

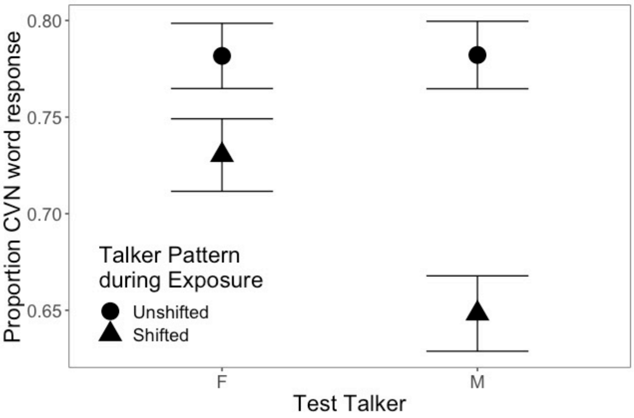
Immediately following exposure, participants completed a word completion task (**test phase**). This task follows [8]: on a given critical trial, listeners hear a syllable fragment containing a nasal vowel, gated into noise. Listeners then selected one of two minimal pair choices to complete the lexical item (either a CVC or CVN, corresponding to the minimal pair option for that syllable). In the test phase, listeners heard *both voices* produce CV syllables containing nasal vowels. Since listeners had already heard both voices in exposure, we expect the likelihood of identifying nasal vowels as CVC or CVN items to vary based on whether they heard that particular voice as shifted or unshifted in exposure.

Participants heard an equal number of control trials with an oral vowel, so their responses would not be biased toward nasal lexical choices. There were 40 test trials (2 repetitions of ten minimal pairs*2 talkers), plus 40 filler trials.

2.2. Results

Listener responses were coded for completing the fragment as either the CVN (=1) or the CVC minimal pair item (=0) in each test trial. Listeners' mean responses for identifying word fragments as the CVN item are provided in Figure 1. A logistic mixed-effects regression model was fitted to the responses using the *glmer()* function in *lme4* [1] using *R* [10]. The model included two fixed effects predictors, and the two-way interaction, to test the effect of exposure to talker-specific vowel coarticulatory patterns on subsequent vowel categorization. First, Talker Exposure Pattern, a categorical variable with two levels (unshifted voice, shifted voice [base level]), tested whether hearing either that voice as unshifted (i.e., CVC-CṼN phonetic pattern) or that voice as shifted (i.e., CṼC-CṼ̃N phonetic pattern) in exposure influenced the likelihood of lexical completion. Third, Speaker was included as a categorical variable with two levels (Male voice or Female voice [base level]) to test whether hearing the shifted phonetic pattern in one of the apparent talkers' voices influenced adaptation behavior. The model included by-participant random intercepts and by-participant random slopes for Speaker and Talker Exposure Pattern (within-subject factors).

Figure 1: *Test Phase*. Pooled proportion of CVN responses to syllables with nasal vowels, after exposure to one voice with a shifted nasality pattern (either M-shifted or F-shifted; between-subjects).



Overall, there was a main effect of Talker Exposure Pattern: listeners were reliably more likely to select the CVN minimal pair for nasal vowel syllables produced by the unshifted voice in exposure (78%), relative to responses in the shifted voice (69%) (z=3.5, p<.01). This main effect is seen in Figure 1. There was not a significant two-way interaction between Speaker Gender and Talker Exposure Pattern (z=-3.3, p=.5). Although the magnitude of the shift appears larger in the apparent-Male voice than for the apparent-Female voice, this was not reliably different. No other main effects were significant.

These results indicate that listeners generate a speaker-specific representation for phonetic-to-sound category mapping for vowel nasality patterns. After hearing a category-shifted speaker, who produced nasalized vowels in CṼC words and hypernasalized vowels in CṼ̃N words, listeners were more likely to categorize that speaker's nasal vowels as signaling CVC words, relative to nasal vowels in the unshifted talker's voice. Hence, listeners adapted to talker-specific vowel-nasality systems signaling nasal lexical contrasts.

3. EXPERIMENT 2

Experiment 2 examines whether the *distribution* of a shifted phonetic pattern within a single talker's productions influences patterns of perceptual adaptation. As in Experiment 1, listeners were exposed to two apparent talkers: one talker is unshifted, while the other talker has shifted nasality patterns (and, they vary in apparent gender). Now, however, the shifted talker produced a *vowel-specific shift*: only the mid vowels are shifted.

There are several possibilities for how changing the distribution of the novel phonetic pattern in exposure will influence listeners' adaptation. Listeners might adapt veridically, displaying the perceptual shift only on the specific vowels by the talker who was shifted in exposure. Alternatively, the target of adaptation might be broader. Listeners might identify the phonological (here, vowel-class) category as the target of adaptation, displaying the shift in mid vowel for both talkers (generalizing across talkers). Or, the target of adaptation might be talker-specific, in which we predict listeners will shift all vowels produced by the mid-vowel-shifted talker (generalizing across vowel categories).

3.1. Methods

Experiment 2 used the same stimuli and procedures from Experiment 1—**only the distribution of the shifted phonetic nasality pattern in the exposure phase differed**. 74 native English-speaking UC Davis undergraduates participated in Experiment 2 (with no reported visual or hearing impairments). One listener group (n=37) was exposed to the shifted phonetic nasality pattern in mid vowels only in the Male voice and the unshifted pattern in low vowels in the Male voice and all words produced by the Female voice; the other group (n=37) was exposed to the shifted phonetic nasality patterns in words with mid vowels produced by the Female voice only and unshifted nasality patterns in all other words by the Female and the Male voice. Our decision to subset the vowel categories in this way (i.e., mid vs. low vowels) reflects patterns of differential degrees of coarticulatory nasality as a function of vowel height cross-linguistically [3].

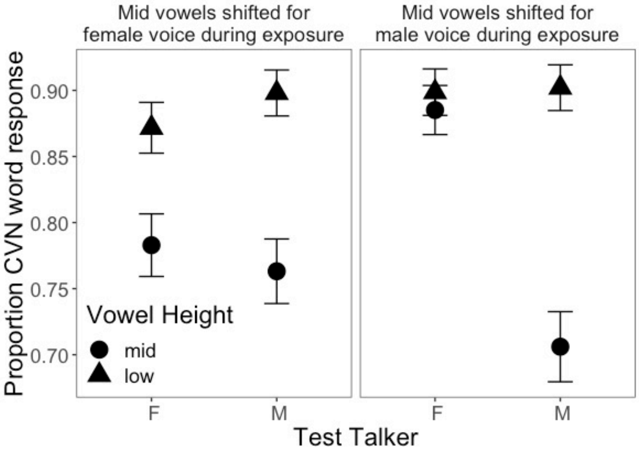
3.2. Results

Listener responses, coded for nasal word responses as in Experiment 1, are provided in Figure 2.

Responses to nasal vowels were analyzed using a mixed-effects logistic regression. The model included main effects of Vowel Height (mid, low), Talker Exposure Pattern (unshifted voice, shifted voice [base level]), tested whether hearing either unshifted (i.e., CVC-CṼN phonetic pattern) or shifted (i.e., CṼC-CṼ̃N phonetic pattern), and Speaker (F, M). All possible interactions were also included. By-participant random intercepts and by-participant random slopes for Vowel Height and Speaker (and their interaction) were included.

Vowel height was a significant predictor of CVN word responses for nasal vowels: listeners were more likely to indicate that nasal low vowels signaled a CVN item (89%) than nasal mid vowels (78%) (z=4, *p*<.001). Even though they were exposed to a shifted nasal system in only a portion of the words with mid vowels in the pre-test, there was not an interaction between Group and Vowel Height (*p*=.6). However, the three-way interaction of Vowel Height, Participant Group, and Talker was significant (*z*=2.1, *p*<.05). This interaction is illustrated in Figure 2: listeners who heard only mid vowel shifted in the female voice in exposure (right panel) displayed mid vowel adaptation in both the female's mid vowel **and** the male speaker's mid vowels; meanwhile, listeners who heard only shifted mid vowels in the male talker's voice in exposure (left panel) did not generalize the shift to the female voice.

Figure 2: *Test Phase*. Pooled proportion of CVN responses to syllables with nasal vowels, after exposure to a shifted voice (either M-shifted or F-shifted; between-subjects condition) in **mid vowels only.**



4. GENERAL DISCUSSION

We find that the distribution of phonetic variation influences perceptual adaptation: Speaker-specific adaptation occurs when two talkers display distinct patterns across all their word productions (Experiment 1). Yet, when a single talker displays a vowel category-specific distribution of a novel phonetic pattern, listeners are more likely to hone in on the linguistic category (here, mid vowels) as the target of the perceptual boundary shift and generalize over talker identity (Experiment 2, for the Female shifted group). Thus, these findings reveal that how variation is distributed in speech between social (here, talker-specific) and linguistic (here, vowel phoneme) categories influences which of those categories listeners generalize over during perceptual adaptation.

The observation that the distribution of phonetic variation during exposure influences whether listeners display veridical adaption or generalize is relevant for proposed mechanisms of perceptual learning. For example, [6] argues that perceptual adaptation is driven by the phonetic patterns present in the input, and does not involve a relaxation of the criteria for a phoneme category. Our results support the fact that adaption is systematically related to the structured variation in the input, however, we suggest that this mechanism is sensitive to how the phonetic variation is distributed over linguistic and social categories, weighing the linguistic category over the social category, in some cases, when they correlate.

We also observed a gender asymmetry in Experiment 2: adaptation was in fact veridical for the male voice (i.e., there was no generalization to the female voice when the male's mid-vowels were shifted); meanwhile, generalization from female-shifted to the male voice in testing was observed in Experiment 2. Notably, this aligns with reports from prior work: for example, [4] found that listeners trained in a fricative category shift on a female voice generalized to a male voice, but not vice versa. Thus, there is evidence of asymmetries in adaptation to male and female voices. One question is whether these are reflective of adaptation to the general social categories, or simply idiosyncratic talker effects. More robust generalization of category shifts from female talkers is a scenario which aligns with observations of sound changes being led by young female speakers in a speech community (e.g., [5]). If listeners reflect sensitivity to socio-indexical categories during generalization of perceptual adaptation, this could be one explanation for how innovative phonetic variants diffuse socially across speech communities. Another potential explanation could stem from the fact that one voice was natural (female) and the other was synthesized from (male): Thus, this could have led listeners to adapt differently for the male voice than for the female voice. These possibilities can be addressed in future work.

Listeners adapt to innovative speech patterns by shifting their perceptual boundary of a sound category from exposure to a novel phonetic variant when presented in a lexical item. In some conditions, that perceptual shift is associated and applied only to the voice in which the shift was initially heard. In other cases, listeners generalize that shift to other talkers even if the innovative pattern was not heard in that particular speaker's voice. The conditions under which adults generalize a perceptual shift is an important avenue to explore because it has implications for sound change: the generalization of a novel sound-to-meaning mapping to different speakers can provide insight into the origin and spread of sound change within a speech community.

5. REFERENCES

[1] Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.

[2] Dahan, D., Drucker, S., & Scarborough, R. (2008). Talker adaptation in speech perception: Adjusting the signal or the representations?. *Cognition*, *108*(3), 710-718.

[3] Hajek, J. (1997). Universals of Sound Change in Nasalisation. Oxford: Blackwell.

[4] Kraljic, T., & Samuel, A. (2007). Perceptual adjustments to multiple speakers. *JML*, *56*(1), 1-15.

[5] Labov, W., Rosenfelder, I., & Fruehwald, J. (2013). One hundred years of sound change in Philadelphia: Linear incrementation, reversal, and reanalysis. *Language*, *89*(1), 30-65.

[6] Maye, J., Aslin, R. N., & Tanenhaus, M. K. (2008). The weckud wetch of the wast: Lexical adaptation to a novel accent. *Cognitive Science*, *32*(3), 543-562.

[7] Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive psychology*, *47*(2), 204-238.

[8] Ohala, J. J., & Ohala, M. (1995). Speech perception and lexical representation: the role of vowel nasalization in Hindi and English. In *Phonology and phonetic evidence,* Ed. by *Connell, Bruce, & Arvaniti Amalia: Cambridge U Press*, 41-60.

[9] Quam, C., Knight, S., & Gerken, L. (2017). The distribution of talker variability impacts infants' word learning. *Laboratory Phonology*.

[10] R Core Team, (2017). R: A language and environment for statistical computing.