JASA
EXPRESS LETTERS

https://doi.org/10.1121/10.0000951

Partial perceptual compensation for nasal coarticulation is robust to fundamental frequency variation

Georgia Zellou,[a] Santiago Barreda, and Bruno Ferenc Segedin
Department of Linguistics, University of California, Davis, 469 Kerr Hall, One Shields Avenue, Davis, California 95616, USA
gzellou@ucdavis.edu, sbarreda@ucdavis.edu, bferencsegedin@ucdavis.edu

Abstract: Listeners show better-than-chance discrimination of nasalized and oral vowels occurring in appropriate consonantal contexts. Yet, the methods for investigating partial perceptual compensation for nasal coarticulation often include nasal and oral vowels containing naturally different pitch contours. Listeners may therefore be discriminating between these vowels based on pitch differences and not nasalization. The current study investigates the effect of pitch variation on the discrimination of nasalized and oral vowels in C_N and C_C items. The $f0$ contour of vowels within paired discrimination trials was varied. The results indicate that pitch variation does not influence patterns of partial perceptual compensation for coarticulation. *© 2020 Acoustical Society of America*

[Editor: Douglas D O'Shaughnessy] Pages: EL271–EL276

Received: 13 January 2020 **Accepted:** 7 March 2020 **Published Online:** 19 March 2020

1. Introduction

Speech is a highly variable and ambiguous signal. One source of variability is coarticulation, or overlap between adjacent articulations during the production of segments. For instance, vowels adjacent to nasal consonants are at least partially nasalized due to an overlapping velum-lowering gesture, and this appears to be cross-linguistically universal (Hajek, 1997). Listeners appear to be sensitive to these covarying patterns, and seem to "factor out" coarticulatory variation when its source is known. For example, /g/ is typically associated with a lower $F3$ than /d/ so that stimuli from a /da/-/ga/ $F3$ continuum are increasingly identified as /ga/ as $F3$ decreases. However, following an /ar/ syllable, listeners are more likely to report hearing /da/ than /ga/ for any step along the continuum (Mann, 1980). Mann and others have suggested that this is because listeners attribute a lower $F3$ to the effect of the preceding rhotic, and so do not attribute this acoustic feature directly to the /da/-/ga/ distinction. Similarly, listeners have more difficulty judging the nasality of vowels when they occur adjacent to a nasal consonant, than when the vowels are heard in isolation, suggesting that listeners may perceptually "ignore" nasality given an appropriate contextual source (Kawasaki, 1986). The tendency to perceptually ignore acoustic information due to coarticulation when attributed to some phonological context has been referred to as "perceptual compensation."

One typical way to assess perceptual compensation is by using a paired discrimination paradigm [e.g., 4-interval forced-choice (4IAX); Beddor and Krakow, 1999]. In critical trials, listeners hear two pair of items containing vowels presented in different consonant contexts: in one pair of items the vowels are acoustically identical, while in a second pair the vowels are acoustically different but occur in appropriate coarticulatory contexts, e.g., CVC-CVN vs CVC-CṼ̲N where the non-underlined vowels are the same vowel token. In such a paradigm, listeners are asked to indicate which pair of words has the more similar (or dissimilar) vowels. Three possible outcomes of this design can be considered. First, listeners will respond using the veridical acoustics and always indicate that the acoustically-different pair (e.g., CVC-CṼ̲N) is more dissimilar. In contrast, if listeners fully perceptually compensate for the nasal coarticulation, they should "erase" the nasality in the appropriate context (e.g., CVC-CṼ̲N), resulting in two pairs of non-nasal sounding vowels. This would lead to an at-chance performance.

However, research over the past couple of decades has refined understandings of perceptual compensation by revealing that listeners do not fully ignore coarticulatory detail and that some of the variability present in the speech signal remains perceptible. Listeners may display what has been called "partial" compensation when completing paired discrimination tasks (Beddor and Krakow, 1999). When this occurs, listeners attribute only some of the acoustic

---

[a] Author to whom correspondence should be addressed.

effects of coarticulation to the consonantal source while the remaining, residual coarticulatory detail is perceived as inherent to the vowel and used during its classification (Beddor *et al.*, 2013). Partial compensation indicates that listeners can still "hear" some of the nasality in the CṼN item, which would lead to an above-chance performance in discriminating the CVC-C$\underline{\tilde{V}}$N vowel pair. Partial perceptual compensation has been observed across a variety of studies examining vowel-to-vowel coarticulation (Fowler and Brown, 2000; Beddor *et al.*, 2002), vowel-to-sibilant coarticulation (Yu and Lee, 2004), and is robust to individual differences in perceptual sensitivity to coarticulation (Zellou, 2017).

Partial compensation is used to make inferences about broad theoretical questions with important implications (e.g., Beddor, 2009). Thus, it is important to confirm that these findings in the literature are robust to some common methodological practices employed in this area of research. Paired discrimination requires close attention to the fine-grained phonetic properties of vowel sounds, as typically the differences between stimulus pairs are subtle. When constructing stimuli for paired discrimination tasks, a common practice is to splice a single vowel into different phonological contexts, meaning that in many cases three of the four vowel sounds in a trial will be the same vowel sound (e.g., Fowler, 2006; Beddor and Krakow, 1999; Zellou, 2017). While prior work reports speech manipulation to match oral and nasalized vowels in duration (e.g., Beddor and Krakow, 1999), as well as amplitude (e.g., Zellou, 2017), across items within paired discrimination trials, they do not report controlling for variation in fundamental frequency ($f0$) across stimulus pairs. Practically speaking, this means that in these cases three of four vowels in a trial have identical $f0$ contours, while the fourth has a different contour. Therefore, listeners might exhibit what appears more veridical discrimination, what appears to be partial compensation, simply by identifying the token with the different $f0$ contour.

The current study investigates whether results supporting partial compensation for coarticulation are robust to pitch variation across discrimination items within a trial. To this end, the study examined whether variation in pitch contours influences performance in paired discrimination trials, especially in critical trials involving comparisons like CVC-CVN vs CVC-C$\underline{\tilde{V}}$N (with the non-identical vowel underlined). The traditional paired discrimination paradigm was employed, across five different $f0$ variation conditions that were either congruent or incongruent, with the expected pattern of results suggesting partial compensation. For example, $f0$ was manipulated so that in some cases spectrally identical vowels had different $f0$ contours, and spectrally-different vowels had identical $f0$ contours. If listeners are reporting differences in spectral properties (i.e., the acoustic features encoding coarticulatory variation), and not source properties (i.e., $f0$, not related to coarticulation), then performance will be similar regardless of whether or not the $f0$ contours reflect spectral differences. Importantly, if patterns of partial compensation are robust to source variation between vowels, this would reinforce a wide range of findings that rely on methods which allow for small but perceptible differences in vowels across pairs.

2. Methods

2.1 Participants

The study was completed by 20 participants (19 female), all native English speakers with no reported hearing impairments, who were recruited from the UC Davis psychology subject pool and received course credit for participation.

2.2 Stimuli

Two sets of CVC-CVN minimal pairs were used to generate the stimuli for the experiment, matching in onsets and place of articulation of the codas: *heb*-*hem* and *hobe*-*home* (/ɛ/ and /o/ were used, following Beddor and Krakow, 1999). A male American English speaker produced two repetitions of each word. Recordings were made using a Shure WH20 XLR head-mounted microphone (Shure, Niles, IL) in a sound-attenuated booth and digitally sampled at 44-kHz. Cross-splicing was then used to create versions of each word containing an oral and nasalized vowel. First, an oral and nasalized vowel was selected and excised from one of the CVC and CVN productions, respectively, of a minimal pair.

The onset consonant and coda from the other productions were also excised. Each isolated oral and nasalized vowel was cross-spliced into both C_C and C_N frames, resulting in eight unique stimuli. All stimuli were amplitude-normalized to 60 dB. These stimuli featured vowels with pitch contours that began at approximately 135 Hz, but ended approximately 10 Hz lower in the nasal syllables compared to the oral syllables (110 vs 100 Hz).

Next, all stimuli were manipulated using Praat (Boersma and Weenink, 2005) to give each item four different $f0$ contours based on the original $f0$ contours of the non-modified stimuli. The first three $f0$ contours began at 135 Hz and finished at 110, 105, and 100 Hz,

https://doi.org/10.1121/10.0000951

decreasing linearly from the beginning to the end of the vowel. The final contour begins at 145 Hz and ends at 100 Hz, also decreasing linearly across the vowel. Voicing during consonant closure was set at 95 Hz. (The use of stimuli containing these four contour types in experimental trials is described in Sec. 2.3.)

Vowels before oral codas tend to be longer than vowels before nasal codas. So, duration between oral and nasalized vowel pairs was neutralized to ensure that listeners would not use differences in duration to distinguish nasal and non-nasal vowels. Following Beddor and Krakow (1999), the differences in length between each vowel pair were calculated and the difference was split. Approximately half of the length-difference was added to the nasal vowels, while about half the length-difference was subtracted from the oral vowels. In lengthening a vowel, a single wave cycle at the midpoint was iterated. In shortening a vowel, full pitch pulses were extracted from near the midpoint.

2.3 Procedure

The manipulations outlined in Sec. 2.2 resulted in stimuli that differed independently in the spectral and *f*0 characteristics. Pairs of stimuli based on the same original token but made to differ in their *f*0 were *spectrally* identical. Conversely, pairs of stimuli could be presented such that they had identical *f*0 contours, but were spectrally dissimilar by virtue of being based on different original sounds. Stimuli presentation utilized a 4IAX paired discrimination task. For each trial, listeners were presented with two pairs of words: one pair of words containing vowels that were spectrally identical and the other pair of words containing spectrally different vowels. The spectrally-different pair >always differed in vowel nasality; one oral vowel and one nasal vowel. Control trials consisted of vowels in matching consonantal contexts across pairs and contained either two CVC minimal pairs (i.e., CVC-CVC$_{\text{[SPECTRALLY-IDENTICAL VOWELS]}}$ vs CVC-CṼC$_{\text{[SPECTRALLY-DIFFERENT VOWELS]}}$) or two CVN minimal pairs (i.e., CṼN-CṼN$_{\text{[SPECTRALLY-IDENTICAL VOWELS]}}$ vs CVN-CṼN$_{\text{[SPECTRALLY-DIFFERENT VOWELS]}}$). Since coda consonant context is held constant in control trials, listeners should display the greatest perceptual sensitivity to differences in vowel nasality across vowels. Test trials featured differing consonantal contexts within-pair, where spectrally-identical pairs contain either oral vowels (i.e., CVC-CVN$_{\text{[SPECTRALLY-IDENTICAL VOWELS]}}$ vs CVC-CṼN$_{\text{[SPECTRALLY-DIFFERENT VOWELS]}}$) or nasalized vowels (i.e., CṼC-CṼN$_{\text{[SPECTRALLY-IDENTICAL VOWELS]}}$ vs CVC-CṼN$_{\text{[SPECTRALLY-DIFFERENT VOWELS]}}$). In these trials, full compensation for coarticulation would result in an at-chance performance. An above-chance performance for test trials would indicate partial compensation.

Test and control trials occurred across five different *f*0 conditions: small congruent (this is the typically-used trial type in prior studies, cf. Beddor and Krakow, 1999, Zellou, 2017), large congruent, neutral, small incongruent, and large incongruent. The configuration of *f*0 contours across these conditions is presented in Table 1. Small congruent trials feature a congruency between pairs that are spectrally identical and pairs with identical *f*0 contours. These trials featured a 10 Hz difference in *f*0 contour for the spectrally-different vowels, potentially biasing listeners toward identifying these vowels as different. This *f*0 differences reflects the small (but perceptible, Jongman *et al.*, 2017) amount of variation across the natural vowels, and represents the sort of uncontrolled *f*0 variation often included in this experimental paradigm. The incongruent condition featured the opposite pattern: the spectrally-identical vowels contain different *f*0 contours, while the spectrally-different vowels were given identical *f*0 contours. In these trials, a reliance on *f*0 would result in a reversal of the discrimination performance seen in the congruent condition: a greater likelihood of identifying the spectrally-different vowels as similar. In the

Table 1. Schematic of *f*0 conditions across trials. The underlined vowel represents the spectrally-different vowel in the trial. Each row represents a different *f*0 condition. The bolded vowel represents the vowel with the different *f*0 contour. Numbers indicate initial and final *f*0 values, in Hertz. In congruent trials, different *f*0 contours align with spectrally-different vowels, while in incongruent trials spectrally-identical vowels are given differing *f*0 contours.

	Pair 1		Pair 2	
	CVC	CVN	CVC	CṼ_N
Large congruent (spectral and larger *f*0 differences align)	145–100	145–100	145–100	**135**–**110**
Small congruent (spectral and natural *f*0 differences align)	135–100	135–100	135–100	**135**–**110**
Neutral (spectral differences only; *f*0 constant across vowels)	135–105	135–105	135–105	135–105
Small incongruent (spectral and natural *f*0 differences mismatch)	**135–110**	135–100	135–100	135–100
Large incongruent (spectral large *f*0 differences mismatch)	**135–110**	145–100	145–100	145–100

neutral condition, listeners cannot use *f*0 information to distinguish vowel pairs at all, and so any compensation seen in this condition will be independent of information related to *f*0 contour. Large congruent and large incongruent conditions contain the same patterns as their small counterparts, but the difference in *f*0 is doubled from 10 to 20 Hz. There were two motivations for generating two levels of pitch variation: First, the large conditions we included in case the original condition was too small. Second, if there is an acoustic effect of *f*0 variation on discrimination patterns, the effect should be linear, which can be seen by adding an additional condition.

If information related to *f*0 contour guides discrimination, then listeners will tend to identify the nasal-oral pair (i.e., the spectrally-different pair; pair 2 in Table 1) as dissimilar in congruent trials and similar in incongruent trials. The overall prediction in this scenario is a pattern of results where "compensation" decreases across *f*0 conditions as in large congruent > small congruent > neutral > small incongruent > large incongruent. On the other hand, if listeners consistently identify the nasal-oral pair as dissimilar across *f*0 conditions, this would indicate that listeners are not relying on source information, suggesting this paradigm is robust to the sorts of small differences in *f*0 seen across natural productions.

Listeners were presented with stimuli over headphones in a sound-attenuated booth and asked to indicate the acoustically-dissimilar vowels. Participants were randomly presented trials in an experimental block that included the 2 possible orderings of the pairs for each of the 5 *f*0 conditions across control and test trials (2 minimal pairs × 2 control trial types + 2 test trial types × 5 *f*0 variation conditions × 2 pair orderings = 80 trials). Two presentations of this block resulted in a total of 160 discrimination trials per participant. Interstimulus intervals (ISIs) within pairs was 500 ms; ISI between pairs (between second and third items) was 750 ms; inter-trial interval was 2 s. Listeners were instructed to determine whether the first or second pair of words contained vowels that sounded most different (binary choice: first pair or second pair), which they indicated via a button press on a response box. Experimental sessions lasted approximately 30 min in length.

3. Results

Results were analyzed using a multilevel Bayesian logistic regression model, fit in R using the *brms* package (Bürkner, 2017). The model predicted "accurate" responses, where these were defined as identifications of the spectrally-different vowel-pairs as being more different. Accuracy was modeled as a function of two categorical predictors and their interaction: trial type (test vs control) and *f*0 condition (large congruent, small congruent, neutral, small incongruent, large incongruent). Random effects were included for all effects and interactions. Although this design focuses on the prediction of correct responses, the random by-subject intercepts control for differences in listener response bias, so that the effects for trial type and *f*0 condition reflect differences in listener sensitivity to the experimental manipulations (DeCarlo, 1998)

Rather than dichotomous hypothesis testing based on *p*-values, Bayesian inference relies on estimating the magnitude and uncertainty of different effects estimates. Credible intervals of parameter values, roughly analogous to confidence intervals, can be established using highest-density intervals (HDIs). For example, the 95% HDI of a parameter indicates that a parameter has a 95% probability of falling within a certain range given the data and model structure. HDIs can also be used to estimate the differences between different linear-combinations of parameters (i.e., contrasts). An attractive quality of Bayesian models is that they allow us to effectively "accept the null hypothesis" by finding that the most likely value for a parameter (given the data and model structure) is either zero, or small enough to not have any practical importance to outcomes (Kruschke, 2011).

Figure 1 presents 95% HDIs of predicted accurate performance classification rates in different listening conditions. There was a consistent difference in performance between control trials (mean = 64.8%, 95% HDI = [60.3%, 68.9%]) and test trials (mean = 56.9%, 95% HDI = [52.3%, 61.3%]), with 7.9% more accuracy in the control trials (95% HDI = [4.2%, 11.8%]). In contrast, there was no consistent difference across *f*0 conditions, with expected accuracy varying only from 59.9% at its lowest (small incongruent) to 61.7% at its highest (small congruent) and large amounts of uncertainty in all effect estimates. The trial type by *f*0 condition interaction also appears to consist primarily of noise, with no discernible pattern on the basis of *f*0 differences. As a result, there is no evidence that the similarity of *f*0 contours between and across stimulus pairs influences listeners' discrimination of vowel sounds across pairs in this paradigm.

4. Discussion

Paradigms that explore the phenomenon of partial perceptual compensation for coarticulation often do not control for secondary acoustic variation in the speech signal, above and beyond the presence of nasalization on vowels adjacent to nasal consonants. Pitch, in particular, is a feature

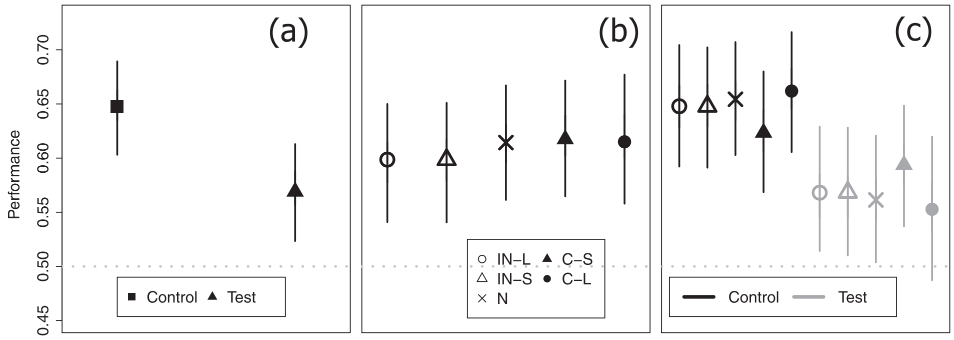

Fig. 1. Points indicate means, lines indicate 95% HDIs for the estimated probability of observing a different response in listening conditions: (a) in control vs testing conditions; (b) across *f*0 condition levels: large incongruent (IN-L), small incongruent (IN-S), neutral (N), small congruent (C-S), large congruent (C-L); (c) *f*0 condition levels individually across each trial type.

that can vary across oral and nasalized vowels for a variety of factors and is something that people are highly sensitive to. The current study investigated whether discrimination patterns used to infer partial perceptual compensation for nasal coarticulation is influenced by *f*0 differences across vowels. Since paired discrimination is a common paradigm by which perceptual compensation is explored, it is appropriate to examine what contribution differences in pitch across vowels in a trial have on discrimination performance. To that end, stimuli pairs that contained either congruent or incongruent coarticulatory and pitch differences across oral and nasalized vowels in appropriate coarticulatory contexts were generated. Across the five *f*0 conditions, it was found that listeners' discrimination performance did not reliably differ. In all conditions, listeners display discrimination performance indicating that they retain some residual acoustic information about the coarticulatory nasalization on the vowel and that guides their above-chance performance. In other words, listeners appear to display perceptual sensitivity to differences across vowels based on coarticulatory cues, and not based on pitch cues. This result replicates prior work finding partial perceptual compensation for coarticulation (Fowler, 2006; Beddor and Krakow, 1999, *inter alia*). Furthermore, the finding in the current study reinforces findings from prior work by displaying that this phenomenon is robust to pitch variation across vowels.

Why is it the case that pitch cues do not bias listeners' partial compensation patterns? One possibility is that the task is vaguely defined; listeners were asked to identify different-sounding vowels and in this task their responses indicate they attend to residual coarticulatory details and not pitch. A different result might be expected if listeners are asked to attend to pitch differences. Furthermore, it was an intentional decision to present listeners with small pitch differences (even in the "large" conditions). One prediction might be that pitch differences that are even larger would lead to a pitch bias in this task. Given that the nature of the task was vague and listeners' default appears to be to attend to coarticulatory details, this supports perspectives that hold that nasalization is informative, meaningful, and helpful in making linguistic decisions (Beddor *et al.*, 2013; Beddor, 2009; Scarborough and Zellou, 2013). Pitch is less helpful in determining the word identity, so in this task, listeners appear to rely on it less. So, listeners' sensitivity to, and retention of, coarticulatory details is functional.

The general finding that listeners exhibit partial perceptual compensation has at least two important theoretical implications. First, listeners' perceptual sensitivity to and use of coarticulatory variation suggests that these details can be encoded in long-term cognitive representations. This is empirically supported by findings of cross-linguistic differences in patterns of produced coarticulatory variation that are, subsequently, found to be mirrored in language-specific patterns of perceptual compensation. For instance, native English, native Thai (Beddor and Krakow, 1999), and native Shona (Beddor *et al.*, 2002) speakers compensate for coarticulation to different extents, reflecting the patterns of coarticulatory variation present in their native languages. Second, the veridical perception of coarticulatory nasalization is a condition that has been argued to be a precursor to phonological change. A listener-based model of sound change (Ohala, 1993) posits that failure to fully attribute coarticulatory variation to its source is one mechanism for grammatical reinterpretation of the speech signal. For example, if a listener veridically hears coarticulatory nasalization on a vowel, one possible reanalysis is that the phonetic property was an intended and inherent aspect of the vowel itself. The finding that partial compensation for coarticulation is robust to pitch variation further underscores the importance of coarticulatory variation in models of speech representations and theoretical claims about the mechanisms of sound change.

Finally, we note that there are several other experimental checks that could be carried out to validate some of the methodological conventions used in this paradigm. For example, because investigations into perceptual compensation typically rely on cross splicing (e.g., Beddor and Krakow, 1999; Fowler, 2006; Zellou, 2017), it is possible that this affects listener responses in (currently) unknown ways. In addition, the reliance on the 4IAX paradigm may be providing a narrow view of the phenomenon. To keep our findings as similar as possible to prior work, we utilized similar stimuli and experimental design methods as prior work. Future work could include different paradigms and different stimuli manipulation methods to explore the effect these variables have on behavioral responses.

References and links

Beddor, P. S. (**2009**). "A coarticulatory path to sound change," Lang. **85**(4), 785–821.

Beddor, P. S., Harnsberger, J. D., and Lindemann, S. (**2002**). "Language-specific patterns of vowel-to-vowel coarticulation: Acoustic structures and their perceptual correlates," J. Phonetics **30**(4), 591–627.

Beddor, P. S., and Krakow, R. A. (**1999**). "Perception of coarticulatory nasalization by speakers of English and Thai: Evidence for partial compensation," J. Acoust. Soc. Am. **106**(5), 2868–2887.

Beddor, P. S., McGowan, K. B., Boland, J. E., Coetzee, A. W., and Brasher, A. (**2013**). "The time course of perception of coarticulation," J. Acoust. Soc. Am. **133**(4), 2350–2366.

Boersma, P., and Weenink, D. (**2005**). Praat. Doing phonetics by computer (Version 5.1).

Bürkner, P. C. (**2017**). "brms: An R package for Bayesian multilevel models using Stan," J. Stat. Software **80**(1), 1–28.

DeCarlo, L. T. (**1998**). "Signal detection theory and generalized linear models," Psychol. Methods **3**(2), 186–205.

Fowler, C. A. (**2006**). "Compensation for coarticulation reflects gesture perception, not spectral contrast," Percept. Psychophys. **68**(2), 161–177.

Fowler, C. A., and Brown, J. M. (**2000**). "Perceptual parsing of acoustic consequences of velum lowering from information for vowels," Percept. Psychophys. **62**(1), 21–32.

Hajek, J. (**1997**). *Universals of Sound Change in Nasalization* (Blackwell, Oxford, United Kingdom), Vol. 31.

Jongman, A., Qin, Z., Zhang, J., and Sereno, J. A. (**2017**). "Just noticeable differences for pitch direction, height, and slope for Mandarin and English listeners," J. Acoust. Soc. Am. **142**(2), EL163–EL169.

Kawasaki, H. (**1986**). "Phonetic explanation for phonological universals: The case of distinctive vowel nasalization," in *Experimental Phonology*, edited by J. J. Ohala and J. J. Jaeger (MIT Press, Cambridge, MA), pp. 81–103.

Kruschke, J. K. (**2011**). "Bayesian assessment of null values via parameter estimation and model comparison," Perspect. Psychol. Sci. **6**(3), 299–312.

Mann, V. A. (**1980**). "Influence of preceding liquid on stop-consonant perception," Percept. Psychophys. **28**(5), 407–412.

Ohala, J. J. (**1993**). "Coarticulation and phonology," Lang. Speech **36**(2–3), 155–170.

Scarborough, R., and Zellou, G. (**2013**). "Clarity in communication: 'Clear' speech authenticity and lexical neighborhood density effects in speech production and perception," J. Acoust. Soc. Am. **134**(5), 3793–3807.

Yu, A. C. L., and Lee, H. (**2014**). "The stability of perceptual compensation for coarticulation within and across individuals: A cross-validation study," J. Acoust. Soc. Am. **136**(1), 382–388.

Zellou, G. (**2017**). "Individual differences in the production of nasal coarticulation and perceptual compensation," J. Phonetics **61**, 13–29.