The Perceptual Time Course of Coarticulation

Patrice S. Beddor, Kevin B. McGowan, Julie Boland, and Andries Coetzee
Department of Linguistics, University of Michigan, Ann Arbor

Abstract

Listeners' moment-by-moment processing as the coarticulated signal unfolds over time was investigated. The target coarticulatory cue was anticipatory vowel nasalization. English-speaking participants' eye movements were monitored as they heard instructions to look at one of two pictured objects on a computer screen. Trials included picture pairs for naturally produced words of the form CṼNC-CVC (e.g., bend-bed), CṼNC-CṼNC (bend-bent), and CVC-CVC (bed-bet).

Results: When participants heard CṼNC (bend), they fixated the correct picture earlier when the competing picture was CVC (bed)–that is, when vowel nasality is disambiguating–than when the competitor was another CṼNC (bent). The earlier the onset of vowel nasalization, the faster the fixations to CṼNC. Moreover, participants' use of Ṽ was sensitive to the different temporal patterns of ṼN triggered by coda voicing. In contrast, for most trials, a non-nasalized V was not similarly helpful for selecting CVC over CṼNC, suggesting that listeners were not simply responding on the basis of familiarity with ṼN and VC sequences.

Conclusion: In online processing, listeners use anticipatory coarticulation, often as soon as it becomes available, in ways that indicate their precise attunement to time-varying information. The results are interpreted as supporting the theoretical perspective that listeners track the gestural information afforded by the dynamics of coarticulation as encoded in the acoustic signal.

Background

- Coarticulatory cues influence listeners' decisions in offline tasks:
  - *Identification*: Coarticulatory vowel nasalization alone, with no N, elicits N percepts from American English listeners (Malécot 1960, Beddor 2009).
  - *Gating*: When listeners hear CV fragments, they are more likely to report that the full word is CṼN (rather than CVC) if the vowel is nasalized (Lahiri & Marslen-Wilson 1991, Ohala & Ohala 1995).
  - *Reaction time*: Listeners' reaction times in identifying coda N (vs. C) are faster if the preceding vowel is nasalized (Fowler & Brown 2000).
- In real-time processing, cross-spliced, mismatched coarticulatory cues slow listeners' responses:
  - *Eye-tracking*: When listeners hear inappropriate vocalic cues to stop place, they initially fixate an image consistent with those cues and are less accurate in fixating the correct image even after they hear a disambiguating stop burst (Dahan *et al.* 2001, Dahan & Tanenhaus 2004).
  - *MEG*: Listeners' neural responses to a coda consonant as N or C are slowed when the preceding vowel has inappropriate nasality (Flagg *et al.* 2006).

Theoretical Goals

- We adopt the theoretical perspective that listeners are interactive participants in processing the input acoustic signal, using the rich, time-varying information in coarticulation to determine what speakers are saying (see also Fowler 1996, Hawkins 2003, Whalen 1984, among others).
- We test whether, as the (appropriately) coarticulated signal unfolds over time, listeners' perceptual assessments evolve in real time. Specifically:
  - Do English-speaking listeners use coarticulatory Ṽ as information about an upcoming N as soon as Ṽ becomes available?
  - Do listeners use the the precise, time-varying cues (such as different ṼN patterns in different voicing contexts) in online processing?
  - Is oral V similarly informative about an upcoming C as soon as V is heard?

The Perceptual Time Course of Coarticulation

To address these questions, and to more generally probe online processing of gestural information encoded in the acoustics of coarticulation, we measured the time course of perception of coarticulatory nasalization using eye tracking.

Method: Tracking eye movements to audio-visual stimuli

Stimuli

T-D	T-NT	D-ND	NT-ND
bet - bed	bet - bent	bed - bend	bent - bend
watt - wad	watt - want	wad - wand	want - wand
set - said	set - scent	said - send	scent - send
let - lead	let - lent	lead - lend	lent - lend
wet - wed	wet - went	wed - wend	went - wend

- *Visual:* Pencil & charcoal pictures of each word.
- *Audio:* Manipulated natural-speech recordings presented over headphones. Each CṼNT-CVT and CṼND-CVD trial had 3 types: light Ṽ, heavy Ṽ, and Deleted-N.

Manipulation 1: ṼN stimuli with two degrees of vowel nasalization:

- *Light nasalization*: 40% nasal vowel, 60% oral
- *Heavy nasalization*: 80% nasal vowel, 20% oral

Manipulation 2: Heavy Ṽ (80% nasal V) but no N ([n] deleted from bent, bend, want, wand, etc.)

Method Cont: Eye-tracking task

Participants

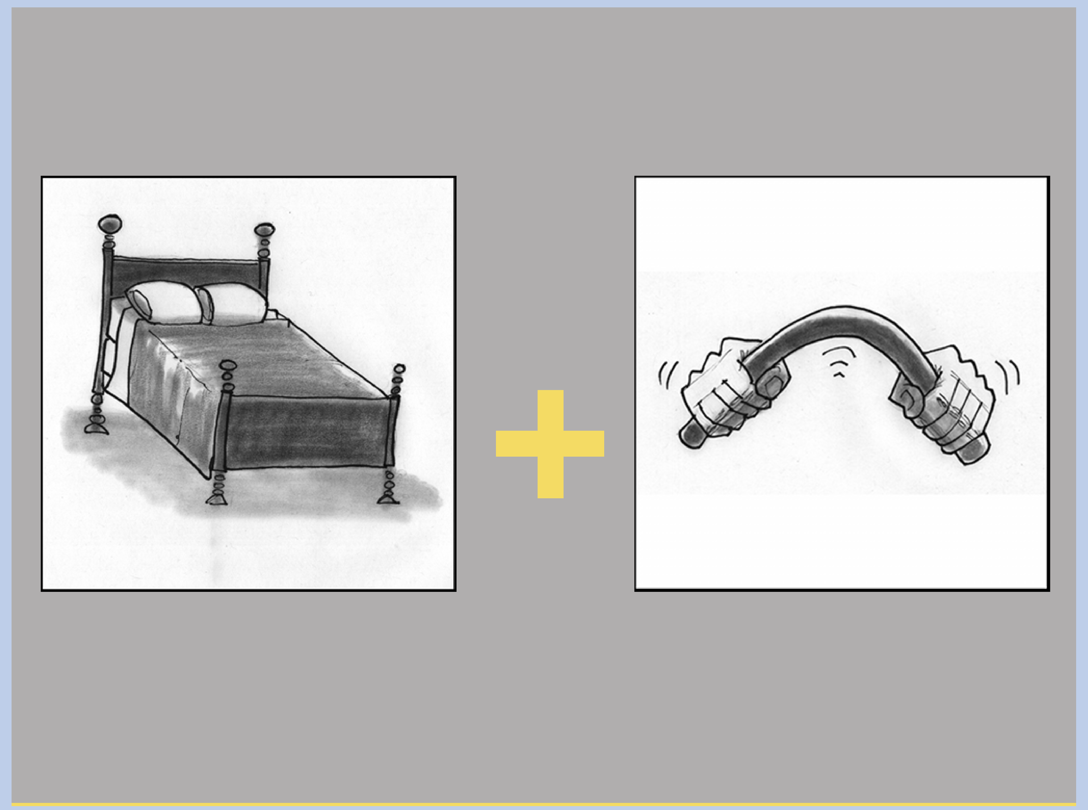
25 native English-speaking undergraduate students at the University of Michigan participated in 2 separate 1 hour sessions each.

Picture familiarization Participants learned the name of each picture prior to the main task.

Audio

For each trial, participants heard:

- Look at the pictures (e.g., *bed* & *bend*);
- Fixate cross;
- Now look at [target] (e.g., *bend*. Cross disappears at onset of word.)



Predictions

- **Hypothesis 1:** Ṽ signals an upcoming N.

Auditory CṼNC

1a. Initial correct fixations to CṼNC (bent) should occur
earlier for CṼNC-CVC (bent-bet)
than for CṼNC-CṼNC (bent-bend)
because Ṽ is not a differentiating cue for CṼNC-CṼNC.

1b. In CṼNC-CVC (bent-bet) trials, initial correct fixations should occur earlier for CṼNC with heavy nasalization than for CṼNC with light nasalization.

- **Hypothesis 2:** Presence of information for a coarticulatory gesture is more informative than its absence. (Listeners are not simply responding on the basis of familiarity with ṼN and VC.)

Auditory CVC

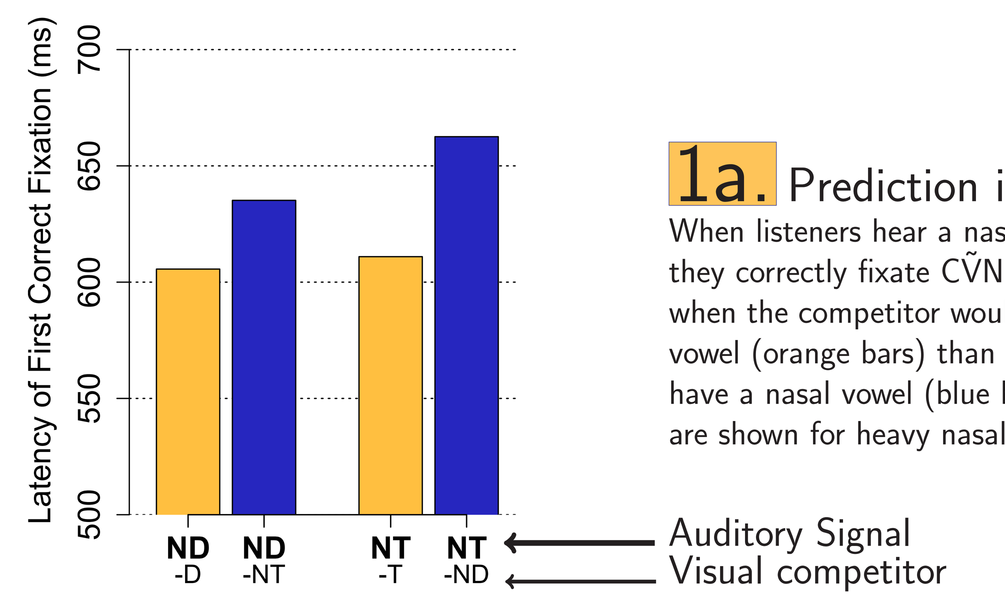
2. Initial correct fixations to CVC (bet) should occur *about the same time* in CVC-CVC (bet-bed) and CVC-CṼNC (bet-bent) trials.

- **Hypothesis 3:** English-speaking listeners use Ṽ in voiceless (bent) more than in voiced (bend) contexts because Ṽ tends to be heavily nasalized and N is often absent in $\tilde{V}NC_{voiceless}$ (e.g., Malécot 1960, Cohn 1990).

Voicing

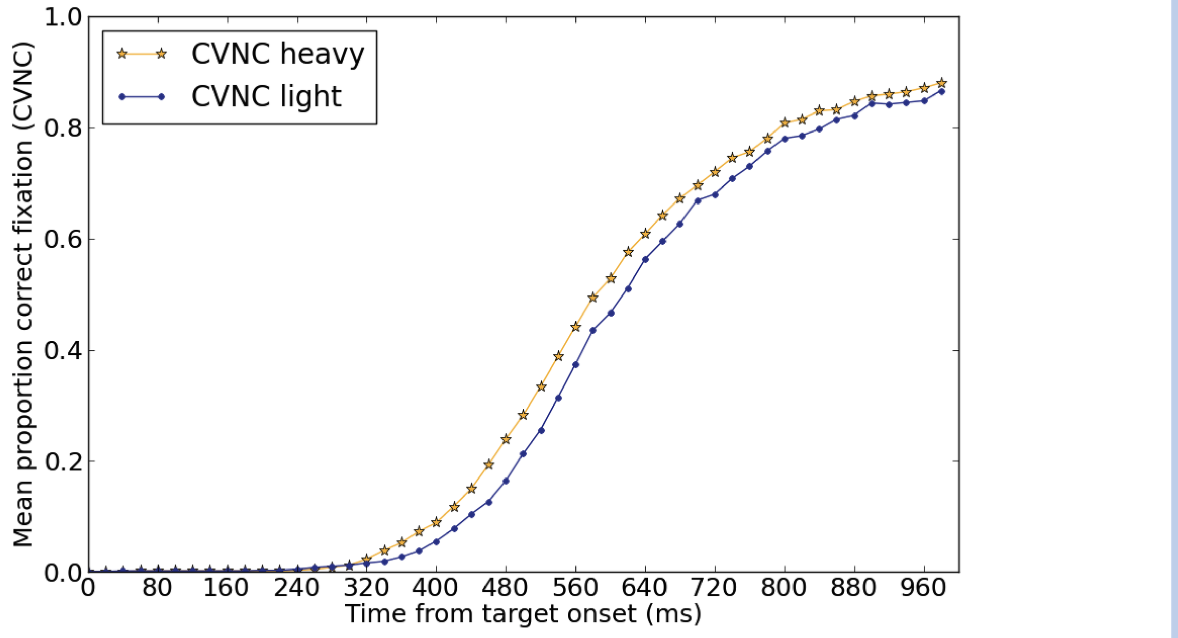
3. Fixations to Deleted-N trials should be less accurate, and show a greater decrease over time, for CṼ(N)D than for CṼ(N)T.

Results: Prediction 1



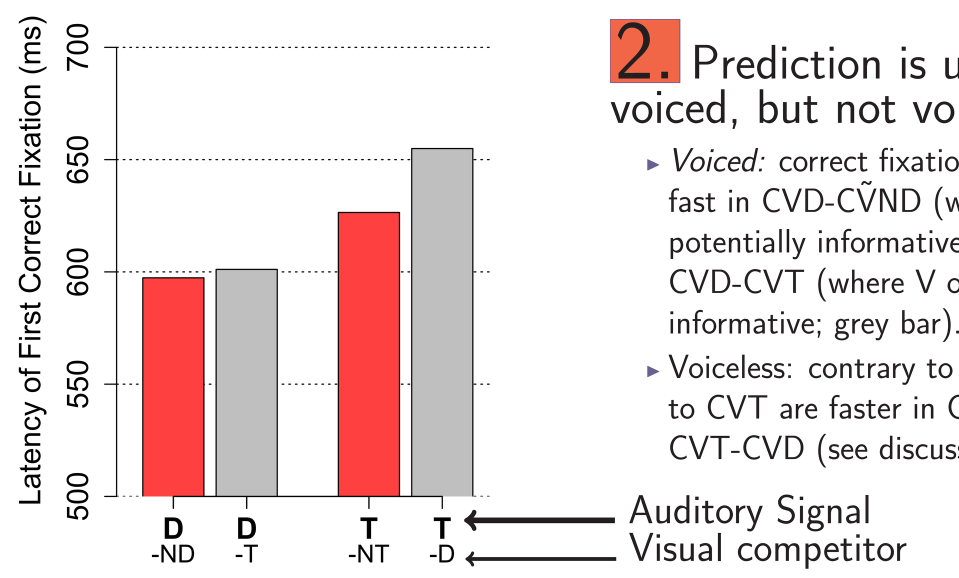
1a. Prediction is upheld: When listeners hear a nasalized vowel, they correctly fixate CṼNC image faster when the competitor would have an oral vowel (orange bars) than when it would have a nasal vowel (blue bars). Results are shown for heavy nasalization.

1b. Prediction is upheld: In CṼNC-CVC (bent-bet) trials, the more nasalized the vowel, the earlier listeners correctly fixate the CṼNC image.



Note: a 200ms delay is generally assumed for programming an eye movement (e.g., Dahan *et al.* 2001).

Results: Prediction 2

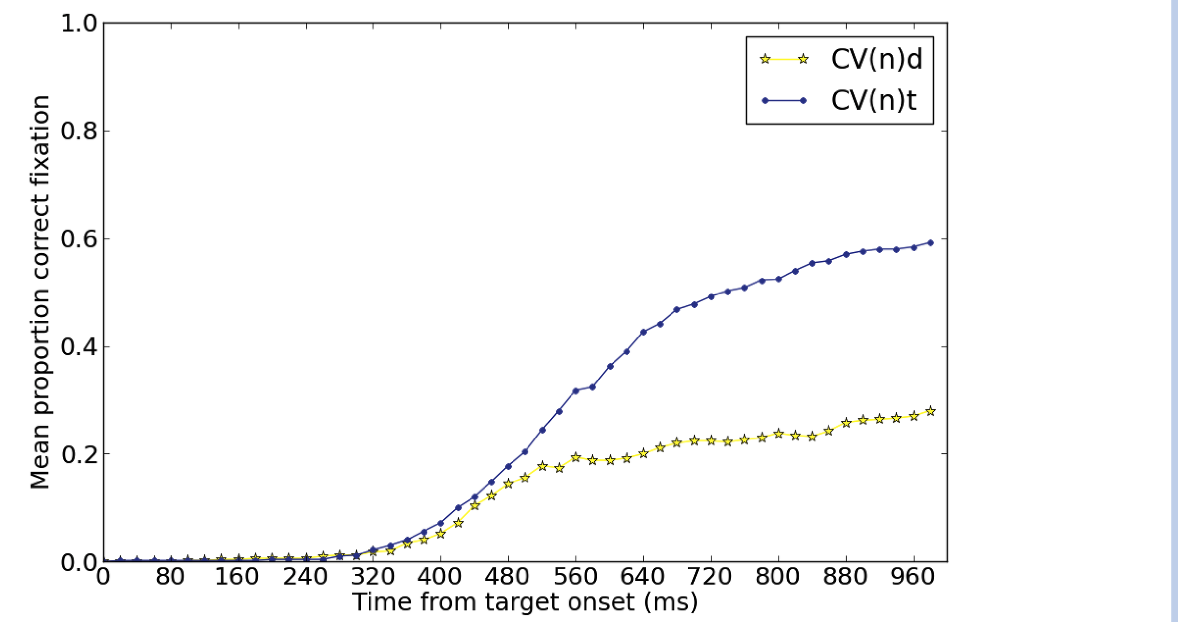


2. Prediction is upheld for voiced, but not voiceless, trials:

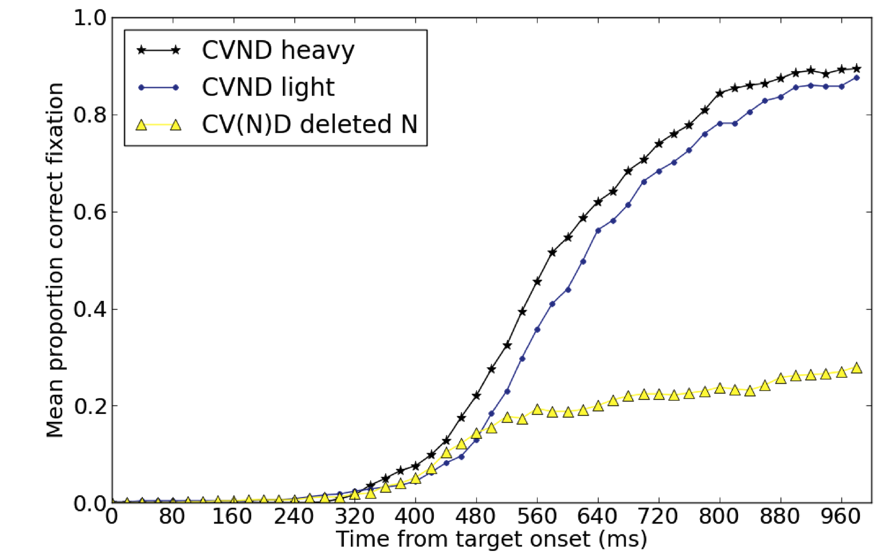
- *Voiced:* correct fixations to CVD are equally fast in CVD-CṼND (where oral V is potentially informative; red bar) and CVD-CVT (where V orality is not informative; grey bar).
- Voiceless: contrary to predictions, fixations to CVT are faster in CVT-CṼNT than in CVT-CVD (see discussion).

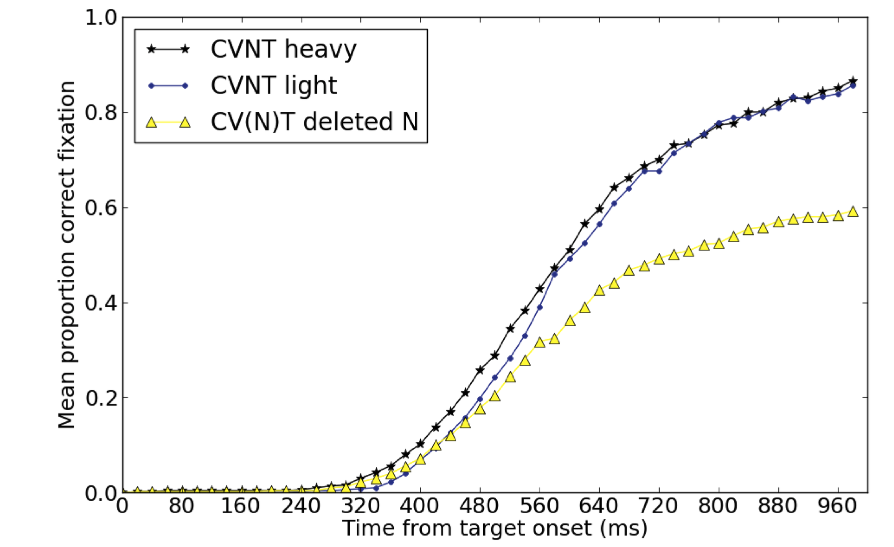
Results: Prediction 3

3. Prediction is upheld: Fixations to Deleted-N trials are overall less accurate and show a greater decrease over time to CṼ(N)D than to CṼ(N)T.



Results: Summary (voiced & voiceless)





Discussion

- **Listeners use Ṽ to plan their eye movement.**

CṼNC

CṼNC was fixated earlier when the vowel in the competitor was expected to be non-nasal than when it would be nasal. The more nasalized the Ṽ, the earlier CṼNC was fixated.

- **Listeners are less likely to use V.**

CVC

CVD was not fixated earlier when the vowel in the competitor was expected to be nasal than when it would be non-nasal. For CVT, oral V did lead to earlier fixations in CVT-CṼNT than in CVT-CVD comparison. The voicing effect is due in part to vowel quality differences (in watt [wat]-want [wɔ̃nt] but not wad [wad]-wand [wãnd]) distinct from nasality.

- **Listeners' use of context-specific coarticulatory Ṽ patterns evolves as the signal unfolds.**

Voicing

In production, English has heavier vowel nasalization and shorter N in ṼNT than ṼND. In perception, Deleted-N (with heavy Ṽ) elicits more fixations to CṼNT than to CṼND. That V is useful in CVT but not CVD may also be partly due to heavy nasalization in ṼNT: non-nasal V is useful only when the competitor would be expected to have especially heavy nasalization.

Conclusion

Coarticulatory cues speed the time course of lexical activation. That listeners actively use these cues as the signal unfolds, in ways sensitive to context-specific time-varying properties, is contrary to models in which coarticulation does not facilitate and even impedes perception (e.g., Lindblom 1990, Tatham & Moreton 2006), and supports the theoretical perspective that listeners closely track the gestural information afforded by the dynamics of coarticulation.

References

Beddor, P. S. 2009. A coarticulatory path to sound change. *Language* 85.785–821.

Cohn, A. C. 1990. Phonetic and phonological rules of nasalization. In *UCLA Working Papers in Linguistics*, volume 76. University of California at Los Angeles.

Dahan, D., Magnuson, J.S., Tanenhaus, M.K., & Hogan, E.M. 2001. Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language & Cognitive Processes* 16.507–534.

——, & Tanenhaus, M.K. 2004. Continuous mapping from sound to meaning in spoken-language comprehension: immediate effects of verb-based thematic constraints. *Journal of Experimental Psychology: Learning, Memory, & Cognition* 30.498–513.

Flagg, E. J., Cardy, J.E.O., & Roberts, T.P.L. 2006. Meg detects neural consequences of anomalous nasalization in vowel¿consonant pairs. *Neuroscience Letters* 397.263–268.

Fowler, C A. 1996. Listeners do hear sounds, not tongues. *Journal of the Acoustical Society of America* 99.1730–1741.

——, & Brown, J. M. 2000. Perceptual parsing of acoustic consequences of velum lowering from information for vowels. *Perception and Psychophysics* 62.21–32.

Hawkins, S. 2003. Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics* 31.373–405.

Lahiri, A., & Marslen-Wilson, W. 1991. The mental representation of lexical form: a phonological approach to the recognition lexicon. *Cognition* 38.245–294.

Lindblom, B. 1990. Explaining phonetic variation: a sketch of the H&H theory. In *Speech Production and Speech Modelling*, ed. by W. J. Hardcastle & A. Marchal, 403–439. Kluwer Academic.

Malécot, A. 1960. Vowel nasality as a distinctive feature in American English. *Language* 36.222–229.

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: Papers in Laboratory Phonology IV*, ed. by B. Connell & A. Arvaniti, 41–60. CUP.

Tatham, M., & Moreton, K. 2006. *Speech Production and Perception.*. Palgrave.

Whalen, D.H. 1984. Subcategorical phonetic mismatches slow phonetic judgments. *Perception and Psychophysics* 35.49–64.

e-mail: beddor@umich.edu or clunis@umich.edu