The Relation between Language Users' Perception and Production Repertoires

Patrice Speeter Beddor

University of Michigan
beddor@umich.edu

ABSTRACT

Understanding the relation between perceived and produced speech for individual language users is important for theories of speech perception and production, and for theories of the phonetic bases of sound change. The research program described in this paper is guided by the hypothesis that a language user's perception and production repertoires are complexly related in ways mediated by wide-ranging (phonetic, phonological, socio-cognitive, other) factors. One aim of this program is to test the long-standing assumption, particularly in the sound change literature, that the phonetic variants that a language user finds especially salient or useful in perception are manifested in that individual's productions. Experimental findings from studies of both stable patterns of coarticulatory variation and sound change in progress are reported that provide some support for this assumption. However, aspects of this work remain in the early stages; ongoing work more systematically explores the interacting factors that influence the production-perception relation.

Keywords: coarticulation, nasalization, tonogenesis, sound change, perception-production relation

1. INTRODUCTION

The relation between speech production and perception is a foundational issue for phonetic theories. Theories of speech production seek to explain how linguistic forms are physiologically realized by the speaker through the controlled, coordinated activities of the articulatory system. Theories of speech perception aim to explain how listeners interpret the resulting acoustic signal as linguistic information. These theories must also account for effective communication: although the correspondences among articulation, acoustics, and percept are complex, speakers produce articulations with acoustic consequences that convey the planned linguistic forms to listeners. Consequently, many theorists postulate a necessarily tight link between production and perception, arguing for parity between the forms of speaking and listening (e.g., [12], [23], [34]).

For some researchers, the postulated close relation is not limited to the requirement of sufficient similarity between signals produced by the speaker and received by the listener, but extends as well to the production-perception relation within the individual language user. Motor Theorists, for example, argue that perceiving speech involves recruitment of the listener's motor system ([10], [22]). Most approaches to exemplar theory assume that a speaker's productions are drawn from exemplar clouds seeded by that individual's perceptual input; these models often assume a perception-production loop in which the phonetic details of the perceived input are reflected in production (e.g., [18], [33]).

The relation between perceived and produced speech for an individual language user is also a central issue for the study of the phonetic bases for sound change. Contemporary experimental approaches to sound change investigate how the phonetic variants in the ambient language might serve as a source of new sound patterns that spread through a speech community. Approaches to sound change in which listeners' selective attention to certain variants arguably contributes to sound change assume, either explicitly or tacitly, that the variants that are particularly salient in perception will be realized in an innovative listener-turned-speaker's own subsequent productions ([3], [15], [25], [29], [30], [39]).

Thus, diverse theoretical perspectives converge in the study of the nature of the link between a speaker-listener's articulatory and perceptual repertoires or grammars. Underlying this study is the solid understanding, grounded in decades of research, that these repertoires differ across language users. Speakers differ, for example, in their strategies for coordinating articulatory movements (e.g., [13], [20], [24], [26]). Listeners make use of the rich, time-varying information in the input, but their use of, and adjustments for, this information is imperfect ([4], [9]) and listener-specific ([2], [17], [38]).

Do a language user's articulatory strategies correlate with that user's particular perceptual weightings? In ongoing research in our laboratory, we are especially interested in this question from the perspective of the role of individuals in the inception of sound change. We hypothesize that a language user's perceptual weightings will be manifested in their productions. This hypothesis does not mean, of course, that we expect a precise mirror between perception and production. Rather, considerably greater malleability is expected for perception: flexibility in perception is essential to comprehension, whereas non-accommodation in production typically need not impair intelligibility ([11], [32]). Findings from speech imitation tasks, for example, which show that many listeners-turned-speakers weakly approximate characteristics of a model speaker's productions (e.g., [1], [16], [21], [28]), are strongly suggestive of a perception-production link, but one that is variable and subject to numerous influences.

My current research program narrows the broad issue of the nature of the link between language users' articulatory and perceptual repertoires to a specific question: do language users who *produce* more innovative variants also weight the innovative property more heavily in *perception*? In this paper, I illustrate this work through two in-progress projects being conducted in collaboration with several colleagues. One project explores a relatively stable pattern of coarticulatory variation in American English and the other studies a sound change in progress in Afrikaans.

2. TIME COURSE OF COARTICULATORY VOWEL NASALIZATION IN PERCEPTION AND PRODUCTION

One component of our lab's research agenda investigates the relation between a listener's dynamic use of coarticulatory information and that language user's own coarticulated productions. Previous studies have generally failed to find a correlation between perception and production of coarticulation for individual participants. For example, participants who produce more extensive coarticulation for a particular dimension do not more accurately perceive differences between the relevant coarticulated variants ([13], [14]) nor do they exhibit larger perceptual boundary shifts for that dimension ([17]). We use a visual world eye-tracking paradigm to study perception under the expectation that, because coarticulation unfolds over time, measures of listeners' moment-by-moment use of coarticulatory information may offer new insights. In our study of vowel nasalization perceived and produced by American English speakers, we hypothesize that participants who do not use vowel nasalization to anticipate an upcoming nasal consonant, and hence who perceptually assign little weight to this coarticulatory property, will be among the speakers who produce less extensive anticipatory nasalization.

2.1. Nasal perception

The perception study in this ongoing project is a (partial) replication of our previous experiments that monitored eye movements as participants heard words with coarticulatory vowel nasalization ([5]), but with a new group of participants whose production is also assessed. Test stimuli in this eye-tracking task are four CV(N)C quadruplets that differ in final voicing (e.g., *bet-bed-bent-bend*; *set-said-scent-send*). Visual stimuli are line drawings corresponding to the critical words. The naturally produced auditory stimuli were manipulated to ensure that stimulus onsets were identical for critical trials and to control for nasalization which, for all words with coarticulatory [Ṽ], began 20% into the vowel. In each trial, participants saw two images on a computer monitor, as illustrated in Figure 1 for *bent-bet*, and were instructed to look at the image corresponding to the (single) word they heard. Visual pairings were of CVNC-CVC (e.g., *bent-bet*), CVNT-CVND (*bend-bent*), and CVT-CVD (*bet-bed*) images. Auditory stimuli were CṼNC ([bɛ̃nt], [bɛ̃nd]), CVC ([bɛt], [bɛd]), and CṼC ([bɛ̃t], [bɛ̃d]). The last stimulus type tests whether [Ṽ] alone, with no N, is sufficient to sustain fixation of the CVNC image.

Figure 1: Nasalization eyetracking task: Sample screen shot for the visual trial *bent-bet*.

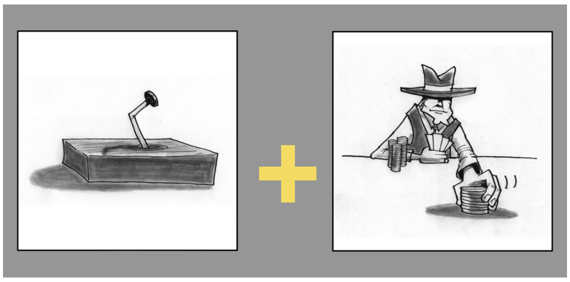
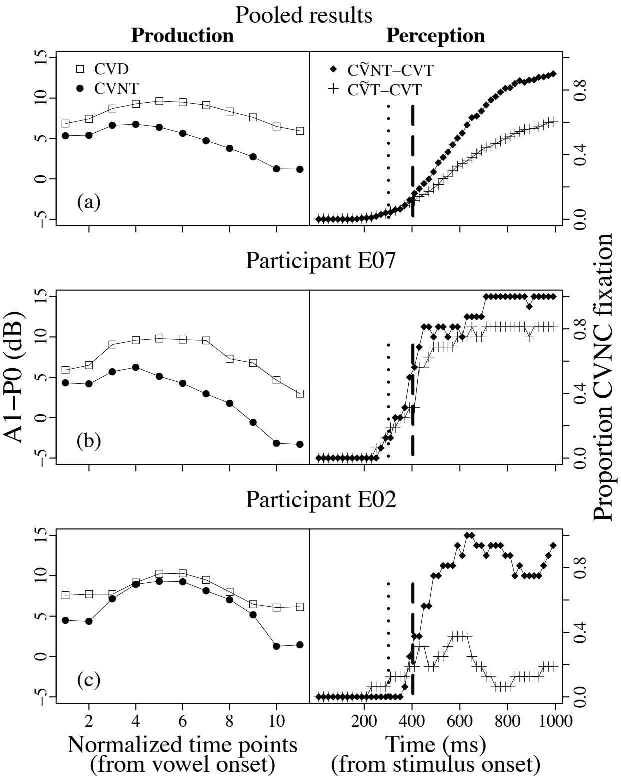


Figure 2a (right panel) gives the pooled proportion fixations over time on CVNC words of 32 listeners for auditory [CṼnt] (diamonds) and [CṼt] (pluses) trials when the competitor image was a CVT word. (For reasons of space, only voiceless results are presented. Voiced [CṼ(n)d] elicited similar results except that listeners were less likely to fixate a CVNC image in response to [CṼd] than to [CṼt].) The results replicate our earlier study ([5]). Overall, listeners began to fixate a CVNC rather than a CVC image (e.g., *bent* rather than *bet*) shortly after coarticulatory nasalization began, and prior to N onset. However, as illustrated in the lower panels of Figure 2, listeners differed in their use of [Ṽ] as information for an upcoming N. For example, on hearing [CṼ], participant E07 (Figure 2b, right) looked at the CVNT image and continued to fixate CVNT even in the absence of a following N. In comparison, the fixation patterns of participant E02 (Figure 2c, right) show more delayed looks to CVNT on hearing [CṼnt] and a much lower proportion of looks on hearing [CṼt]. For E02, [Ṽ] (with no following N) was not sufficient to sustain CVNT percepts. These illustrative differences are pervasive: as in our earlier study, listeners differed substantially and systematically in their attention to coarticulation in their moment-by-moment processing of that information.

Figure 2: Coarticulatory nasalization production (left) and perception (right) pooled across participants (a) and for two individuals (b, c). Production: A1-P0 values across the time course of oral (CVD) and nasal (CVNT) vowels. (A1-P0 decreases as nasalization increases.) Perception: proportion fixations over time on CVNT image for auditory [CṼnt] and [CṼt] trials. Short vertical lines (right panels): onset of vowel nasalization (dotted) and onset of N (or site of N deletion in [CṼt]; dashed) plus 200 ms for eye movement programming delay.



2.2. Nasal production

The same 32 participants were acoustically recorded producing multiple repetitions of the target CVNC and CVC words from the perception study, along with numerous filler words. In general, as the spatial extent of nasality (e.g., size of velopharyngeal coupling) increases, the amplitude of the low-frequency nasal formant (P0) increases and the amplitude of F1 (A1) typically decreases ([6]) for non-high vowels. In our analyses, acoustic nasality was assessed through a series of A1-P0 measures across the duration of each (non-high) vowel for each speaker, automatically calculated via a Praat script.

Figure 2a (left) gives the A1-P0 measures, pooled across participants, for the time-normalized vowels in the CVNT (circles) and CVD (squares) productions. (CVD rather than CVT is the oral comparison because vowel durations are more similar in CVD-CVNT than CVT-CVNT and because many CVT vowels had creaky voicing, creating problems for the A1-P0 calculation.) As expected for American English speakers ([35]), vowels in CVNT words were, on average, nasalized (relative to vowels in CVD) across their duration, with increasing nasalization (decreasing A1-P0) approaching N onset. However, also as expected ([2]), the coarticulatory time course differed across speakers. Participants E07 (Figure 2b, left) and E02 (Figure 2c, left) exemplify some of these differences. E02 was one of five participants whose nasal and oral A1-P0 functions did not significantly diverge (as measured in a series of *t* tests across the vowel) until the final third (or less) of the vowel; E07 was one of 18 participants with significant divergence within the first third of the vowel.

2.3. Perception-production relation

Does a listener's dynamic use of coarticulatory information in perception correlate with that language user's own coarticulated productions? Although American English coarticulatory nasalization appears to be variable but stable (e.g., the patterns observed in recent years are similar to those described by Malécot in 1960 [27]), participant E07 might nonetheless be viewed as an "innovative" listener for whom coarticulatory [Ṽ] provided sufficient information for an upcoming N. In comparison, E02 might be characterized as a more conservative listener who required the consonantal information. As can be seen in Figure 2b and 2c, the innovative listener-speaker produced temporally and spectrally more extensive anticipatory nasalization than did the conservative one. To test whether a similar perception-production relation holds more generally for this group of participants, we reduced each participant's responses to two data points. The perception measure is the average proportion CVNC fixations in the [CṼt] trials (pluses in Figure 2, right) from 220 ms to 900 ms after stimulus onset. The production measure is necessarily a preliminary one. To compare across participants, A1-P0 for CVNT vowels should ideally be normalized based on both maximum A1-P0—from CVC contexts—and minimum A1-P0—from NVN contexts—for each participant. Because we unfortunately lack NVN data in this pilot study, for each participant we calculated a single non-normalized measure that is the average of the oral (CVD)–nasal (CVNT) A1-P0 differences across time points 2-10 (Figure 2, left panels). A linear regression analysis on the production and perception measures for the 32 participants yielded R^2 = 0.103 (p < 0.07). Clearly, most of the variation in one dimension is not accounted for by the other, but the marginally significant trend in the predicted direction is suggestive that our in-progress larger-scale study (with more participants, NVN, and nasal airflow data) might yield a clearer outcome.

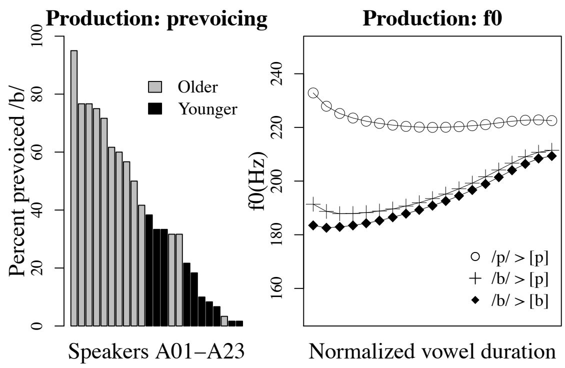
3. THE PERCEPTION-PRODUCTION RELATION IN AN ONGOING SOUND CHANGE IN AFRIKAANS

A second research strand in our lab again investigates whether information heavily weighted in perception is prominent in that language user's productions, but from the perspective of sound changes in progress. Harrington and colleagues' study [15] (see also [19]) of ongoing back vowel fronting in Southern British English, for example, showed that, compared to older speakers, younger language users not only produced more fronted vowels and showed fewer consonantal influences on those vowels but also perceptually adjusted less for consonantal effects. Our study of emergent tonogenesis in Afrikaans, led by collaborator Andries Coetzee [7], similarly compares the production and perception of older and younger speakers, although we also investigate this relation at the level of the individual language user. Of interest is that Afrikaans, which is traditionally described as contrasting prevoiced and voiceless unaspirated stops, is undergoing a change whereby phonologically voiced stops often lack prevoicing. The phonological voiced-voiceless contrast is not neutralized, but is rather maintained through the f0 contour of the post-stop vowel. We hypothesized that language users who produced particularly large voiced-voiceless f0 differences—and little to no voice onset time difference—would weight f0 more heavily in perception. We expected as well that younger participants' responses would be especially likely to show this more innovative pattern.

3.1. Voicing production

Thirteen female speakers over 40 years old and 10 female speakers under 25 years produced acoustic recordings of a large set of words beginning with oral stops and other (filler) consonants. For all /p b t d/-initial words produced by all participants, VOT and f0 across the duration of the vowel were calculated. Figure 3 (left) gives the percent phonological /b/ tokens produced with prevoicing (negative VOT) for each of the 23 speakers; on average, older speakers were three times more likely than younger speakers to produce /b/ with prevoicing.

Figure 3: Production of stop voicing contrasts in Afrikaans. Left: For each younger (black bars) and older (grey) participant, percent initial phonological /b/ (of 60 stimuli) produced with prevoicing. Right: f0 contours for vowels following /p/ (circles), devoiced /b/ (pluses), and prevoiced /b/ (diamonds), pooled across the 23 participants. (Older and younger speakers' f0 contours did not differ.)



All speakers, regardless of whether their /b/ VOTs corresponded to prevoiced [b] or voiceless [p], maintained a robust /b/-/p/ distinction in their post-stop f0 contours. Unlike VOT values, these contours did not show an age effect; older and younger speakers' f0 patterns differed in neither the temporal nor spectral extent of the influence of the preceding stop. The pooled results in Figure 3 (right) show that f0 was determined by phonological voicing: post-stop f0 is relatively high after phonological /p/ (circles) and low after both prevoiced [b] (diamonds) and voiceless [p] (pluses) realizations of phonological /b/. (Alveolars showed the same patterns for VOT and f0.) Thus, at this stage of this change in progress in Afrikaans, in production f0 is stable and depends on phonological rather than phonetic voicing. VOT is variable, with younger speakers devoicing phonological voiced stops more than older speakers do.

3.1. Voicing perception

To test Afrikaans speakers' relative perceptual weighting of stop voicing and post-stop f0, the same 23 participants responded to a stop identification task in which these two properties were orthogonally varied. Two f0 continua were created, labial /bas-pas/ and alveolar /du:r-tu:r/, varying in seven equal-sized f0 steps; endpoints matched the f0 profiles of original versions of these words produced by a female native Afrikaans speaker. For each continuum, voicing was manipulated in three steps: full, reduced (1.5 sd below average intensity and duration), and no voicing, thereby creating 21 stimuli for each of the labial and alveolar series.

Figure 4 gives the identification responses of the older (upper panel) and younger (lower) participants for the labial series. (The alveolar series showed the same basic outcome.) For both age groups, when no voicing was present (pluses), varying f0 provided sufficient information for the /b/-/p/ contrast. When full (diamonds) or reduced (circles) voicing was present, listeners were, on average, less likely to identify the stop as /p/ even at the highest f0 value.

Figure 4: Perception of voicing in Afrikaans. Identification responses of older (upper panel) and younger (lower) listeners to the f0-varying /bas/-/pas/ continuum for three degrees of stop voicing (full, reduced, and none).

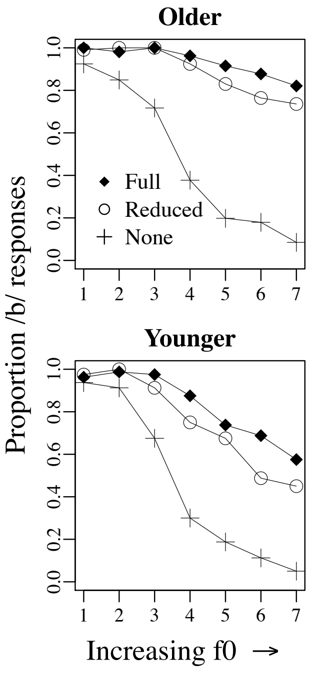


Figure 4 also shows that voicing had a (moderately) greater influence on the older participants' stop judgments. For example, across the three steps at the high f0 end of the continua (stimuli 5-7), both older and younger groups of listeners responded /b/ 15% of the time when no voicing was present. When these same stimuli were fully voiced, older and younger listeners responded /b/ 87% and 68% percent of the time, respectively.

3.1. Perception-production relation

f0 is a stable property of the Afrikaans /b/-/p/ contrast: post-stop f0 differentiates /b/ and /p/ in older and younger speakers' productions and it perceptually serves as reliable information for differentiating /b/ from /p/ for both age groups. VOT, however, is variable and age-dependent: older speakers are more likely than younger speakers to produce prevoiced /b/ and their perceptual judgments of stops are more influenced by the presence of voicing. The relation between heavy weighting of voicing in perception and greater likelihood of prevoicing in production also emerges for individual participants. For example, Figure 5 (upper panel) shows that for older participant A01, who was the most consistent producer of prevoiced realizations of /b/, the presence of voicing overrides f0 in perception: this participant judged all stimuli with full or partial voicing as voiced. In comparison, younger participant A23, who devoiced /b/ 98% of the time and who produced no VOT distinction between /b/ and /p/, perceptually judged most stimuli with high f0 as voiceless regardless of stop VOT. (Participants A01 and A23 correspond to the left- and right-most bars, respectively, in Figure 3, left.)

In assessing whether this type of perception-production relation holds more generally for the Afrikaans data, we again reduced each participant's responses to two data points. The production measure was that participant's mean VOT value for /b/. The perception measure was the percent voiced responses to the three highest f0 steps (stimuli 5-7) of the continua with full or reduced voicing. A linear regression analysis showed a modest, but significant correlation between a participant's production and perception measures (Figure 6, $R^2 = 0.16$, $p < 0.05$).

For this sound change in progress, we had hypothesized that the more innovative cue for voicing, f0, would be variable across participants. However, we have apparently missed this expected stage in the change. All participants in this study produced large f0 differences between phonological /p/ and /b/ (independent of phonetic voicing) and perceptually relied on f0 information when no voicing was present. Thus, rather than finding that language users who produce more innovative (f0) variants also heavily weight the innovative property in perception, the data instead show that the users who produce more *conservative* (voiced) variants are especially likely to attend perceptually to the conservative property.

Figure 5: Production (left) and perception (right) of the Afrikaans /b/-/p/ contrast for two participants. Participant A01 (upper panel) produced more prevoiced /b/s and attended more to voicing in perception than did participant A23 (lower panel).

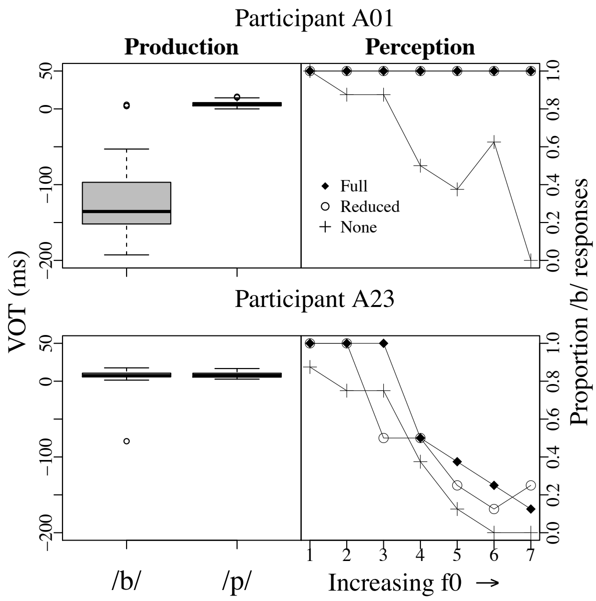
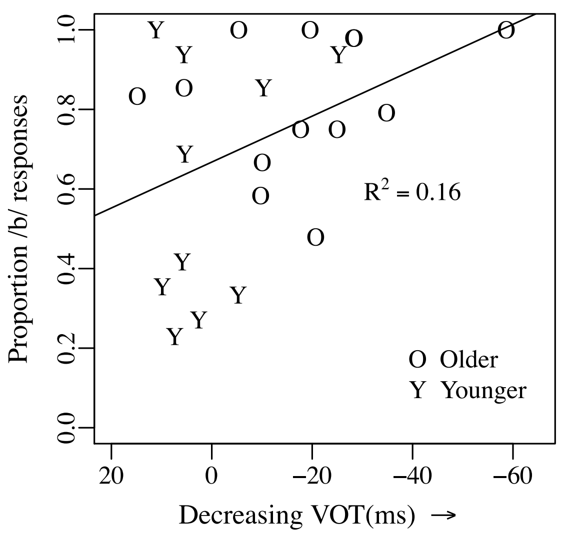


Figure 6: Production-perception relation for Afrikaans voicing. Scatterplot shows, for each participant, the relation between mean produced VOT for /b/ and percent /b/ perceptual responses to stimuli with full or reduced voicing.



4. DISCUSSION

The relation between perceived and produced speech for an individual language user is important for theories of speech perception and production, and for theories of the phonetic precursors for sound change. In my work on this relation, I am especially studying phonetic cue weighting, working under the hypothesis that the perceptual weights that listeners assign to the covarying phonetic properties for a phonological contrast will be manifested in those individuals' productions of that contrast. Although my collaborators and I remain in the early stages of this program, the results to date indicate that it is a fruitful framework to pursue. Our study of nasalization shows that, as acoustic information for nasality unfolds in real time, some American English-speaking listeners reliably use coarticulated [Ṽ] to anticipate an upcoming nasal consonant, while others do not. And there is a trend in the data we have collected and analyzed thus far for participants who are especially attentive to [Ṽ] in perception to produce more heavily coarticulated vowels. (I reiterate, though, that the preliminary production measure, while appropriate for within-speaker comparisons, may both under- and over-estimate some across-speaker differences.)

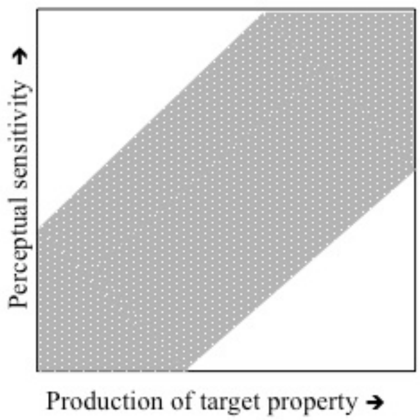
Study of emergent tonogenesis in Afrikaans shows that Afrikaans-speaking listeners were also variably attentive to phonetic information—in this case, to prevoicing information when differentiating voiced from voiceless stops. For some (mostly older) listeners, prevoicing overrode post-stop f0 cues whereas for other (mostly younger) listeners prevoicing carried little perceptual weight. The former listeners were, in their own speech, especially likely to produce prevoicing.

In an entirely separate set of studies conducted in our lab, Harim Kwon [21] has used an imitation task to investigate cue weighting in perception and production. This work tested imitation of enhanced VOT and enhanced f0 by speakers of Seoul Korean, a variety in which, similar to Afrikaans, an earlier phonation distinction is being replaced by a post-stop f0 distinction. Kwon found that the Korean-speaking participants who were especially accurate in discriminating VOT differences between Korean lax and aspirated stops were more likely to produce VOT enhancement in imitation.

4.1. Flexible perceivers

Although these findings are encouraging, continued study of listeners' selective attention and these listeners-turned-speakers' articulatory strategies will need to become more sensitive to the many factors that can influence this relation. For example, we need to build into our approach a mechanism that allows for greater variation in perception than in production. At present, in testing for a correlation between perceptual and articulatory weightings, we are effectively hypothesizing that language users will not be reliable producers but insensitive perceivers of a target property nor will they be weak or inconsistent producers but attentive perceivers of that property. That is, in the schematic perception–production space in Figure 7, our current approach tests the hypothesis that the non-shaded regions of the figure will be under-populated.

Figure 7: Schematic representation of relations between production and perception. Non-shaded, lower-right region: reliable producers but insensitive perceivers. Non-shaded, upper-left region: weak or inconsistent producers but attentive perceivers.



That reliable producers / insensitive perceivers—lower-right region of Figure 7—should be highly exceptional is consistent with our expectations (but see section 4.2 for an important caveat). However, speech perception is malleable and adaptive; listeners perceptually retune depending on phonetic context, speaker, speaking rate, novel experiences, and much more. Consequently, listeners are expected to be sensitive to information that is available in the unfolding acoustic signal and should, for reasons sketched out in the Introduction, exhibit greater flexibility in perception than in production. Therefore, inconsistent or weak producers / reliable and attentive perceivers of a target property—as in the upper-left region of the figure—are to be expected. And our findings bear this out, as clearly demonstrated for Afrikaans devoicing in Figure 6. The different perceptual responses of two younger participants in that study, both of whom produced voiceless realizations of /b/ 98% of the time (corresponding to the two right-most bars in Figure 3, left panel), provide a particularly clear example of what I judge to be an unsurprising "mismatch" between perception and production. One of these consistent devoicers (participant A23 in Figure 5, right) was perceptually insensitive to voicing: most perceptual stimuli with high f0 were heard as /p/ even when voicing was present. The other highly consistent devoicer fell at the other perceptual extreme, judging as /b/ nearly all stimuli with full or reduced voicing. (These two participants correspond to the left-most "Y"s in Figure 6.) Neither the Afrikaans devoicing nor the American English nasalization study, though, showed evidence of language users who were highly consistent producers of the target property (voicing, vowel nasalization) but who did not systematically attend to that property in perception.

4.2. Some other factors that mediate the perception-production relation

The reliability of cues for a target property likely also influences the relation between perceived and produced speech. In our study of nasalization, for example, to date we are finding that individual participants' perceptual use of [Ṽ] is related to their production of [Ṽ] in voiceless (VNT), but not voiced (VND), contexts. A possible contributing factor is that vowel nasalization in English is temporally and spatially more extensive, and the nasal consonant is shorter, in VNT than in VND sequences, making [Ṽ] a more reliable and arguably more important (for some, likely phonologized [35], [36]) property in voiceless contexts. Continued work in this area may benefit from weighting phonetic cues as a function of their reliability ([5], [37]).

Many non-phonetic factors also mediate the perception-production relation, especially when that relation is assessed from the perspective of the initiation of sound change. Study of language users' cognitive processing styles ([38, [39]) and the work of many researchers on social influences on accommodation ([31]) and sound change ([8]) all point towards influences that may heighten or attenuate the perceivers' attention or speakers' articulatory adjustments. In ongoing work in our lab, we are particularly interested in the possibility that otherwise expected links between a language user's production and perception could substantially change if the target phonetic property were socially stratefied. For example, our study of American English vowel nasalization, in which we are finding that participants who reliably use coarticulatory nasalization in perception have a tendency to produce more heavily coarticulated vowels, draws largely on a Midwestern variety in which nasalization is socially neutral. However, in Afrikaans, anticipatory vowel nasalization is a socially indexed property: speakers from different ethnic groups produce systematically different nasalization patterns. Andries Coetzee and I are in the early stages of an Afrikaans version of the nasalization study, which will allow us to compare perception and production of socially differentiated nasalization in Afrikaans to that of idiosyncratic, socially neutral patterns of nasalization in English.

5. CONCLUSION

The overarching hypothesis that motivates the research program described here is that a language user's perception and production repertoires are complexly related in ways that are mediated by both phonetic and non-phonetic factors. Within this broad agenda, a more specific aim to date has been to test the assumption in the sound change literature that the phonetic variants that an innovative—or conservative—language user finds to be especially salient in perception will be correspondingly prominent in that individual's own productions. The two studies reported here, which examine a sound change in progress (Afrikaans tonogenesis) and a relatively stable pattern of variation (American English vowel nasalization), provide preliminary support for this assumption. Continued work in this area will explore how interacting pressures (e.g., phonetic and social) combine to determine a language user's perceptual weights and articulatory targets.

6. ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation under Grant Number BCS-1348150 to Patrice Beddor and Andries Coetzee; any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation. This work reflects and depends on the essential contributions of many collaborators and research assistants, including Andries Coetzee, Anthony Brasher, Julie Boland, Kevin McGowan, Cameron Rule, Kerby Shedden, Will Styler, and Daan Wissing. I also thank Rebecca Scarborough and Will Styler, who shared their Praat script for calculating A1-P0, and numerous audiences for their valued feedback, especially colleagues in the U-M Phonetics-Phonology Research Group.

7. REFERENCES

[1] Babel, M. 2012. Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *J. Phonetics* 40, 177-189.

[2] Beddor, P.S. 2009. A coarticulatory path to sound change. *Language* 85, 785-821.

[3] Beddor, P. S. 2012. Perception grammars and sound change. In: Solé, M.-J., Recasens, D. (eds), *The Initiation of Sound Change: Production, Perception, and Social Factors*. Amsterdam: John Benjamins, 37-55.

[4] Beddor, P. S., Harnsberger, J. D., Lindemann, S. 2002. Language-specific patterns of vowel-to-vowel coarticulation: acoustic structures and their perceptual correlates. *J. Phonetics* 30, 591-627.

[5] Beddor, P. S., McGowan, K. B., Boland, J. E., Coetzee, A. W., Brasher, A. 2013. The time course of perception of coarticulation. *J. Acoust. Soc. Am*. 133, 2350-2366.

[6] Chen, M. Y. 1997. Acoustic correlates of English and French nasalized vowels. *J. Acoust. Soc. Am*. 102, 2360-2370.

[7] Coetzee, A. W., Beddor, P. S., Wissing, D. 2014. The emergence of tonogenesis in Afrikaans. *J. Acoust. Soc. Am.* 135, 2421 (abstract).

[8] Foulkes, P., Scobbie, J. M., Watt, D. 2010. Sociophonetics. In: Hardcastle, W. J., Laver, J., Gibbon, F. E. (eds), *The Handbook of Phonetic Sciences*. Malden, MA: Wiley-Blackwell, 703-754.

[9] Fowler, C. A., Brown, J. M. 2000. Perceptual parsing of acoustic consequences of velum lowering from information for vowels, *Perception & Psychophysics* 62, 21-32.

[10] Galantucci, B., Fowler, C. A., Turvey, M. T. 2006. The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review* 13, 361-377.

[11] Garrett, A., Johnson, K. 2013. Phonetic bias in sound change. In: Yu, A. C. L. (ed), *Origins of Sound Change*. Oxford: Oxford University Press, 51-97.

[12] Goldstein, L., Fowler, C. A. 2003. Articulatory phonology: a phonology for public language use. In: Schiller, N. O., Meyer, A. S. (eds), *Phonetics and Phonology in Language Comprehension and Production*. Berlin: Mouton de Gruyter, 159-207.

[13] Grosvald, M. 2009. Inter-speaker variation in the extent and perception of long-distance vowel-to-vowel coarticulation. *J. Phonetics* 37, 173-188.

[14] Grosvald, M., Corina, D. 2012. The production and perception of sub-phonemic vowel contrasts and the role of the listener in sound change. In: Solé, M.-J., Recasens, D. (eds), *The Initiation of Sound Change: Production, Perception, and Social Factors*. Amsterdam: John Benjamins, 77-100.

[15] Harrington, J., Kleber, F., Reubold, U. 2008. Compensation for coarticulation, /u/-fronting, and sound change in standard southern British: an acoustic and perceptual study. *J. Acoust. Soc. Am*. 123, 2825-2835.

[16] Honorof, D., Weihing, J., Fowler, C. A. 2011. Articulatory events are imitated under rapid shadowing. *J. Phonetics* 39, 18-38.

[17] Kataoka, R. 2011. Phonetic and cognitive bases of sound change. Ph.D. dissertation, University of California, Berkeley.

[18] Kirchner, R. 2012. Modeling exemplar-based phonologization. In: Cohn, A. C., Fougeron, C., Huffman, M. (eds), *The Oxford Handbook of Laboratory Phonology*. Oxford: Oxford University Press, 332-344.

[19] Kleber, F., Harrington, J., Reubold, U. 2012. The relation between perception and production of coarticulation during a sound change in progress. *Language and Speech* 55, 383-405.

[20] Krakow, R. A. 1989. The articulatory organization of syllables: a kinematic analysis of labial and velar gestures. Ph.D. dissertation, Yale University, New Haven, CT.

[21] Kwon, H. 2015. Cue primacy and spontaneous imitation: is imitation phonetic or phonological? Ph.D. dissertation, University of Michigan, Ann Arbor.

[22] Liberman, A. M., Mattingly, I. G. 1985. The motor theory of speech perception revised. *Cognition* 21, 1-36.

[23] Liberman, A. M., Whalen, D. H. 2000. On the relation of speech to language. *Trends in Cognitive Sciences* 4, 187-196.

[24] Lin, S. S. 2011. Perception and production of prosodically varying inter-gestural timing in American English laterals. Ph.D. dissertation, University of Michigan, Ann Arbor.

[25] Lindblom, B., Guion, S., Hura, S. Moon, S.-J., Willerman, R. 1995. Is sound change adaptive? *Rivista di Linguistica* 7, 5-36.

[26] Lubker, J., Gay, T. 1982. Anticipatory labial coarticulation: experimental, biological, and linguistic variables. *J. Acoust. Soc. Am*. 71, 437-448.

[27] Malécot, A. 1960. Vowel nasality as a distinctive feature in American English. *Language* 36, 222-229.

[28] Nielsen, K. 2011. Specificity and abstractness of VOT imitation. *J. Phonetics* 39, 132-142.

[29] Ohala, J. J. 1981. The listener as a source of sound change. In: Masek, C. S., Hendrick, R. A., Miller , M. F. (eds), *Papers from the Parasession on Language and Behavior*. Chicago: Chicago Linguistic Society, 178-203.

[30] Ohala, J. J. 1993. The phonetics of sound change. In: Jones, C. (ed), *Historical Linguistics: Problems and Perspectives*. London: Longman, 237–278.

[31] Pardo, J. S. 2006. On phonetic convergence during conversational interaction. *J. Acoust. Soc. Am*. 119, 2382-2393.

[32] Pardo, J. 2012. Reflections on phonetic convergence: speech perception does not mirror speech production. *Language and Linguistics Compass* 6/12, 753-767.

[33] Pierrehumbert, J. 2001. Exemplar dynamics: word frequency, lenition and contrast. In: Bybee, J., Hopper, P. (eds), *Frequency Effects and the Emergence of Linguistic Structure.* Amsterdam: John Benjamins, 137-157.

[34] Schwartz, J.-L., Basirat, A., Ménard, L., Sato, M. 2012. The Perception-for-Action-Control Theory (PACT): a perceptuo-motor theory of speech perception. *J. Neurolinguistics* 25, 336-354.

[35] Solé, M.-J. 1992. Phonetic and phonological processes: the case of nasalization. *Language and Speech* 35, 29-43.

[36] Solé, M.-J. 1995. Spatio-temporal patterns of velopharyngeal action in phonetic and phonological nasalization. *Language and Speech* 38, 1-23.

[37] Toscano, J. C., McMurray, B. 2010. Cue integration with categories: Weighting acoustic cues in speech using unsupervised learning and distributed statistics. *Cognitive Science* 34, 434-464.

[38] Yu, A. C. L. 2010. Perceptual compensation is correlated with individuals' "autistic" traits: implications for models of sound change. *PLoS ONE* 5(8), e11950.

[39] Yu, A. C. L. 2013. Individual differences in socio-cognitive processing and the actuation of sound change. In: Yu, A. C. L. (ed), *Origins of Sound Change: Approaches to Phonologization*. Oxford: Oxford University Press, 201-227.