Modulation of speech gestures through prosody or sound change: A commentary

MARIE K. HUFFMAN

Stony Brook University

1. Introduction

Prosodic effects are a key source of sound variation, as prosodic factors such as stress, accent and prosodic phrasing have been shown to affect the extent, duration and coordination of speech gestures. In July 2010 the LabPhon 12 conference included three talks in a session entitled "Modulation of speech gestures through prosody or sound change". Two of the talks (Choe and Redford in this issue, and Tilsen 2011) examined prosodic effects on the incidence of speech errors. Both papers show the importance of multiple levels of prosodic representation to speech planning, and argue for the importance of dynamic activation of linguistic units during speech planning. These papers reveal how such errors are a natural part of the way the speech planning process works. A variety of factors can lead to ambiguous or non-canonical outputs. In the extreme, the result is deviant enough to be labeled an error. However, there are many kinds of variation which are not heard as errors, but nonetheless are detected by listeners and stored as linguistically relevant. The third paper (Hualde et al. 2011) discusses how a specific form of variation, reduction, can lead to sound change in exactly this way. Fine phonetic analysis of ongoing voiceless stop voicing in Spanish inspires an account of sound change that brings together both Neogrammarian insights and modern views of lexical representations, exemplar theory and speech processing. Here too we see evidence of the importance of prosody, as prosodic effects appear to help drive the process by which conventionalized sound variation leads to lexical change. In this discussion I will highlight some of the leading ideas in these papers and identify promising directions for future research which they inspire.

2. Choe and Redford: The distribution of speech errors in multi-word prosodic units

Choe and Redford argue that prosodic units above the level of the word are used in the planning of speech production. The evidence comes from speech errors elicited using read tongue twisters. Choe and Redford examine the correlation between

JLP 3 (2012), 27–35
DOI 10.1515/lp-2012-0003

1868-6346/12/03–27

number of (segmental) errors and position within the utterance and within "intonation units". Intonation units are empirically observed prosodic domains that correspond roughly to intermediate phrases and intonation phrases. In these data, there are more speech errors non-initially in IUs, and marginally more errors later in the utterance, with the special exception that IU-final, utterance-final position shows fewer errors than would otherwise be predicted. Choe and Redford suggest that the relatively low number of errors IU-initially, and the accumulation of errors over the course of an IU, may be explained in terms of a type of speech planning activation gradient, in which activation of words in an IU is strong initially, but then decays. Later in the IU, activation is weak enough that "noise" (interference) from other items in the speech plan can lead to activation of the wrong form. One important contribution of this work is that it establishes prosodic domains more firmly within models of speech planning. As prosody has become so central to accounts of linguistic structure, it is critical that prosodic considerations be brought to bear within all models of linguistic planning and processing. This paper helps make the case for this shift.

An additional critical insight is the identification of activation gradiency as a source of variability in speech outputs. This is a provocative notion which accords well with views of activation in other areas of speech production (see Section 3 below). On the other hand, while Choe and Redford propose the IU as "the" unit of activation for speech planning, this does not seem justified theoretically or empirically. For one thing, given the richness of linguistic structure, it is not clear, *a priori*, why there would be one unit that is used as *the* organizing domain for speech planning. Furthermore, their data show evidence of an utterance level activation gradient. If activation contours define relevant domains for speech planning, then it would seem that the utterance must also be considered. Their Figure 3 shows that the proportion of anticipation errors is higher in final IUs, compared to non-final IUs. With less to anticipate later in the utterance, this is a somewhat surprising result. However, this result would make sense if there were a fairly constant anticipatory effect, but a gradient utterance level activation effect. In other words, if perseveration effects from already activated forms are weaker later in the utterance, then with a constant rate of anticipation, relatively more anticipation errors will occur later in the utterance, until the very utterance final position, where there is nothing more to anticipate. The relevance of the utterance as a prosodic domain in speech production has already been established in articulatory and acoustic studies (e.g., Fougeron and Keating 1997, Cho and Keating 2009). It is not a surprise then, that utterance level effects are apparent in Choe and Redford's speech error data. Given this fact, though, we must wonder whether there is actually one single unit of speech planning at all.

On the other hand, one might argue that the utterance level effect is a methodological artifact. Choe and Redford's data comes from speech errors in read sentences which were rehearsed mentally before being spoken. This may have made the utterance a more important unit in the planning of this speech than would

normally be the case. A different error distribution might be expected for natural speech based on the fact that a whole utterance is usually not fully rehearsed before it is spoken[1]. A similar concern might be raised about data like that of Fougeron and Keating (1997), which also involved read speech and utterances, the content of which was fully known before they were spoken. However, given that their results showed evidence for multiple levels of prosodic domains, we return to the argument that if several levels of prosodic structure are involved, there is no obvious principled reason to assume the others are not involved as well. The degree of activation of different prosodic domains may well differ with the sentence, the situation, and even the speaker, but there is no reason to propose an essentially different account for staged speech errors as opposed to natural ones or naturally produced error-free speech.

Finally, we address a second noteworthy trend in Choe and Redford's data, which is the observation that utterances with more IUs also had more errors. Choe and Redford propose that their speakers may have parsed more difficult utterances into smaller units (= more IUs). This is intuitively true for common tongue twisters. Consider the relative ease of saying the following parses of the Seashell tongue twister, where the vertical line indicates a prosodic break:

(1) a. She sells| sea shells| down by the | sea shore
    b. She sells sea shells| down by the sea shore
    c. She sells sea shells down by the sea shore

However, on their general account, more errors might not be expected with more domains. If activation refreshes at an IU boundary, as they suggest, then we would expect overall stronger activation of the intended items, since each IU brings strong(er) activation, and more/shorter IUs should lead to less accumulated activation decay. This in turn should mean less chance of interference from other forms in the plan, since activation will not have waned as much, which should result in *fewer* errors. This contradiction suggests something else is involved. It may be that the difficulty of the utterance led Choe and Redford's speakers to produce a difference in prosodic phrasing, as they suggest, but that the source of difficulty itself contributed to errors. The next paper we discuss may offer part of the answer. To anticipate slightly, the metrical structure of the sentences may have made some of them more error prone, for some speakers. Whether the prosodic phrasing chosen was a cause or an effect is a question for future research.

3. Tilsen: Metrical regularity facilitates speech planning and production

Tilsen (2011) presents experimental and modeling data in support of the argument that metrical regularity affects the nature of speech production gestures. Evidence from four-word nonsense utterances with similar syllabic structure but different stress patterns shows that there are more (segmental) errors when the stress

alternation is not consistent. Tilsen concludes that metrical regularity affects speech production planning by influencing the activation levels of prosodic units. The broader claim is that speech planning involves dynamical activation functions (associated with prosodic units) which can interact. When these functions are phased in such a way as to reinforce each other, activation is strengthened, leading to faster and/or more accurate retrieval from working memory. When functions interfere with each other, activation can be weakened. For the case of metrically induced speech errors, Tilsen's central claim is that irregular metrical structure leads to interference between activation functions for words (or feet) diminishing their amplitude, which can lead to either mis-selection or failure to pass threshold for execution by the speech production system. Similarly, reinforcement between functions can arise from metrical regularity.

This very interesting proposal has a number of implications. First, the extension of dynamical modeling from gestural systems (as in a task-dynamic model of articulatory phonology) to prosodic effects on speech planning brings an appealing consistency to the different components of the speech modeling enterprise. In addition, given Tilsen's claim that planning systems with these activation functions can be associated with units at "any level of the prosodic hierarchy" (p. 307), we have a new testable hypothesis, which is that interference and reinforcement effects will be seen for levels of prosodic phrasing above the word, such as the units discussed by Choe and Redford. In fact, there is some evidence in Tilsen's data suggesting that his speakers did in fact employ something like an intermediate phrase, and that this phrasing affected error rates. Tilsen reports that hesitation errors were significantly reduced when Word 2 and Word 4 had identical initial consonants, as in the examples in (2):[2]

(2) m**ee**tida **p**eetida seetida **p**eetida

Perhaps segmental similarity favored additional phrasing which provided IU-like bounding for planning system oscillation effects. In other words, the repeated forms in these sets may have led speakers to favor parsing of words as in (3):

(3) [m**ee**tida p**ee**tida] [s**ee**tida p**ee**tida]

Further support for this phrasing is the fact that transposition errors were more common between Word 2 and Word 3 than between Word 3 and Word 4. So, again, with supra-word parsing as in (4), this effect would make sense if the prosodic domains indicated by the brackets showed activation contours like that evidenced in Choe and Redford's work.

(4) [Word 1 Word 2][Word 3 Word 4]

We would probably need the additional assumption that anticipation errors are more likely than perseveration errors (as suggested by Dell et al. 1997). On this view, activation would be relatively weak on Word 2 because it is domain final,

whereas Word 3 activation would be particularly strong, being domain initial. This would predict that Word 3 would affect Word 2, and would seem to favor anticipation errors (Tilsen does not detail the directionality of error of this type). On the other hand, considering Word 3 and Word 4, the former, being domain initial, and therefore strongly activated, would resist interference from Word 4. Word 3 might affect Word 4, but again if anticipations are favored, then this effect could be weaker than the effect Word 3 could have on Word 2. While highly speculative, the most important point here is that the model and the data inspire additional questions for future research and that there seems to be promise of accord between results from research like that of Tilsen and of Choe and Redford.

A final question for both studies is the status of phrasal level prominence. Just as it is likely that Tilsen's subjects parsed the nonsense words they spoke into prosodic domains, it is also likely that they assigned relative degrees of prominence to words within these domains. Choe and Redford's subjects also had some freedom to assign prominence within IUs. Neither study addresses the question of what we may call phrasal stress, but it seems highly likely that stress matters. Turning again to the famous Seashell twister, compare the standard form, in (5a), in which a prominence falls on the second *sea*, to a modified version, in (5b) in which *see* is not accented, but a following word, *Shorty* is:

(5) a. Shé sells séashells by the séa shore
b. Shé sells séashells to see Shórty

The case in (5b) seems much *less* likely to generate errors, even though it is rhythmically very similar to (5a), and if anything is *less* metrically regular. The reason is probably that similarly prominent items are more likely to interact/interfere with each other, an effect documented for spontaneous speech errors by Fromkin (1971) and subsequent researchers.

Another interesting implication of Tilsen's approach is that it makes general predictions about how prosodic structure could affect gestural activation. These insights could in turn give us new understanding of patterns in fine phonetic detail. Particularly interesting is the notion that a more metrically irregular context might lead to reduction or, in the extreme, omission, of articulatory gestures. Tilsen suggests that if languages differ in relative metrical regularity, they might also differ in the likelihood that gestural reductions and omissions will occur. Going a step further, then, if gestural reduction is one impetus for sound change (see Section 4 below), then we come to the tantalizing idea that metrical structure may help contribute to the stability, or lack thereof, of segmental contrasts in a language. Conversely, in cases where prominence-induced gestural effects lead sound change (e.g., as Jacewicz et al. [2006] argue for English vowels), again metrical regularity may be a factor affecting the likelihood and time course of sound change. We turn now to a discussion of reduction/lenition more generally as a source of the variability that can lead to sound changes. Again, we find that prosody appears to play a crucial role.

4. Hualde, Simonet and Nadeau: Consonant lenition and phonological recategorization

Hualde et al. (2011) examine ongoing phonetic variation in Spanish as a means of gaining insight into the first steps in sound change. Some varieties of Spanish show voicing and constriction lenition in intervocalic voiceless stops. Hualde et al. examine one such variety, Majorcan Spanish. They document the fine phonetic details of this lenition process, which show that voicing produces greater phonetic overlap between contrasting stops, but does not neutralize the distinction. While a historically earlier stop voicing process shows a word boundary effect in modern Spanish, Hualde et al. show that in Majorcan Spanish, voiceless stop lenition applies intervocalically, without regard to word boundaries, suggesting that the word boundary effect is a later development. Thus, the first "change" is a general, contextually defined change, here a lenition, which produces variably voiced/lenited voiceless stops. Eventually, they argue, this type of phonetic variation could be further conventionalized as a systematic pattern of pronunciation, and later, recast as a categoricalized change in the representation. The Majorcan data, along with recent research on the processing of speech and the insights of exemplar theory, give us a clearer picture of how this might come about.

For one thing, there is compelling evidence that we can actually *hear* fine phonetic differences of the sort that Hualde et al. report. Indirect evidence is found in the growing body of literature on perceptual learning (e.g., Norris et al. 2003, Kraljic and Samuel 2007) in which exposure to phonetically ambiguous sounds in the context of a known lexical item can shift a category boundary to include more non-standard tokens of the category. Even more compelling is evidence like that reported by McMurray et al. (2009), who report that eye-gaze patterns show gradient perceptual effects of continuous VOT variation. In this work, subjects view a display of four items while hearing a word that has a stop with ambiguous VOT. Early in the stimulus, the intended word is not evident, as only later parts of the word disambiguate between the two choices (e.g., *barricade* or *parakeet*). Their data show that even in first looks to displayed items, there are gradient effects of VOT. That is, higher VOTs garner more looks to the voiceless form (e.g., *parakeet*) while lower VOTs favor *barricade*. Crucially, the subjects' inclination to look at either candidate item was significantly affected by VOT, in a linear fashion. That is, even in the range of VOT values that fit neither category, we don't get simple guessing behavior; rather, subjects show sensitivity to, and make use of, the fine details. Furthermore, after the point of disambiguation, at which point both choices should be equally quickly selected, in fact subjects were slower to fixate on a final choice when the VOT of the initial sound was higher. So, for cases that began more [b]-like, even after the lexical information (-*keet*) clearly indicated the choice, subjects were still slower to settle on the "parakeet" than when the disambiguating information indicated an initial [b] (hence, *barricade*). McMurray et al. suggest that the ambiguous VOT led to activation of both candidate forms, so that

even at the point of disambiguation, there was competition between them which slowed subject response.

Thus, we can hear intermediate voicing values, such as those which might be created by reduction processes like those in Majorcan Spanish. How then do we get from here to a categorical sound change? Hualde et al. argue that this low level variation gets conventionalized as contextually determined variation, and from there can become a categorical change. The sequence they propose is as in (6):

(6) Steps in a Model of sound change (from Hualde 2011)
   a. "Online" effect: gesture reduction and overlap: /apa/ [apa] ~ [aba] (variable degree of voicing) ~ [aφa] (variable closure)
   b. Conventionalization: /apa/ [aba]
   c. Phonemic recategorization: [aba] /-p-/ > [aba] /-b-/

Phonetic variation of a variety of sorts can be produced due to gestural overlap and/or reduction (6a), here illustrated by stop voicing or weakening in degree of constriction. One such effect becomes conventionalized (6b), with a specific context, here voicing of stops in intervocalic position. Finally, words with this conventionalized variant may be restructured with the conventionalized variant recategorized (6c). Here, a nonce form /apa/ being restructured as /aba/.

How a language moves from step (6a) to step (6b) is not very clear. If a change is going to conventionalize, its application "across the board" would make sense because there is a clear context to which the change can be attributed; that is, the context that led to the effect in the first place. On the other hand, what leads low level phonetic variation to become more consistent in this way? The fact that the low level variation has a context gives it some consistency, but something must still occur in the language to allow this type of variation to be tolerated to the degree that it becomes a salient and common enough feature that speakers (and presumably, especially learners) establish it as a regular pattern. Right now, for example, the voiceless stop lenition effect is evident only about 20% of the time in Majorcan Spanish. Presumably there are forces constraining this lenition which would have to shift to allow it to become so common that it is treated as a regular sound pattern of the language. While perceptual learning studies document that speakers can tolerate phonetic ambiguity in speech, it is not immediately evident what would lead the incidence of such ambiguity to increase to the point that a low level lenition effect would become formally instituted in the language. Sociolinguistic considerations may be crucial here, as it is otherwise unclear why the language would spontaneously change its degree of tolerance for phonetic variation surrounding a contrast. Conventionalization is hypothesized to be general, applying to all sounds in the relevant context (as in an across the board Neogrammarian change). This makes sense since conventionalization is generalization across many words with a similar anomaly (in a similar context). For example, many words may have a voiced stop between vowels, even though other words (or other speakers) have a voiceless stop. Since many words are involved, and they all have

this same property, the generalization must be about what they *have in common*, which is the sound and its context.

Recategorization, on the other hand, is hypothesized to be gradual, occurring one lexical item at a time. Hualde et al. have a preliminary proposal for how recategorization takes place; that is, how we get from step (6b) to step (6c). They argue that even among the conventionalized variants (e.g., voiced voiceless stops), there is a range of degrees of the effect, and the items with the most extreme variants are the first to "cross over," recategorizing with a new representation. The more extreme variants arise due to a variety of factors, but in the case of Spanish stop voicing, Hualde et al. suggest that prosody is a crucial factor. They argue that lenition is more extensive, for example, in post-tonic position. Words with voiceless stops in post-tonic position will be the first to get recategorized, because they frequently show strong voicing effects. After occurring in such forms, recategorization may spread through the effect of frequency, analogy, or other factors. Hualde et al. hypothesize that the more extreme variants are recategorized and stored as such, possibly alongside a non-shifted representation (e.g., both [polítigo] and [polítiko] for *político* 'politician'). From that point, if the newer form somehow became dominant, for example through more frequent occurrence, it could become the (primary) form acquired by children. This process is a lexically based one, as the frequency and/or strength of a particular representation of a possible pronunciation for a word is necessarily a property of that word only. That a prosodic effect could help determine the beginning of the recategorization process should not be surprising, but such claims have until now been relatively few, and the idea opens up a vast range of additional sources of explanation of the course taken by sound changes[3]. The additional assumption that multiple competing representations of a word co-exist for speakers during the stage before recategorization is also an interesting one which merits additional extensive investigation.

5. Conclusion

The papers reviewed here showcase new approaches to studying and modeling variation. The speech error studies demonstrate the role multiple levels of linguistic representations play in speech planning and they advance new views of the activation of linguistic content during speech planning. The study of Spanish stop voicing reveals the sorts of conditions in which sound change may take hold, and discusses critical ways in which different aspects of linguistic knowledge can influence the ultimate result. More generally, models of variable activation in both speech production and speech perception promise to provide an improved understanding of speech planning and sound change. Models of speech planning will be further enriched by including consideration of the full range of prosodic influences, including multiple levels of prosodic phrasing and prominence. Likewise, models

of sound change should continue to be refined by new accounts of how speech variation is produced and processed.

Correspondence e-mail address: marie.huffman@stonybrook.edu

Notes

1. It seems likely that the degree to which later material is "pre-prepared" in natural speech may depend in part on the syntactic structure.
2. In both examples, underlined syllables are those which were targeted for word stress within the non-word items; bolded letters indicate the consonants that were identical in the word sets showing the effect under discussion.
3. See also Beckman et al. (1992) on how prosodic effects can influence gestural overlap, another major source of phonetic variability which can lead to sound change.

References

Beckman, Mary E., Kenneth de Jong, Sun-Ah Jun & Su Ar Lee. 1992. The interaction of coarticulation and prosody in sound change. *Language and Speech* 35. 45–58.

Cho, Taehong & Patricia Keating. 2009. Effects of initial position versus prominence in English. *Journal of Phonetics* 37(4). 466–485.

Choe, Wook Kyung & Melissa A. Redford. (this issue) The distribution of speech errors in multi-word prosodic units. *Laboratory Phonology* 3(1).

Dell, Gary, Lisa Burger & William Svec. 1997. Language production and serial order: A functional analysis and a model. *Psychological Review* 104(1). 123–147.

Fromkin, V. A. 1971. The nonanomalous nature of anomalous utterances. *Language* 47. 27–52.

Fougeron, Cécile & Patricia A. Keating. 1997. Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America.* 101(6). 3728–3740.

Hualde, José Ignacio, Miquel Simonet & Marianna Nadeu. 2011. Consonant lenition and phonological recategorization. *Laboratory Phonology* 2(1). 301–329.

Jacewicz, Eva, Robert A. Fox & Joseph Salmons. 2006. Prosodic prominence effects on vowels in chain shifts. *Language Variation and Change* 18(3). 285–316.

Kraljic, Tanya and Arthur G. Samuel. 2007. Perceptual adjustments to multiple speakers. *Journal of Memory and Language* 56(1). 1–15.

McMurray, Bob, Michael K. Tanenhaus & Richard N. Aslin. 2009. Within-category VOT affects recovery from "lexical" garden paths: Evidence against phoneme-level inhibition. *Journal of Memory and Language* 60(1). 65–91.

Norris, Dennis, James M. McQueen & Anne Cutler. 2003. Perceptual learning in speech. *Cognitive Psychology* 47(2). 204–238.

Tilsen, Sam. 2011. Metrical regularity facilitates speech planning and production. *Laboratory Phonology* 2(1). 185–218.