

PHONETIC ENCODING OF PROSODIC STRUCTURE

Patricia A. Keating

Dept. of Linguistics, University of California, Los Angeles

ABSTRACT: Speech production is highly sensitive to the prosodic organization of utterances. This paper reviews some of the ways in which articulation varies according to prosodic prominences and boundaries.

INTRODUCTION

Prosody is the organization of speech into a hierarchy of units or domains, some of which are more prominent than others. That is, prosody serves both a grouping function and a prominence marking function in speech. Prosody includes (1) phrasing, which refers to the various size groupings of smaller domains into larger ones, such as segments into syllables, syllables into words, or words into phrases; (2) stress, which refers to prominence of syllables at the word level; (3) accent, which refers to prominence at the phrase level; and distributions of tones (pitch events) associated with any of these, such as (4) lexical tones associated with syllables; or (5) intonational tones associated with phrases or accents. These various aspects of prosody together define and reflect the prosodic structure of an utterance. For the purposes of this paper, we can consider any one interval of speech that is grouped into a single prosodic domain and ask, At what level of prosodic structure does this domain occur? What speech events occur at the beginning and end of this domain? What is in the domains before and after this one? How prominent is this domain relative to those neighbours? All of this information will be relevant phonetically.

The phonetic dimensions that are most obviously involved in realizing prosody are pitch, duration, and loudness, which are generally thought of as the suprasegmental dimensions. But the phonetic dimensions that are typically thought of as more segmental than suprasegmental also serve to realize prosodic distinctions. For example, it is well-known that vowel quality varies not only with phonemic vowel identity, but also with such suprasegmental factors as stress and length (Lehiste 1970). Put generally, the phonetic realization of an individual speech segment (vowel or consonant)'s phonological properties depends in part on that segment's position in the entire prosodic structure. It's useful in this regard to think of each segment as a set of phonetic features occupying a terminal node at the bottom of a very large tree of the prosodic structure. The exact pronunciation of any one such feature will depend on the other features in that segment, features of neighbouring segments, *and the position of the feature in the overall tree*. Thus segmental phonetic dimensions are as much about prosody as are the traditional suprasegmental dimensions.

In Levelt, Roelofs & Meyer (1999)'s important model of planning for speech production, a distinction is made between (1) *phonological encoding*, or generating a complete phonological representation, including prosody, from lexical entries and syntactic structure; and (2) *phonetic encoding*, which specifies the surface phonetic shape of the phonological representation. Levelt *et al.*, relying on the traditional distinction between segmental and suprasegmental phonological representations and speech parameters, envision segmental and suprasegmental planning for both phonological and phonetic encoding as virtually independent, with perhaps some late, minor mutual influences. In particular, phonetic encoding of segments largely consists in the retrieval of stored precompiled gestural scores for syllables. As discussed at length in Keating & Shattuck-Hufnagel (2002), the missing ingredient from their model is the close link between prosodic structure and segmental phonetic properties. We stress that even if phonetic encoding relies on stored syllable plans, the work of phonetic encoding has just begun with their retrieval, as adjustments to them are required on the basis of all kinds of prosodic information; and even if phonetic encoding relies on stored exemplars, then that retrieval operation itself must be highly sensitive to prosodic structure and the retrieved exemplar may still require further processing. The present paper, like Keating & Shattuck-Hufnagel (2002), defends this claim, by reviewing a variety of ways in which phonetic encoding must be sensitive to prosodic structure, here focusing especially on results from my own laboratory.

My special interest here is what is thought of as a strength relation between positions and phonetic realizations. The general idea is that some prosodic positions are stronger than others, and segments in stronger positions are pronounced with greater strength than are segments in weaker positions. Thus in turn segment strength serves as an indicator of positional strength. Not surprisingly, prominent syllables are prosodically strong. It turns out that other strong positions are the beginnings of the various sized prosodic domains, so segments in domain-initial positions are generally strengthened, while segments within those domains are not. Because of this relation, segment strength also serves as an indicator of local coherence vs. disjuncture in speech: a strengthened segment indicates a break, the start of a new domain, while domain-internal spans of segments are not interrupted by strengthening.

What exactly does strengthening mean? Strengthening is articulatory, meaning that the articulations themselves are stronger, or more extreme. For consonants, for example, the primary oral constriction is more extreme, meaning that the primary articulator moves farther from a neutral position, into a more extreme position which reduces the size of any mouth opening. Such strengthening (often called *fortition* in the historical linguistics literature) is the opposite of *weakening* (or *lenition*), by which a more reduced primary consonant articulation results in a greater mouth opening. In historical sound changes (e.g. Hock 1991), consonant strengthenings and weakenings are observed in different prosodic positions; for example, word-initial consonant strengthening may change an approximant consonant into an obstruent, or a continuant into a non-continuant. Word-initial segments are also preserved more often than other segments, since lenition in weak positions often leads to complete loss. Glottalization of word-initial vowels at the beginnings of prosodic domains (Dilley *et al.* 1996), which gives them a more consonantal quality, can also be seen as a strengthening, though of a very different sort (Pierrehumbert & Talkin 1992).

As will be seen in more detail below, the beginnings of prosodic domains and prominent syllables are both locations for strengthenings. These are complemented by domain-final lengthening, the well-known phenomenon where segments at the ends of domains have longer durations (e.g. Wightman *et al.* 1992). The total effect is that at end of one domain there is a slowing down, then at the beginning of the next domain, a strong attack, with another strong moment associated with any prominence. At the word level, languages may license or distribute inherently stronger segment types in stronger positions, and phonetic strength will also be modulated by position. Durations will similarly be modulated by position. Overall, then, there is a tendency for words (and larger units, too) in a given language to have a phonological and/or phonetic shape conditioned by the prosodic structure of the language.

EDGE EFFECTS

In my work (with various collaborators) I have been most interested in domain-initial articulatory strengthening, that is, strengthening associated with the beginnings of prosodic domains. I have put forward a specific claim about how this strengthening works through the whole of prosodic structure: that it is cumulative, in the sense that the higher in the prosodic tree an initial position is, the stronger that position and the segment in it. The empirical support for this claim is somewhat mixed, as I will make clear below, but there is an interesting range of data that seem to work this way.

At UCLA we have primarily used electropalatography (EPG) to infer the strength of segment articulation. With the Kay Elemetrics EPG system, a speaker wears a custom-made false palate embedded with 96 contact electrodes. When the tongue touches any electrodes, a circuit is completed, current flows, and the contact is thereby registered. A computer samples the contact over the entire palate every 10 msec, and each frame of data shows which electrodes were contacted at that time. Our general method is to construct speech materials that put a test consonant into different prosodic positions, and then take the simplest measure of strength, namely the maximum amount of contact between tongue and palate found during that consonant in each condition. This measure ignores where on the palate, and when during the consonant, this peak contact occurs, but those aspects can be measured as follow-ups. For stop consonants, we also generally measure the duration of the stop seal, that is, the amount of time that the vocal tract is completely sealed off by the stop occlusion.

In our first study, Fougeron & Keating (1997), we looked at English /n/ and /o/ in initial and final positions in several domains, as articulated by three speakers in reiterant speech. Our original

hypothesis motivating this study did not involve domain-initial strengthening; we expected to see some kind of final or progressive weakening. Yet what we found in our data for /n/ was fairly cumulative domain-initial strengthening: each speaker made 3 or 4 pairwise cumulative distinctions between /n/s in domain-initial positions. However, no speaker made all possible distinctions, and there was no pairwise distinction that was made by all three speakers. There was a trend towards strengthening (meaning less contact) of the vowels in initial /no/ syllables; and strengthening of /n/ in domain-final syllables was occasionally found. Domain-final vowel strengthening was observed, but it was not as strongly cumulative, in that phrase-final /o/s at different levels were not distinguished by amount of opening. Also, acoustic duration of /o/ did not vary much according to domain. In contrast, acoustic durations of /n/ (and as measured by Keating *et al.* 2003b, also articulatory durations of /n/) were more consistently cumulative than linguopalatal contact; nonetheless, correlations between /n/ durations and contact were minimal to modest. Finally, many explicit tests for articulatory *declination* – a progressive weakening over an utterance – failed to provide any evidence at all for it.

We then followed up with a study of domain-initial strengthening in three other languages which differ in their prosodic properties: French, Korean, and Taiwanese. In this study we used real-word utterances rather than reiterant speech. Overall, each language shows cumulative initial strengthening. In fact, the surprising thing is how similar the results are for the three languages, despite their prosodic differences. The pattern was most consistent for Korean, which we had predicted could show the most articulatory strengthening, as its domain beginnings are generally thought to be prosodically strong. What we found is that all the languages had fairly consistent cumulative initial lengthening, and in Korean strengthening is most related to that lengthening. In our English data, the correlations of contact with duration were low to modest; in French there was a stronger relation; but in Korean the correlations were very high. Strengthening in Korean seems to be related to how much time is available for the articulation: in Cho & Keating (2001) we showed that up to 80 msec, the amount of contact is a function of the time, with the peak contact coming at the end of the consonant and shorter consonants undershooting their target; but above 80 msec, there is no additional contact. Thus I would say that in Korean, there is little if any independent effect of strengthening apart from lengthening. In that the other languages are not like this, initial strengthening seems to be a separate effect from lengthening, though both effects are sensitive to prosodic position. Furthermore, in light of the Korean pattern, it seems possible that the greater opening of the final vowels in the Fougeron & Keating study of English could be due to final lengthening, rather than an independent strengthening.

Other follow-up studies on these languages from our lab have been Fougeron (2001) on French, and Cho & Jun (2000) and Kim (2001) on Korean. One finding worth noting from the French study is that domain-initial strengthening is limited to only the very first segment in the domain – thus in a /kl/ cluster, only the first C is strengthened, and the vowel /i/, in initial position, shows a limited strengthening. This very local effect is perhaps consistent with the fact that in French, final lengthening is also limited in extent (Fletcher 1991), though in English it can extend back to a stressed vowel, as seen by Wightman *et al.* (1992). A second finding was that the sibilant fricative /s/, unlike other consonants, varied very little across prosodic positions, presumably because the production of sibilance constrains the articulation. This result was not unexpected given Byrd's (1996) comparison of English /s/ in onset vs. coda position.

We have also in a limited way compared domain-initial with domain-final consonants in English, using symmetrical nonsense words like /tɛbɔbɛt/ whose first and last consonant was one of /t d n l/ (Keating *et al.* 1999). These test words occurred at the beginning or the end of an utterance, so that the word-initial consonants occurred utterance-initially vs. utterance-medially, while the word-final consonants occurred utterance-medially vs. utterance-finally. The maximum EPG contact depended on position in both the word and the utterance. Overall, as expected, the consonants had more contact at the beginnings of words than at the ends, and also as expected, word-initial consonants had more contact when they were also at the beginning of an utterance. However, utterance-final consonants had more contact than other word-final consonants. That is, there is no cumulative domain-final weakening of consonants; instead we see some strengthening at the end of the largest domain. The role of domain-final lengthening in this apparent strengthening deserves further study.

Other researchers have contributed to our knowledge about initial strengthening in a variety of languages, including Gordon (1999), Lavoie (2001), Tabain (2003), Tabain *et al.* (2003), Onaka (2003). Most such studies have found an overall tendency, but not a perfect pattern, of cumulative

domain-initial strengthening. An exception is Byrd *et al.* (2000) on Tamil, which shows initial lengthening but no effect of prosodic position on the extent of articulation, evidence that again supports the independence of lengthening and strengthening as prosodically conditioned effects on speech articulation. And, Byrd & Saltzman (1998) found a very different result for English than we did: comparing lip movements at the boundaries of what were probably three different prosodic domains, they found that displacement to the postboundary consonant was highly correlated with duration.

Prosodic boundaries also affect coarticulation, in that interactions generally occur between neighbouring segments that are close, and boundaries serve to separate segments. The fact that within-word coarticulation is greater than cross-word coarticulation has already been exploited in the infant perception literature, where it has been shown by Johnson & Jusczyk (2001) that infants are sensitive to this difference in that they can use lack of coarticulation as a cue to the presence of a word boundary. Cho (2002, in press) used EMA to show that larger, phrasal, boundaries inhibit cross-word coarticulation more than smaller boundaries do. This inhibition was only partly attributable to lengthening around the boundary.

PROMINENCE EFFECTS

As mentioned earlier, it is also known that segments in prominent positions are strengthened. Many previous studies have been concerned with the articulation of prominent vowels, including the direction of displacement for prominent high vowels, and differences between prominence and edge positions (e.g. Edwards *et al.* 1991, Beckman *et al.* 1992, Fletcher & Vatikiotis-Bateson 1994, de Jong 1995, Harrington *et al.* 2000, Cho 2002, Erickson 2002). Other studies (reviewed in Epstein 2002, 2003) have been concerned with the phonation qualities associated with prominence.

Some of our EPG work on English has looked at linguopalatal contact with prominence, both lexical stress and accent. A 3-syllable test word such as /tɛbɔbɛt/ had either initial or final "lexical" stress and was either focally accented or not. Neither stress nor accent affected the amount of contact for the lingual consonants, but they did make the consonants longer; and both stress and accent made the vowels more open.

We have also investigated the optical correlates of prominence. In Keating *et al.* (2003a), three Californian speakers read disyllables differing in lexical stress, and sentences differing in location of focal accent (phrasal stress). Video, audio, and face movements were recorded (the face movements using a 3-D optical tracking system and 20 small retro-reflectors on the speaker's face). Several measurements were made of movements of points on the head, eyebrow, lips, and chin, and these measures were tested to see whether they varied as a function of stress/accent. Although in general most measures did vary, phrasal accent affected more measures, and with larger differences between accented vs. unaccented, compared to lexical stress. For example, the eyebrow moved with phrasal accent but not with lexical stress. A subsequent visual perception experiment using these tokens showed that although both lexical stress and phrasal accent were perceived above chance, phrasal accent was perceived more reliably. Correlation analyses of the production and perception data for phrasal accent revealed that *chin opening movements* (displacements and velocities) best accounted for perception performance. This result supports the tradition that the jaw is the articulator most associated with stress/accent: though the lips and the mouth opening the lips define were also visible to perceivers, perceivers apparently did not rely on that part of the face as much as on the chin, which is the visible part that moves most like the jaw.

CONCLUSION

In sum, when a speaker plans for the phonetic aspects of speech production, prosodic structure organizes the treatment of possibly every feature in every segment, and the interactions of segments. One aspect of this dependence is the relation between the strength of a prosodic position, and the phonetic strength of a segment in that position. A theory of phonetic encoding that incorporates this basic fact is a major challenge, but an important one.

ACKNOWLEDGEMENTS

This research was supported by an NSF Linguistics grant to P. Keating (#95-11118) and an NSF KDI grant to L. Bernstein *et al.* (#99-96088). All participation of human subjects was approved by UCLA and, where relevant, the House Ear Institute. Collaborations with all my co-authors on the various papers cited here, and the contributions of the former students cited, are gratefully acknowledged.

REFERENCES

- Beckman, M.E., Edwards, J. & Fletcher, J. (1992) "Prosodic structure and tempo in a sonority model of articulatory dynamics" in Docherty, G.J. & Ladd, D.R. (eds.) *Papers in laboratory phonology II: gesture, segment, prosody*, Cambridge University Press 68–86.
- Byrd, D. (1996) "Influences on articulatory timing in consonant sequences" *Journal of Phonetics* **24** 209-244.
- Byrd, D. & Saltzman, E. (1998) "Intragesural dynamics of multiple prosodic boundaries" *Journal of Phonetics* **26** 173-200.
- Byrd, D., Kaun, A., Narayanan, S. & Saltzman, E. (2000) Phrasal signatures in articulation. In M. B. Broe and J. B. Pierrehumbert (eds.) *Papers in Laboratory Phonology V: Acquisition and the Lexicon*, Cambridge University Press 70-87.
- Cho, T. (2002) *The Effects of Prosody on Articulation in English*, New York: Routledge.
- Cho, T. (in press) "Prosodically-conditioned strengthening and vowel-to-vowel coarticulation in English" *Journal of Phonetics*.
- Cho, T. & Jun, S.A. (2000) "Domain-initial strengthening as enhancement of laryngeal features: Aerodynamic evidence from Korean" in J. Boyle, J-H. Lee, and A. Okrent (eds.), *Chicago Linguistic Society* **36** 31-44 [also in *UCLA Working Papers in Phonetics* **99** 57-69].
- Cho, T. & Keating, P. (2001) "Articulatory and acoustic studies on domain-initial strengthening in Korean" *Journal of Phonetics* **29** 155-190.
- de Jong, K. (1995) "The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation" *Journal of the Acoustical Society of America* **97** 491-504.
- Dilley, L., Shattuck-Hufnagel, S. & Ostendorf, M. (1996) "Glottalization of word-initial vowels as a function of prosodic structure" *Journal of Phonetics* **24** 423-444.
- Edwards, J., Beckman, M.E. & Fletcher, J. (1991) "The articulatory kinematics of final lengthening" *Journal of the Acoustical Society of America* **89** 369-82.
- Epstein, M. (2002) *Voice Quality and Prosody in English*. Unpublished UCLA dissertation.
- Epstein, M. (2003) "Voice quality and prosody in English" *Proceedings of the 15th International Congress of Phonetic Sciences* 2405-2408.
- Erickson, D. (2002) "Articulation of Extreme Formant Patterns for Emphasized Vowels" *Phonetica* **59** 134-149.
- Fletcher, J. (1991) "Rhythm and final lengthening in French" *Journal of Phonetics* **19** 193-212.
- Fletcher, J. & Vatikiotis-Bateson, E. (1994) "Prosody and intrasyllabic timing in French" *Ohio State University Working Papers* **43** 41-46.
- Fougeron, C. (1999) "Prosodically conditioned articulatory variations: a review" *UCLA Working Papers in Phonetics* **97** 1-74.

Proceedings of the 6th International Seminar on Speech Production, Sydney, December 7 to 10, 2003.

- Fougeron, C. (2001) "Articulatory properties of initial segments in several prosodic constituents in French" *Journal of Phonetics* **26** 45-69.
- Fougeron, C. & Keating, P. (1997) "Articulatory strengthening at edges of prosodic domains" *Journal of the Acoustical Society of America* **101** 3728-3740.
- Gordon, M. (1999) "The effect of stress and prosodic phrasing on duration, acoustic amplitude and air flow of nasals in Estonian" *UCLA Working Papers in Phonetics* **92** 151-159.
- Harrington, J., Fletcher, J. & Beckman, M. (2000) "Manner and place conflicts in the articulation of Australian English consonants" in M. Broe & J. Pierrehumbert (eds.) *Papers in Laboratory Phonology V: Language Acquisition and the Lexicon*, Cambridge University Press 40-51.
- Hock, H. H. (1991) *Principles of Historical Linguistics*, 2nd edition, Berlin: Mouton de Gruyter
- Johnson, E. & Jusczyk, P. (2001) "Word segmentation by 8-month-olds: When speech cues count more than statistics" *Journal of Memory and Language* **44** 1-20.
- Keating, P., Baroni, M., Mattys, S., Scarborough, R., Alwan, A., Auer, E. & Bernstein, L. (2003a) "Optical Phonetics and Visual Perception of Lexical and Phrasal Stress in English" *Proceedings of the 15th International Congress of Phonetic Sciences* 2071-2074.
- Keating, P., Cho, T., Fougeron, C. & Hsu, C-S. (2003b) "Domain-initial strengthening in four languages" in J. Local, R. Ogden & R. Temple (eds.) *Papers in Laboratory Phonology VI*, Cambridge University Press 143-161.
- Keating, P. & Shattuck-Hufnagel, S. (2002) "A Prosodic View of Word Form Encoding for Speech Production" *UCLA Working Papers in Phonetics* **101** 112-156.
- Keating, P., Wright, R., & Zhang, J. (1999) "Word-level asymmetries in consonant articulation" *UCLA Working Papers in Phonetics* **97** 157-173.
- Kim, S. (2001) "Domain initial strengthening of Korean fricatives /s/ and /s*/" *Harvard Studies in Korean Linguistics* **IX** 164-173.
- Lavoie, L. (2001) *Consonant strength: Phonological patterns and phonetic manifestations*, New York: Garland.
- Lehiste (1970) *Suprasegmentals*, Cambridge MA: MIT Press.
- Levelt, W. J. M., Roelofs, A. & Meyer, A. S. (1999) "A theory of lexical access in speech production" *Brain and Behavioral Sciences* **22** (1) 1-38.
- Onaka, A. (2003) "Domain-initial strengthening in Japanese: An acoustic and articulatory study", *Proceedings of the 15th International Congress of Phonetic Sciences* 2091-2094.
- Pierrehumbert, J. & Talkin, D. (1992) "Lenition of /h/ and glottal stop" in G. Docherty, & D. R. Ladd (eds.) *Papers in laboratory phonology I: gesture, segment, prosody*, Cambridge University Press, 90-117.
- Tabain, M. (2003). "Effects of prosodic boundary on /aC/ sequences: articulatory results" *Journal of the Acoustical Society of America* **113** 2834-2849.
- Tabain, M., Perrier, P. & Savariaux, C. (2003), "A kinematic study of prosodic boundary effects on /i/ articulation in French", *Proceedings of the 15th International Congress of Phonetic Sciences* 2617-2620.
- Wightman, C., Shattuck-Hufnagel, S., Ostendorf, M. & Price, P. (1992) "Segmental durations in the vicinity of prosodic phrase boundaries" *Journal of the Acoustical Society of America* **91** 1707-17.