운율구조에 의한 음성 발화와 인식: 개관

조태홍
한양대학교 영어영문학부

Prosodically-driven speech production and perception: A review

Taehong Cho
Division of English Language and Literature
Hanyang University, Seoul, Korea

tcho@hanyang.ac.kr

Abstract

This paper summarizes some recent findings with respect to how prosodic structure is manifested in fine-grained phonetic details and how such phonetic manifestation of prosodic structure may be exploited in spoken word recognition.

1.Introduction

In speech production, speakers produce a string of words to form an utterance in a systematically organized way, such that some words may be grouped together within a phrase while some others may be produced separately and distinctly, each forming a single phrase; and yet some words may be produced with more prominence relative to others. Such a grouping of words and distribution of relative prominence among words manifest a prosodic structure of the utterance. A single sentence (with the same lexical content and syntactic structure), however, may be produced with different prosodic structures, every time it is pronounced even by the same speaker. Some determining factors among many others may include what kind of syntactic structure the sentence is built on; what kind of informational structure the utterance conveys in a particular discourse situation; how many syllables or words are available to form one chunk; and how fast the utterance is produced ([18,19,23]). As such, prosodic structure has been widely recognized as an essential element of speech production as it conveys a great deal of both linguistic and extralinguistic structural information. This paper summarizes some recent studies with respect to how prosodic structure is manifested in fine-grained phonetic details and how the fine-grained yet systematic phonetic variation stemming from prosodic structure may be used in spoken word recognition.

2.Phonetic correlates of prosodic structure

In pace with the growing awareness of the role of prosodic structure in speech production, a large body of phonetic studies of the past two decades or so have increasingly demonstrated the importance of fine-grained phonetic details in building up differential prosodic structures of utterances.

2.1. Domain-final phenomena

One of the most conspicuous phonetic hallmarks of prosodic structure has perhaps been found in the temporal dimension of phonetic realizations. A great deal of studies has reported that a domain-final segmental element is realized with systematic yet fine-grained lengthening either acoustically or articulatorily, such that the degree of final lengthening is closely correlated with the level of prosodic structure or the prosodic boundary strength (e.g., [4,7,12,27]).

In addition to the temporal expansion (which is usually accompanied by intonational boundary markings, [2,25]), more recent studies have demonstrated that the final element may undergo spatial expansion, as well. For instance, it has been shown that the amount of linguopalatal contact (as measured by Electropalatography, EPG) for the preboundary vowel decreases as the boundary level

increases ([15]). A decreased EPG contact indicates more vocalic opening at the end of larger prosodic domains. Subsequently, it has been reported that the amount of lip opening is larger domain-finally than domain-medially for both /a/ and /i/ in English ([4,7]); the tongue position is consistently higher for /i/ but consistently lower for /a/ in domain-final position, as compared to their domain-medial counterparts ([6]); and C-to-V displacement in the tongue movement for French /a/ in the ta#C context is larger before a larger prosodic boundary ([26]).

Cho (2004) has also shown that domain-final vowels such as English /a, i/ resist coarticulation with the following vowel more at a higher level of prosodic juncture. It is proposed that prosodically-conditioned V-to-V coarticulatory reduction is another type of strengthening that occurs in prosodically strong locations. The prosodically-driven coarticulatory patterning is taken to be part of the phonetic signatures of the hierarchically-nested structure of prosody ([5]).

2.2. Domain-initial phenomena

These findings together demonstrate various domain-final phonetic phenomena which may come in package as fine-grained, yet systematic phonetic hallmarks of prosodic structure. Yet another line of research has focused on fine-grained phonetic markings of prosodic structure coming from the other side of prosodic juncture, namely domain-initial position. In an EPG study, Fougeron & Keating ([15]) has shown that degree of linguopalatal contact for domain-initial /n/ in English increases progressively as the prosodic level moves up in the prosodic hierarchy (from Word-initial, Intermediate-Phrase-initial to Intonational-Phrase-initial, if one follows a model of prosodic organization [2]; see also [1] for a prosodic transcription system in American English, ToBI [Tone and Break Index]). This phenomenon has been referred to as domain-initial articulatory strengthening, and similar effects, though with some language-by-language variation, have been found in EPG studies on other languages such as Korean ([9,21]), French ([14]), Taiwanese ([17]) and Japanese ([24]), and in magnetometer studies on labial articulation in English ([3,7]). See also [20] for cross-linguistic comparisons of EPG studies. (Note that the term domain-initial strengthening is generally used to refer to any phonetic patterning arising in domain-initial position, including spatial and temporal expansion of articulation due to prosodic boundaries, whereas the term domain-initial articulatory strengthening is referred specifically to spatial expansion.)

2.3. Prosodic strengthening and featural enhancement

Questions have been posed as to how prosodic strengthening phenomena are linked to enhancement of features in the phonological system of a given language. As one of the first attempts to explore this issue, Cho & Jun (2000) investigated how consonantal features are realized in prosodically strong locations by examining the three-way contrastive stops in Korean (i.e., fortis, lenis, aspirated) [8]. The results of an acoustic/aerodynamic study suggest that fortis and aspirated stops are strengthened in a way that enhances underlying features of the stops (e.g., [constricted glottis] and [spread glottis], respectively) i.e., paradigmatic/phonemic enhancement. On the other hand, the lenis stop, arguably unspecified for either of these features, is nonetheless strengthened, but this time presumably to enhance the consonantality that results in a greater CV contrast, i.e., syntagmatic enhancement.

Another example is found in Cho (2005) who investigated the effects of accent and prosodic boundaries on the production of English vowels (/a, i/), by concurrently examining acoustic vowel formants and articulatory maxima of the tongue, jaw, and lips obtained with EMA (Electromagnetic Articulography) ([6]). The results demonstrate that prosodic strengthening (due to accent and/or prosodic boundaries) has differential effects depending on the source of prominence (in accented syllables vs. at edges of prosodic domains; domain-initially vs. domain-finally). The results are interpreted in terms of how the prosodic strengthening is related to phonetic realization of vowel features. For example, when accented, /i/ was fronter in both acoustic and articulatory vowel spaces (enhancing [-back]), accompanied by increase in both lip and jaw

openings (enhancing sonority). By contrast, at edges of prosodic domains (especially domain-finally), /i/ was not necessarily fronter, but higher (enhancing [+high]), accompanied by increase only in the lip (not jaw) opening. This suggests that the two aspects of prosodic structure (accent vs. boundary) are differentiated by distinct phonetic patterns. Further, it implies that prosodic strengthening, though manifested in fine-grained phonetic details, is not simply a low-level phonetic event but a complex linguistic phenomenon, closely linked to enhancement of phonological features and positional strength which may license phonological contrasts.

Cho & McQueen (2005) further asked a question as to whether domain-initial strengthening is constrained by language-specific phonological system of a given language--i.e, whether domain-initial strengthening phenomena varies cross-linguistically depending on the phonological system of languages [10]. They thus examined acoustic effects of domain-initial strengthening in Dutch as well as its language-specificity in terms of realization of phonetic features in comparison with that in English. In this study, prosodic influences on phonetic realizations of four Dutch consonants (/t d s z/) were examined. Sentences were constructed containing these consonants in word-initial position; the factors such as lexical stress, phrasal accent and prosodic boundary were manipulated between sentences. Eleven Dutch speakers read these sentences aloud. The patterns found in acoustic measurements of these utterances (e.g., VOT, consonant duration, voicing during closure, spectral center of gravity, burst energy) indicate that the low-level phonetic implementation of all four consonants is modulated by prosodic structure. One of the important findings was that shorter VOTs were found for /t/ in prosodically stronger locations (stressed, accented, and domain-initial), as opposed to longer VOTs in these positions in English. This suggests that prosodically-driven phonetic realization is bounded by language-specific constraints on how phonetic features are specified with phonetic content: Shortened VOT in Dutch reflects enhancement of {-spread glottis}, while lengthened VOT in English reflects enhancement of {+spread glottis}.

The results reported in Cho & McQueen suggest that there are cross-linguistic differences in the prosodic modulation of segment realization. Based on the language-specific pattern we observed in the phonetic realization of the voicing contrast (e.g., shortened VOT in Dutch voiceless stops vs. lengthened VOT in English voiceless stops in stronger prosodic positions), it was proposed that the phonetics-prosody interface is modulated by the language-specific phonetic component of the grammar in which phonetic features (e.g., {+/-spread glottis]) are specified with phonetic content. Prosodic structure in a given language thus appears to influence the realization of the phonetic information that is relevant for lexical distinctions in that language.

3.The role of prosodic strengthening in spoken word recognition

Thus far, we have seen how an abstract prosodic structure is phonetically manifested from the perspective of speech production. A question naturally follows as to whether and how the systematic phonetic variation stemming from prosodic structure is used in speech comprehension. This section discusses this issue in connection with the role of domain-initial strengthening in spoken word recognition.

The idea that domain-initial position has a special status in speech recognition is not new: It has often been considered to be an informationally rich locus in speech processing (see [16]). For instance, in the Cohort model (e.g., [22]), word onsets play a critical role in determining which words are considered during the recognition process: Words which begin in the same way as the input word, and only those words, are considered as alternative lexical hypotheses. The precise function of the phonetic consequences of domain-initial strengthening on word onsets in speech perception, however, has not previously been examined. Researchers who have engaged in phonetic studies of domain-initial strengthening have merely speculated on its communicative functions. The simple assumption has been that the speaker signals prosodic structure via articulatory domain-initial strengthening, and the listener makes use of the acoustic consequences of

this articulatory signature in comprehension (see, e.g., [4,13]). More specifically, Fougeron and Keating (1997) discussed the possible benefits that listeners might receive from domain-initial strengthening, including assistance with lexical segmentation and lexical access ([15]). They speculated that, since domain-initial strengthening entails increased articulatory contrast between segments straddling a prosodic boundary, this contrast could contribute to marking that boundary, and thus help listeners to parse the incoming speech signal into words and phrases. Likewise, Cho and Jun (2000) interpreted the pattern of consonantal domain-initial strengthening that they observed as being related to the enhancement of phonological features and phonological contrasts, and hypothesized that these kinds of enhancements could ultimately facilitate word recognition through augmenting lexical distinctions ([8]).

Cho, McQueen and Cox ([11]) moved beyond these speculations, and investigated directly the role of domain-initial strengthening in speech comprehension. They explored the role of the acoustic consequences of domain-initial strengthening in spoken-word recognition. In two cross-modal identity-priming experiments, listeners heard sentences and made lexical decisions to visual targets, presented at the onset of the second word in two-word sequences containing lexical ambiguities (e.g., bus tickets, with the competitor bust). These sequences contained Intonational Phrase (IP) or Prosodic Word (Wd) boundaries, and the second word's onset (e.g., [tI]) was spliced from another token of the sequence in IP- or Wd-initial position. Acoustic analyses showed that the IP-initial consonants were articulated more strongly than the Wd-initial consonants. In Experiment 1, related targets were post-boundary words (e.g., tickets). No strengthening effect was observed (i.e., identity priming effects did not vary across splicing conditions). In Experiment 2, related targets were pre-boundary words (e.g., bus). There was a strengthening effect (stronger priming when the post-boundary onsets were spliced from IP-initial than from Wd-initial position), but only in Wd-boundary contexts. These were exactly the conditions where the phonetic detail associated with domain-initial strengthening could assist listeners in lexical disambiguation. A general conclusion of Cho et al [11], is that domain-initial strengthening is one of many acoustic cues used in the segmentation of continuous speech.

4. Conclusion

This paper has reviewed some recent studies on the relationship between prosodic structure and phonetic realization and between phonetic manifestation of prosodic structure and speech comprehension. The findings support a more general hypothesis that speakers signal prosodic structure via systematic yet fine-grained phonetic details, and listeners use these cues to prosodic structure in decoding continuous speech. Further researches on these issues will certainly shed more light on the relationship between speech production and speech perception that are intertwined, being further modulated by prosodic structure. An important fact to be kept in mind, however, is that not all acoustic phenomena found in speech production are perceptually relevant, which leads to a question as to the extent to which prosodically-conditioned phonetic details in speech production are indeed exploited by listeners.

References

[1] Beckman, M. E. & Elam, G. A. (1997). Guidelines for ToBI labelling, version 3.0. Unpublished Manuscript. Ohio State University.

[2] Beckman, M. E. & Pierrehumbert, J. (1986). Intonational structure in Japanese and English. Phonology Yearbook, 3, 255-309.

[3] Byrd, D. & Saltzman, E. (1998) Intragestural dynamics of multiple prosodic boundaries, Journal of Phonetics, 26, 173-200.

[4] Cho, T. (2002). The Effects of Prosody on Articulation in English. New York: Routledge.

[5] Cho, T. (2004). Prosodically-conditioned strengthening and vowel-to-vowel coarticulation in English. Journal of Phonetics, 32, 141-176.

[6] Cho, T. (2005) Prosodic strengthening and featural enhancement: Evidence from acoustic and articulatory realizations of /a,i/ in English. Journal of the Acoustical Society of America

117(6), 3867-3878.

[7] Cho, T. (in press). Manifestation of prosodic structure in articulation: Evidence from lip kinematics in English. In Laboratory phonology 8: Varieties of phonological competence.

[8] Cho, T. & Jun, S. (2000). Domain-initial strengthening as featural enhancement: Aerodynamic evidence from Korean. In Chicago Linguistics Society, 36, 31-44.

[9] Cho, T. & Keating, P. (2001). Articulatory and acoustic studies of domain-initial strengthening in Korean. Journal of Phonetics, 29,155-190.

[10]Cho, T. & McQueen, J. (2005).Prosodic influences on consonant production in Dutch: Effects of prosodic boundaries, phrasal accent and lexical stress. Journal of Phonetics 33(2) 121-157

[11]Cho, T., McQueen, J. & Cox, E. (in press).Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English. Journal of Phonetics.

[12]Edwards, J. E., Beckman, M. E., & Fletcher, J. (1991). The articulatory kinematics of final lengthening. Journal of the Acoustical Society of America, 89, 369-382.

[13]Fougeron, C. (1999). Prosodically conditioned articulatory variations: A review. UCLA Working Papers in Phonetics, 97, 1-74.

[14]Fougeron, C. (2001). Articulatory properties of initial segments in several prosodic constituents in French. Journal of Phonetics, 29,109-135.

[15]Fougeron, C. & Keating, P. A. (1997). Articulatory strengthening at edges of prosodic domains. Journal of the Acoustical Society of America, 106, 3728-3740.

[16]Gow, D.W., Melvold, J., & Manuel, S. (1996). How word onsets drive lexical access and segmentation: Evidence from acoustics, phonology and processing. Proceedings of ICSLP (CD-ROM). Philadelphia, USA.

[17]Hayashi, W., Hsu, C., & Keating, P. (1999). Domain-initial strengthening in Taiwanese: A follow-up study. UCLA Working Papers in Phonetics,97, 152-156.

[18]Jun, S.-A. (1993). The Phonetics and Phonology of Korean Prosody. Ph.D. Dissertation. Ohio State University.

[19]Keating, P. A. & Shattuck-Hufnagel, S. (2003). A prosodic view of word form encoding for speech production. UCLA Working Papers in Phonetics, 101, 112-156.

[20]Keating, P. A., Cho, T., Fougeron, C. & Hsu, C. (2003). Domain-initial strengthening in four languages. In Papers in laboratory phonology 6: Phonetic interpretations (J. Local, R. Ogden, & R. Temple, editors), pp. 145-163. Cambridge, UK: Cambridge University Press.

[21]Kim, S. (2003). Domain-initial strengthening of Korean fricatives. Harvard Studies in Korean Linguistics, 9.

[22]Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. Cognition, 25, 71102.

[23]Nespor, M. & Vogel, I. (1986). Prosodic Phonology. Dordrecht: Foris.

[24]Onaka, A. (2003). Domain-initial strengtehning in Japanese: An acoustic and articulatory study. In Proceedings of the 15th International Congress of Phonetic Sciences, pp. 2091-2094. Barcelona, Spain.

[25]Pierrehumbert, J. (1980). The Phonology and Phonetics of English Intonation. Ph.D. dissertation, MIT.

[26]Tabain, M. (2003). Effects of prosodic boundary on /aC/ sequences: Acoustic results. Journal of the Acoustical Society of America, 113, 516-531.

[27]Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M. & Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. Journal of the Acoustical Society of America, 91, 1707-1717.