Prosodic Strengthening in Speech Production and Perception: The Current Issues*

Taehong Cho**

ABSTRACT

This paper discusses some current issues regarding how prosodic structure is manifested in fine-grained phonetic details, how prosodically-conditioned articulatory variation is explained in terms of speech dynamics, and how such phonetic manifestation of prosodic structure may be exploited in spoken word recognition. Prosodic structure is phonetically manifested in prosodically important landmark locations such as prosodic domain-final position, domain-initial position and stressed/accented syllables. It will be discussed how each of the prosodic landmarks engenders particular phonetic patterns, how articulatory variation in such locations are dynamically accounted for, and how prosodically-driven fine-grained phonetic detail is exploited by listeners in speech comprehension.

Keywords: prosody, articulation, speech production, speech perception, task dynamics, gestural model, fine-grained phonetic detail

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 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 (see Shattuck-Hufnagel & Turk, 1996 for a review). 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

---

* This work was supported by the Korea Research Foundation Grant funded by the Korean Government (MOEHRD) (KRF-2005-041-A00245).

** Department of English Language and Literature, Hanyang University

(e.g., Nespor & Vogel, 1986; Pierrehumbert & Hirshberg, 1990; Jun, 1993; Keating & Shattuck-Hufnagel, 2003). 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 reviews 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. Some space will also be devoted to introducing a mass-spring gestural model (task dynamics model) to discuss how prosodically-driven articulatory variation can be accounted for from a dynamical perspective of speech production.

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. The phonetic correlates of prosodic structure have been found primarily in three prosodically important landmark locations such as domain-initial position, domain-final position and nuclear-pitch accented position. In what follows, some phonetic phenomena that arise with these locations will be discussed.

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 have 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 at prosodic junctures (e.g., for preboundary (domain-final) lengthening: Edwards, Beckman & Fletcher, 1991; Wightman, et al., 1992; Gussenhoven & Rietvelt, 1992; Berkovits, 1993; Byrd, 2000; Cambier-Langeveld, 2000; Byrd, Krivokapić & Lee, 2006; Cho, 2006 for postboundary (domain-initial) lengthening: Fougeron & Keating, 1997; Byrd & Saltzman, 1998; Byrd, Kaun, Narayanan & Saltzman, 2000; Cho & Keating, 2001; Keating, Cho, Fougeron & Hsu, 2003; Cho & McQueen, 2005; Byrd, et al., 2006; Cole, Kim, Choi & Hasegawa-Johnson, 2007; *inter alia*).

In addition to the temporal expansion (which is usually accompanied by intonational boundary markings, (Beckman & Pierrehumbert, 1986; Pierrehumbert, 1980)), 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 (Fourgeron & Keating, 1997). 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 (Cho, 2006); the tongue position is consistently higher for /i/ but consistently lower for /a/ in domain-final position, as compared to their domain-medial counterparts (Cho, 2005); and C-to-V displacement in the tongue movement for French /a/ in the *ta#C* context is larger before a larger prosodic boundary (Tabain, 2003b).

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.

2.2 Transboundary phenomena

Another type of boundary-related temporal expansion has recently been pointed out that encompasses both the pre- and post-boundary components. Byrd (2000), for example, examined temporal characteristics of 'transboundary' articulatory movement that spans a prosodic boundary —i.e., temporally expanded V-to-C lip closing movement which passes over a prosodic boundary from the pre-boundary vowel to the postboundary consonant (as in *Momma#Mimi*, where # is some prosodic boundary). Another example of such a transboundary effect comes from transboundary vowel-to-vowel (henceforth V-to-V) vocalic movements of the tongue in C1V1#C2V2 contexts: V1-to-V2 tongue movements crossing over the intervening consonant can be assumed to be realized under a direct influence of prosodic boundary, hence temporal expansion. From a strict linear view point of segmental phonology, V2 in #C2V2 is not immediately after the prosodic boundary. From an articulatory perspective, as proposed in Articulatory Phonology (Browman & Goldstein, 1990, 1992), however, vocalic and consonantal gestures are on separate gestural tiers with consonantal gestures being superimposed on vocal gestures (cf. Öhman, 1966; Fowler, 1980), such that the articulatory gesture for V2 starts from the offset of the preboundary V1 and reaches its postboundary vocalic target beyond the intervening consonant across a prosodic boundary. The vocalic gesture is indeed largely independent of the consonantal gesture, especially when the consonantal gestures involve lips, as their articulation minimally interferes with the vocalic lingual articulation. The results of recent studies suggested that the boundary-adjacent lengthening may be most effectively manifested in the transboundary articulatory movements that are under a direct influence of the intervening prosodic boundary (see also Byrd, et al., 2006 and Cho, 2006, for similar results in English, and Tabain, 2003a, b and Tabain & Perrier, 2005, for French data).

2.3 Domain-initial phenomena

Aforementioned 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 (1997) 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 (Beckman & Pierrehumbert 1986; see also Beckman & Elam, 1997, for a prosodic transcription system in American English, ToBI [Tones and Break Indices]). 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 (Cho & Keating, 2001), French (Fougeron, 2001), Taiwanese (Hayashi, Hsu & Keating, 1999) and Japanese (Onaka, 2003), and in magnetometer studies on labial articulation in English (Byrd & Saltzman, 1998; Cho, 2006). See also Keating, et al. (1997) 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.)

2.4 Accent-induced prominence

As mentioned above, stress/accent-induced prominence is another essential element of prosodic structure, and therefore accent-induced phonetic correlates are considered to be another phonetic hallmark of prosodic structure (de Jong 1991; Beckman, et al., 1992; Fowler, 1995; Erickson, 2002; Mooshammer & Fuchs, 2002; Cho, 2006; *inter alia)*. Beckman, Edwards & Fletcher (1992), for example, reported that the jaw opening gesture under accent is associated with an increase in duration and displacement, though without a substantial increase in peak velocity. In a kinematic study of the jaw and the tongue movement in English, de Jong (1995a), for example, showed that certain accented segments are produced with more extreme articulatory movements in a direction that results in an enhancement of distinctive features of the segments, based on which a theory of *localized hyperarticulation* has been proposed. It predicts that all phonemically distinctive contrasts will be directly affected by stress. (De Jong expanded Lindblom's (1990) notion of hyperarticulation, applying not only to extended discourses, but also (locally) to individual syllables, and thus to characterize the extreme articulation localized to stressed/accented syllables.)

In investigating articulatory kinematics of jaw movement, Fowler (1995) found that there was a strong tendency towards larger, longer, and faster movement under accent. Cho (2006) also reported that the C-to-V lip opening gesture under nuclear pitch accent is associated with an increase in almost all measured kinematic variables including displacement, movement duration and peak velocity. Based on a previous assumption that stress consists of a global increase in production effort (e.g., Öhman, 1967; Lehiste, 1970), Fowler suggested that "[p]erhaps, it is not that kinematic adjustments occur to accommodate a pitch accent, but rather that the perceptually salient pitch accent highlights *global effort* increases for the listeners"(emphasis added, p. 369). In other words, the global effort hypothesis postulates that it is not that a particular dynamical setting induces accent-induced kinematic variations but that accent-induced kinematic variations is driven by the prominence maximization principle, which is somewhat comparable to de Jong's hyperarticulation hypothesis.

3. Prosodic strengthening and featural enhancement

Both boundary- and accent-induced phonetic hallmarks have been collectively referred to as 'prosodic strengthening' which is defined as spatial and temporal expansion that arises with prosodic landmark locations (e.g., Cho, 2005; Cho & McQueen, 2005). 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 just discussed in the previous section, de Jong's localized hyperarticulation predicts enhancement of phonological features in one of the prosodic landmark locations—i.e., in an accented syllable. In supporting the hyperarticulation hypothesis, de Jong (1995a) showed that the tongue position for English vowel /U/ is backer, but not necessarily higher under stress, being accompanied by lower jaw position. (The extended hyperarticulation hypothesis has also been endorsed by Beckman and her colleagues (e.g., de Jong, Beckman & Edwards, 1993)). This model also appears to explain at least in part Macchi's (1985) observation that non-low vowels /u,i,E/ under stress tend to have lower jaw positions but not necessarily lower tongue body position. Thus, the data in de Jong (1995a) and Macchi (1985) suggest that the jaw always lowers (leading to the sonority expansion, making the sound louder) while the tongue position is in direction of enhancing vowel's distinctive features.

More recently, a comparable maximization of phonological contrast has also been bound in boundary-related positions. As one of the first attempts to explore this issue, Cho & Jun (2000) investigated how consonantal features are realized in domain-initial position by examining the three-way contrastive stops in Korean (i.e., fortis, lenis, aspirated). 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 (structural) enhancement.

Another example of prosodically-driven featural enhancement 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). 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 vary cross-linguistically depending on the phonological system of languages. 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.

4. Speech dynamics and prosodic strengthening

Thus far, I have outlined previous studies that illuminate how prosodically-induced phonetic variation can be accounted for in terms of prosodic strengthening. This section discusses how prosodically-driven articulatory variation can be accounted for in a speech dynamic model, called a mass-spring task dynamic model (Saltzman & Munhall, 1989; see Hawkins (1992) for an overview for non-specialists).

Quite a few speech researchers (e.g., Edwards, et al., 1991; de Jong, 1991; Beckman, et al., 1992; Harrington, Fletcher & Roberts, 1995; Saltzman, 1995; Byrd & Saltzman, 1998, 2003; Byrd, Kaun, Narayanan & Saltzman, 2000; Byrd, 2000; Byrd, et al., 2006) have examined kinematics of articulatory movements and suggested that prosodically conditioned articulatory variation may be controlled by a particular dynamical parameter setting in the framework of a mass-spring task dynamical model. In the next section, I will first give a brief review of a mass-spring task dynamical model, and then more specific questions that have been addressed in the literature.

4.1 Task dynamic model and dynamical parameters

As the term task dynamic implies, it describes articulatory movement in terms of the task to be executed, using dynamics that are specific not to the articulators that are executing the task, but to the task itself. The term task can be thought of as the goal of articulatory movement, such as forming an appropriate oral constriction at a location in the vocal tract in order to produce an oral fricative (e.g., /ʃ/). Such a movement pattern or an articulatory action that forms a linguistically significant constriction is called a 'gesture.' The dimensions along which the articulatory action or goal for constrictions is executed are called 'tract variables.' Some of these tract variables include lip aperture (LA), tongue-body constriction location (TBCL), and tongue-body constriction degree (TBCD). In Articulatory Phonology (Browman & Goldstein, 1990, 1992), the articulatory gesture is viewed as the phonological unit, and the information about the gestural structures (including the information about the task to be done) is fed into the task dynamic model.

In the task dynamic model, the articulatory gesture is described in terms of the behavior of the abstract 'mass' which is connected to a 'spring' and a 'damper' in a critically damped mass-spring system. As Hawkins (1992) describes, it is as if one end of the spring is attached to the mass, and the other end is held at the target location. Then, as the target location (the other end of the spring) changes, the spring is stretched, and the mass is pulled towards the target location. The mass-spring gestural model is critically damped, so that the mass never reaches the target location, and is not pulled back to its original location, but rather it stays in the target region, continuously and slowly reaching the equilibrium position of the spring. In short, in a critically damped mass-spring gestural model, the mass does not oscillate, but asymptotes towards the equilibrium position, such that the gesture is generally realized as a one-directional movement towards the target.

In the model, the gesture is defined as a dynamical system specified with a set of parameter values. Relevant dynamical parameters include target (underlying amplitude), stiffness (or natural frequency), damping ratio, intergestural timing, and activation time. Characteristics of the articulatory movements that result from executing gestures depend on values of these parameters specified for a given gesture. The relationship among dynamical parameters in a critically damped mass-spring system is described in the following differential equation (Browman & Goldstein, 1990; Hawkins, 1992):

$$m\ddot{x} + b\dot{x} + k(x - x_0) = 0$$

where m = mass associated with the task variable
b = damping of the system
k = stiffness of the spring
x_0 = equilibrium position of the spring (the target)
x = instantaneous value of the task variable (current location of the mass)
$\dot{x}$ = instantaneous velocity of the task variable
$\ddot{x}$ = instantaneous acceleration of the task variable
$(x - x_0)$ = instantaneous displacement of the task variable

(from Hawkins, 1992:17)

In this equation, the mass (m) is always set to be constant, and the damping ratio is also usually constant. Therefore, once we know parameter values for the stiffness (k) and the target (x0), we can solve the equation for the tract variable x (e.g., lip aperture) associated with a given gesture.

Crucially, the model assumes that any systematic articulatory or kinematic variation is interpreted as consequences of dynamical parameter settings. Thus, in theory, any systematic kinematic variations arising from prosodic conditions should be accounted for by a particular dynamical mechanism.

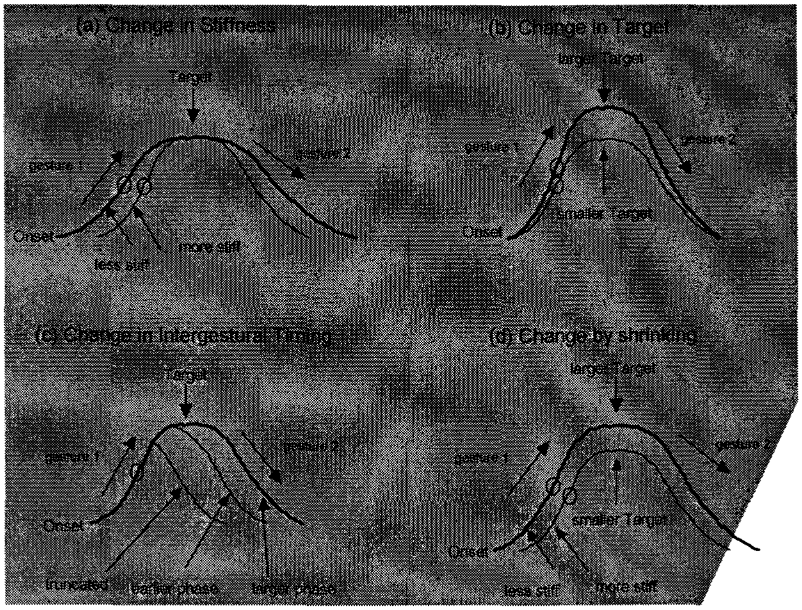

Figure 1. Hypothetical movement trajectories that correspond to a change in each parameter. (a) show change in stiffness (b) change in target (c) change in intergestural timing,; and (d) change by shrinking. Empty circles indicate the timepoint of the peak velocity attainment.

Some researchers (de Jong, 1991; Beckman, et al., 1992; Byrd, et al., 2000) have provided useful summaries of the kinematic consequences of various mass-spring equation parameter manipulations. Figure 1 shows schematized movement trajectories that correspond to changes in four dynamical parameters—i.e., stiffness, target, intergestural timing, and shrinking.

(1) *Stiffness*. Variation in articulatory movement duration is thought to be controlled by the stiffness parameter: the stiffer the spring (the articulator), the faster the movement. An idealized pattern in a pure change in stiffness is visualized in Figure 1a. If stiffness is the only parameter underlying kinematic differences, there should be a change in peak velocity, but not in displacement (i.e., the amount of the spatial distance that the articulator travels). This is because in the differential equation of the mass-spring dynamical system, the slope of the line relating displacement (in the x axis) to peak velocity (in the y axis) is equivalent to =(k/m)1/2 (where k = stiffness, m = mass), and therefore, a change in stiffness (k) will change the slope value (ω)

for displacement/ velocity, but not the displacement value.

Further, the time-to-peak-velocity (acceleration duration, the interval from the movement onset to the attainment of peak velocity) will vary as stiffness changes. Time-to-peak-velocity would be longer when the movement slows down due to decreased stiffness. (Note that some investigators (e.g., Byrd & Saltzman, 1998; Byrd, et al., 2000; and Byrd, 2000) suggested that one of the importance kinematics measures that best characterizes the dynamical stiffness parameter is absolute time-to-peak-velocity.

(2) *Target (underlying articulatory amplitude).* A change in target induces a change in displacement with the interval from the onset to the target being held constant. In a pure target change, peak velocity and displacement changes proportionally without a change in duration. This is because in an increase in target with stiffness being held constant, articulators have to travel farther with no extra time (as can be seen in Figure 1.b).

(3) *Intergestural timing or truncation.* The truncation parameter is based on the assumption that the articulatory movement towards the target can be 'truncated' by an early activation of the following gesture, which keeps the articulatory movement from reaching its assumed target, as can be seen in Figure 1c. Thus, under a pure change in intergestural timing, there should be no change in peak velocity because, for example, the effect of a substantially earlier following gesture is to prevent the preceding gesture from reaching its target (i.e., truncation of the preceding gesture) while stiffness and target specifications remain unchanged.

(4) *Shrinking.* Shrinking can be defined as a change in both target and stiffness which are scaled proportionally. Shrinking can be thought of as a unique dynamical parameter that may underlie prosodically conditioned kinematic variation (see Harrington, et al. (1995), and Byrd, et al. (2000)). As can be seen in Figure 1d, in a pure proportional change in target and stiffness, there would be a proportional increase in both duration and displacement, which results in no change in peak velocity.

Under the basic assumptions just described above, researchers have attempted to characterize systematic variation due to prosodic factors by pinpointing one of these dynamical parameter settings as an underlying mechanism. For example, in an effort to characterize accent-induced articulatory variation, some researchers (e.g., Edwards, et al., 1991; Beckman, et al., 1992) examined the kinematics of jaw movements, and found that under accent, the jaw opening gesture is associated with an increase in duration and displacement without a substantial increase in peak velocity. Based on this, they suggest that a change in intergestural timing is the major dynamical mechanism that governs accent-induced kinematic variations. Harrington, et al. (1995) further tested the intergestural timing hypothesis by comparing the movement trajectory of the unaccented jaw opening and closing gestures for the Australian English word 'barb' against simulated movement trajectories representing truncated and linearly shrunk unaccented gestures. The actual movement trajectory of the unaccented gestures were found to match more closely with the trajectory simulated by truncation (intergestural timing) than to the one simulated by shrinking. This was interpreted as favoring the intergestural timing account.

However, the intergestural timing account has been controversial. For example, the actual kinematic data reported in Harrington, et al. (1995) showed that the peak velocity for the jaw opening gesture was consistently faster under accent for all three speakers, which cannot be fully accounted for by the intergestural timing account alone. (Note that under a pure change in intergestural timing, no change in peak velocity is expected.) In fact, Harrington, et al. also noted the possibility that a combination of truncation and shrinking underlies the articulation of unaccented vowels.

De Jong (1991) also examined differences in the kinematics of jaw movement for the English words 'on' and 'to' as a function of accent, and found that what is commonly observed for both 'on' and 'to' under accent is a greater displacement, suggesting that if anything, the target (or amplitude) is bigger when accented. However, de Jong himself did not entirely rule out the intergestural timing account, adding that "[i]n more extremely stressed cases this amplitude [target] change can be accompanied by a timing change—highly accented consonant gestures being initiated later than their accented counterparts" (de Jong, 1991:125).

Furthermore, in investigating articulatory kinematics of jaw movement for accented 'Pope' and 'Pipe,' Fowler (1995) found that for 'Pope,' all three speakers exhibited larger, longer, and faster movement under accent. For 'Pipe,' one speaker showed the same pattern as for 'Pope,' while another speaker showed longer, faster, but not larger movement; the third speaker showed larger, faster, but not longer movement. It is only this last speaker's pattern, and only for 'Pipe,' that shows one clear case that can be interpretable in terms of a mass-spring gestural model: the larger, faster, but not longer movement supports the target account. In all other cases, no dynamical parameter setting could be identified as an underlying mechanism for accentuation. The lip opening and closing kinematic data in Cho (2006) also demonstrated larger, longer, and faster articulation under accent which suggests that not a particular dynamical parameter setting can be pinpointed to account for accent-induced articulatory variation.

Turning to the questions regarding boundary-induced articulatory variations, researchers have been interested in whether the same dynamical account of accent-induced kinematic variations may apply to the boundary-induced kinematic variations. In fact, a few relevant studies (Edwards, et al., 1991; Byrd & Saltzman, 1998; Byrd, et al., 2000; Byrd, 2000) converge on the conclusion that the dynamical mechanism governing kinematic variations at edges of prosodic domains differs from the one that governs accent-induced variations. A change in the stiffness parameter has been taken to be a dynamical mechanism governing boundary-induced kinematic variations (as opposed to either intergestural timing or target changes that have been taken to be most likely responsible for accent-induced articulatory variation). For example, Byrd & Saltzman (1998) investigated the lip opening and closing gestures for English bilabial /m/ at edges of prosodic domains, and found that movement duration and time-to-peak velocity are highly correlated. Based on this, they suggested that a local stiffness change may be the source of variation in boundary-adjacent lengthening: The higher the boundary, the less the gestural stiffness, which is consistent with the interpretation made by Edwards, et al. (1991) for the domain-final jaw closing gesture and by Byrd (2000) for the transboundary tongue movement from /a/ to /i/.

4.2 The abstract prosodic gesture, the 'π-gesture'

In an effort to characterize the prosodic edge effect, Byrd and her colleagues (Saltzman, 1995; Byrd, et al., 2000; Byrd, 2000; Byrd & Saltzman, 2003) have suggested that there might be some sort of prosodic boundary gestures that are governed by prosodic constituency in the task dynamics model. This abstract and non-tract variable 'prosodic' gesture is called 'π-gesture,' which was initially hypothesized to affect degree of stiffness in non-prosodic articulatory gestures over its activation period, roughly in proportion to the strength of the boundary.

More crucially, Byrd & Saltzman (2000) suggested that a clock-slowing implementation of π-gestures may affect the activation time course (rather than the stiffness) of all the dynamical parameters. Consequently, boundary strength determines degree of clock slowing, such that the gestures adjacent to a strong prosodic boundary will get slower, and possibly spatially larger, as shown in their simulations. Relatedly, with respect to the π-gesture's temporal domain, Byrd (2000) assumes that the π-gesture is anchored at a prosodic boundary (i.e., edges of prosodic domains), such that its clock-slowing effect is stronger near or at the domain-edges, dwindling farther from the edge.

The theory of the π-gesture, however, does not yet provide a full-fledged account of postboundary (domain-initial) articulation. First, the exact scope of the π-gesture's influence in the temporal dimension is not yet clear. In an earlier study, Byrd (2000) suggested that the lengthening of the transboundary V1-to-V2 movement duration at prosodic junctures was due primarily to preboundary lengthening. Byrd claimed it to be a consequence of a stronger effect of the π-gesture on the preboundary V1: that is, in a C1V1#C2V2 sequence V1 is closer to the prosodic boundary than V2 is, so that V1 is more heavily influenced by the π-gesture than V2 is. However, Cho (2002, 2006) showed that boundary-induced durational variation is evident quite equally in both the preboundary and the postboundary lengthening. This is in fact compatible with the results of Byrd and her colleagues' later study (Byrd, et al., 2006) which specifically investigated the temporal scope of the prosodic boundary effects by examining the tongue tip opening and closing movement for coronal consonants. They showed that temporal effects on both preboundary and postboundary articulation are by and large equivalent, especially when it comes to articulation immediately adjacent to the prosodic boundary, although compensatory shortening was observed for the second and the third syllables. What remains unclear is then exactly how far the effects of the π-gesture can be extended around prosodic junctures and exactly how the scope is determined.

Another area of the π-gesture theory that needs improving is how it captures possible spatial expansion at prosodic boundaries. A large body of experimental work that supported the π-gesture hypothesis has characterized the nature of articulation at prosodic boundaries primarily in temporal dimension, as the central clock, which controls the rate of articulatory activation of constriction gestures, is slowed down (Byrd, 2000; Byrd & Saltzman, 2003; Byrd, et al., 2006). But other studies have also shown that boundary-induced articulatory variation can also engender spatial expansion as well (Keating, et al., 1997; Cho, 2006; Tabain, 2003b). Byrd & Saltzman (2003) indeed showed that domain-initial spatial expansion of consonantal articulation (e.g., Keating, et al., 2003; Fougeron, 2001; Cho & Keating, 2001) can be simulated by a clock-slowing implementation of the π-gesture which may reduce articulatory overlap between domain-initial consonantal gesture and the following vocalic gesture, hence untruncated articulatory target of the consonant gesture.

The π-gesture model surely provides a possible way to unify dynamical accounts for symbolic prosodic and segmental units by relating them to the abstract π-gesture and the tract-variable articulatory gestures, respectively. It is hoped that its advocates develop the model more fully to account for both temporal and spatial variation at prosodic boundaries in a cognitively explanatory and computationally implementable way.

5. The role of prosodic strengthening in spoken word recognition: A case of domain-initial strengthening

We have so far seen how an abstract prosodic structure is phonetically manifested from the perspective of speech production. A question that naturally follows is then 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 Gow, Melvold & Manuel, 1996). For instance, in the Cohort model (e.g., Marslen-Wilson, 1987), 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. More specifically, Fougeron & Keating (1997) discussed the possible benefits that listeners might receive from domain-initial strengthening, including assistance with lexical segmentation and lexical access: 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, hence facilitation of lexical access. Likewise, Cho & 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. (See also Kim, 2004a, b, for the use of the increased amplitude in domain-initial position in artificial language learning by Korean listeners.)

Cho, McQueen & Cox (2007) 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., [tɪ]) 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. (2007), is that domain-initial strengthening is one of many acoustic cues used in the segmentation of continuous speech, and that more broadly speakers signal prosodic structure in a systematic fine-grained phonetic detail and listeners make use of it in speech comprehension.

6. 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

Beckman, M. E. & Elam, G. A. 1997. *Guidelines for ToBI Labelling*, version 3.0. Unpublished Manuscript. Ohio State University.

Beckman, M. E. & Pierrehumbert, J. 1986. "Intonational structure in Japanese and English." *Phonology Yearbook*, 3, 255-309.

Beckman, M.E., Edwards, J. & Fletcher, J. 1992. "Prosodic structure and tempo in a sonority model of articulatory dynamics." In G. Docherty and B. Ladd (eds.), *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*, pp. 68-86. Cambridge: Cambridge University Press.

Berkovits, R. 1993. "Utterance-final lengthening and the duration of final-stop closures." *Journal of Phonetics*, 21, 479 - 489.

Browman, C. & Goldstein L. 1990. "Tiers in articulatory phonology, with some implications for casual speech." In J. Kingston and M.E. Beckman (eds.), *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech*, pp. 341 - 376. Cambridge: Cambridge University Press.

Browman, C. & Goldstein, L. 1992. "Articulatory phonology: An overview." *Phonetica*, 49, 155-180.

Byrd, D. 2000. "Articulatory vowel lengthening and coordination at phrasal junctures." *Phonetica*, 57, 3-16.

Byrd, D. 2006. "Relating prosody and dynamic events: Commentary on the papers by Cho, Navas, and Smiljanić." In L. Goldstein, D. Whalen and C. Best (eds.), *Laboratory Phonology 8: Varieties of Phonological Competence*, pp. 549-561. Berlin/New York: Mouton de Gruyter.

Byrd, D., Kaun, A., Narayanan, S. & Saltzman, E. 2000. "Phrasal signatures in articulation." In M. Broe and J. Pierrehumbert (eds.), *Acquisition and the Lexicon: Papers in Laboratory Phonology V*, pp. 70-88. Cambridge, UK: Cambridge University Press.

Byrd, D., Krivokapić, J. & Lee, S. 2006. "How far, how long: On the temporal scope of prosodic boundary effects." *Journal of the Acoustical Society of America*, 120, 1589-1599.

Byrd, D. & Saltzman, E. 2003. "The elastic phrase: modeling the dynamics of boundary-adjacent lengthening." *Journal of Phonetics*, 31, 149-180.

Cambier-Langeveld, T. 2000. *Temporal Marking of Accent and Boundaries*. Ph.D. diss., University of Amsterdam (LOT series 32).

Cho, T. 2002. *The Effects of Prosody on Articulation in English*. New York: Routledge.

Cho, T. 2004. "Prosodically-conditioned strengthening and vowel-to-vowel coarticulation in English." *Journal of Phonetics*, 32, 141-176.

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.

Cho, T. 2006. "Manifestation of prosodic structure in articulation: Evidence from lip kinematics in English." In L. Goldstein, D. Whalen and C. Best (eds.), *Laboratory Phonology 8: Varieties of Phonological Competence*, pp.519-548. Berlin/New York: Mouton de Gruyter.

Cho, T. & Jun, S. 2000. "Domain-initial strengthening as featural enhancement: Aerodynamic evidence from Korean." In *Chicago Linguistics Society*, 36, 31-44.

Cho, T. & Keating, P. 2001. "Articulatory and acoustic studies of domain-initial strengthening in Korean." *Journal of Phonetics*, 29, 155-190.

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, 121-157

Cho, T., McQueen, J. & Cox, E. 2007. "Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English." *Journal of Phonetics*, 35, 210-243.

Cole, J., Kim, H., Choi, H. & Hasegawa-Johnson, M. 2007. "Prosodic effects on acoustic cues to stop voicing and place of articulation: Evidence from Radio News Speech." *Journal of Phonetics*, 35, 180-209

de Jong, K. 1991. *The Oral Articulation of English Stress Accent*. Ph.D. diss., Ohio State University.

de Jong K. 1995a. "The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation." *Journal of the Acoustical Society of America*, 97, 491-504.

de Jong, K. 1995b. "On the status of redundant features: the case of backing and rounding in American English." In B. Connell and A. Arvaniti (eds.), *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV*, pp. 68-86. Cambridge, UK: Cambridge University Press.

de Jong, K., Beckman, M. E. & Edwards, J. 1993 "The interplay between prosodic structure and coarticulation." *Language and Speech*, 36, 197-212.

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.

Erickson, D. 2002. "Articulation of extreme formant patterns for emphasized vowels." *Phonetica*, 59, 134-149.

Fougeron, C. 1999. "Prosodically conditioned articulatory variations: A review." *UCLA Working Papers in Phonetics*, 97, 1-74.

Fougeron, C. 2001. "Articulatory properties of initial segments in several prosodic constituents in French." *Journal of Phonetics*, 29, 109-135.

Fougeron, C. & Keating, P. A. 1997. "Articulatory strengthening at edges of prosodic domains."

Journal of the Acoustical Society of America, 106, 3728-3740.

Fowler, C. A. 1980. "Coarticulation and theories of extrinsic timing." *Journal of Phonetics*, 8, 113-133.

Fowler, C. A. 1995. "Acoustic and kinematic correlates of contrastive stress accent in spoken English." In F. Bell-Berti and L. Raphael (eds.), *Producing Speech: Contemporary Issues: For Katherine Safford Harris*, pp. 355 - 373. Melville, Long Island, New York: AIP Publishing.

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.

Gussenhoven, G. & Rietveld, A. C. M. 1992. "Intonation contours, prosodic structure and preboundary lengthening." *Journal of Phonetics*, 20, 283-303.

Harrington, J., Fletcher, J. & Roberts, C. 1995. "Coarticulation and the accented/unaccented distinction: Evidence from jaw movement data." *Journal of Phonetics*, 23, 305-322.

Hawkins, S. 1992. "An introduction to task dynamics." In G. Docherty and B. Ladd (eds.), *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*, pp. 9-25. Cambridge: Cambridge University Press.

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

Jun, S.-A. 1993. *The Phonetics and Phonology of Korean Prosody*. Ph.D. diss., Ohio State University.

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.

Keating, P. A., Cho, T., Fougeron, C. & Hsu, C. 2003. "Domain-initial strengthening in four languages." In J. Local, R. Ogden and R. Temple (eds.), *Papers in Laboratory Phonology 6: Phonetic Interpretations*, pp. 145-163. Cambridge, UK: Cambridge University Press.

Kim, S. 2003. "Domain-initial strengthening of Korean fricatives." *Harvard Studies in Korean Linguistics*, 9, 164-173.

Kim, S. 2004a. *The Role of Prosodic Phrasing in Korean Word Segmentation*. Ph.D. diss., University of California, Los Angeles.

Kim, S. 2004b. "The Role of Prosodic Cues in Word Segmentation of Korean." In *Proceedings of the 8th International Conference on Spoken Language Processing*, Interspeech 2004. Jeju, Korea.

Lehiste, I. 1970. *Suprasegmentals*. Cambridge, Mass.: MIT Press.

Lindblom, B. 1990. "Explaining phonetic variation: A sketch of the H & H theory." In W.J. Hardcastle and A. Marchal (eds.), *Speech Production and Speech Modeling*, pp. 403-440. Dordrecht: Kluwer Academic Publishers.

Macchi, M. J. 1985. *Segmental and Suprasegmental Features and Lip and Jaw Articulators*. Ph.D. diss.. New York University.

Marslen-Wilson, W. D. 1987. "Functional parallelism in spoken word-recognition." *Cognition*, 25, 71-102.

Mooshammer, C. & Fuchs, S. 2002. "Stress distinction in German: Simulating kinematic parameters of tongue tip gestures." *Journal of Phonetics*, 30, 337-355.

Nespor, M. & Vogel, I. 1986. *Prosodic Phonology*. Dordrecht: Foris.

Öhman, S. E. G. 1966. "Coarticulation in VCV utterances: Spectrographic measurements." *Journal* *of the Acoustical Society of America*, 39, 151-168.

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.

Pierrehumbert, J. 1980. *The Phonology and Phonetics of English Intonation.* Ph.D. diss., MIT.

Pierrehumbert, J. & Hirschberg, J. 1990. "The meaning of intonational contours in the interpretation of discourse." In R.R. Cohen, J. Morgen, and M.E. Pollack (eds.), *Intentions in Communication*, pp. 271-311. Cambridge, MA: MIT Press.

Saltzman, E. 1995. "Intergestural timing in speech production: Data and modeling." In *Proceedings of the XIII International Congress of Phonetic Sciences*, 2, 84-91.

Saltzman, E. & Munhall, K. G. 1989. "A dynamical approach to gestural patterning in speech production." *Ecological Psychology*, 1, 333-82.

Shattuck-Hufnagel, S. & Turk, A. E. 1996. "A prosody tutorial for investigators of auditory sentence processing." *Journal of Psycholinguistic Research*, 25(2), 193-247.

Tabain, M. 2003a. "Effects of prosodic boundary on /aC/ sequences: Acoustic results." *Journal of the Acoustical Society of America*, 113, 516-531.

Tabain, M. 2003b. "Effects of prosodic boundary on /aC/ sequences: Articulatory results." *Journal of the Acoustical Society of America*, 113, 2834-2849.

Tabain, M. & Perrier, P. 2005. "Articulation and acoustics of /i/ in preboundary position in French." *Journal of Phonetics*, 33, 77-100.

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.

received: October 25, 2007
accepted: November 25, 2007

▲ Taehong Cho
Department of English Language and Literature, Hanyang University
Haengdang-dong 17, Seongdong-gu, Seoul, 136-790, Korea
Homepage: http://tcho.hanyang.ac.kr
Tel: +82-2-2220-0746
Email: tcho@hanyang.ac.kr