Cho (in press, 2011). Laboratory phonology. In N.C. Kula, B. Botma, & K. Nasukawa (Eds.), The Continuum Companion to Phonology. London/New York: Continuum.

14 Laboratory Phonology

Taehong Cho

Chapter Overview

1. The Phonetics-Phonology Interface and the Emergence of Laboratory Phonology

What is phonetics and what is phonology? Answering this question is not as simple as it might seem since both disciplines are concerned with speech sounds and sound systems of human language. A common answer to this question would be that phonetics deals with the description of the physical properties of speech sounds that are continuous and gradient in temporal and spatial dimensions, whereas phonology is concerned with abstract aspects of the sound system, describing how the abstract representations of speech units are formed and organized in a categorical way to make contrastive distinctions among them. The distinction between these two subfields of linguistics crystallized in the tradition of the *Sound Pattern of English* (SPE) (Chomsky and Halle 1968). Within this tradition, an extreme view is that the phonetic component is not part of the grammar but 'physical' and 'automatic', and thus something that can be studied outside the realm of linguistics.

However, the severance between phonetics and phonology has been steadily eroding since the 1980s as researchers began to become aware, thanks to meticulous and scientific methodologies adopted in studying speech sounds, of the importance of scalar and gradient aspects of speech in understanding the linguistic sound system. In particular, non-contrastive phonetic events, which had traditionally been understood to be beyond the speaker's control (i.e. as low-level automatic physiological phenomena), have been reinterpreted as part of the grammar since they turn out to be either systematically linked with phonological contrasts or governed by language-specific phonetic rules. For example, non-contrastive vowel duration in English is in fact closely tied with the phonological voicing contrast of the postvocalic consonants. Similarly, stop epenthesis between a nasal and a homorganic fricative (e.g. insertion of [t] in *prince*), which was once considered an inevitable automatic effect of producing nasal-fricative sequences, is now understood to be controlled by the speaker and thus driven by a language-specific phonetic rule. For example, Fourakis and Port (1986) have observed that American English speakers insert a stop most of the time while English speakers in South Africa seldom do.

The necessity of language-specific phonetic rules can also be exemplified in the way languages choose phonetic values for the same phonological categories (cf. Keating, 1984; Kingston and Diehl, 1994; Cho and Ladefoged, 1999). For example, in investigating variation and universals in VOT patterns of voiceless stops in 18 languages, Cho and Ladefoged observed that although languages choose one of the three possibilities (unaspirated, aspirated and heavily aspirated) for the degree of aspiration of voiceless stops, they differ not only in choosing VOT values for unaspirated and aspirated stops (e.g. some languages have extremely polarized VOT values for the two phonological categories while others do not), but they also choose VOT values arbitrarily when they have only one voiceless stop category. Based on these observations, Cho and Ladefoged concluded that the phonetic output of a grammar must contain language-specific components, which differ in the target values they assign for the timing between the oral (stop release) gesture and the laryngeal (voicing) gesture for VOT.

Another important finding is that a number of phonological processes which were traditionally assumed to result in complete neutralization are in fact phonetically partial or gradient, showing that phonetic outputs resulting from some phonological processes may not be identical to phonetic outputs that are not phonologically modified. A good example of incomplete neutralization is the case of stop epenthesis mentioned above. Fourakis and Port (1986) show that the inserted [t] in *prince* is phonetically different from the underlying /t/ in *prints*. Another type of incomplete neutralization is found in some assimilatory processes. For example, an electropalatographic (EPG) study by Zsiga (1995) showed that palatalization of English /s/ before a palatal glide /j/ (e.g. /s/ → [ʃ] /__ j, as in *confe*[ʃ]*ion*) can be gradient when the /s + j/ sequence is created post-lexically across word boundaries (e.g. *confess your*). Similarly, the [p] in *righ*[p] *berry* (*right berry*), which results from assimilation to /b/, is indeed acoustically different from the underlying /p/ in *ripe berry* (e.g. Gow, 2002).

A number of experimental studies, including the aforementioned ones, have led to a gradual consensus that many, if not all, phonetic events that are expressed by gradient and scalar spatio-temporal dimensions are in fact controlled by the speaker and should be considered as being governed by either the phonological system or by the 'phonetic grammar' of the language (e.g. Keating, 1985).

The increasing awareness of language-specific phonetic rules and the close relation between gradient and categorical aspects of speech have thus sparked a fundamental rethinking as to where we should draw the line between phonetics and phonology, or whether a line between them is even necessary (e.g. Keating, 1985, 1990; Cohn, 1998, 2006, 2007; Hayes et al., 2004, among others). Articulatory Phonology is an example of a theoretical framework in which phonetics and phonology are integrated with a set of unified formal mechanisms – that is, phonological contrasts are directly expressed by coordination of articulatory gestures in temporal and spatial dimensions (e.g. Browman and Goldstein, 1992). Another theoretical framework that assumes no boundary between phonetics and phonology is so-called phonetically driven phonology, in which phonological representations are not viewed as categorical but instead as embodying functional phonetic events such as non-contrastive duration, formant transitions and consonant release characteristics (see among others Steriade, 1993b, 1999c; Flemming, 1995; Boersma, 1998; Kirchner, 1998).

Thus far, I have discussed the phonetics-phonology interface at some length since a proper understanding of the sound systems of human language cannot be achieved without understanding their relationship (for this point, see also Hamann, this volume). Indeed, many phoneticians and phonologists have attempted to bridge phonetics and phonology in the past couple of decades, which has led to the inception and consolidation of the interdisciplinary 'cooperative' research community called 'Laboratory Phonology' (see Cohn, in press, for a thorough overview of the history and the significance of Laboratory Phonology). The Laboratory Phonology community, which started as a small conference in 1987, is now considered to be a particular 'approach to investigating human sounds and sound systems, taking as foundational the premise that progress will be achieved more successfully through integrated methodologies' (Cohn, in press). The concerns of Laboratory Phonology are aptly summarized by the three fundamental research questions that Beckman and Kingston (1990: 1) asked at the first Laboratory Phonology conference back in 1987:

> First, how, in the twin processes of producing and perceiving speech, do the discrete symbolic or cognitive units of the phonological representation of an utterance map into the continuous psychoacoustic and motoric functions of its phonetic representations? Second, how should the task of

> explaining speech patterns be divided between the models of grammatical function that are encoded in phonological representations and the models of physical or sensory function that are encoded in phonetic representations? And third, what sorts of research methods are most likely to provide good models for the two components and for the mapping between them?

Beckman and Kingston (1990: 3) go on to highlight the necessity of a so-called hybrid research methodology in answering these questions successfully, that is, a methodology that requires 'experimental paradigms that control for details of phonological structure [and] observational techniques that go beyond standard field methods'. Since then, researchers from different disciplines, including phoneticians, phonologists, psycholinguists, cognitive scientists and speech technologists, have been working towards this goal. Laboratory Phonology has now become codified as a truly multi-disciplinary approach bridging the gap between more theoretical and more empirical approaches to the investigation of the sounds and sound systems of natural language (cf. Cohn, in press).

2. The Phonetics-Prosody Interface

Among the various research questions and issues that have been actively addressed by the Laboratory Phonology community are those regarding the relationship between phonetics and prosody. In fact, it may not be too much to say that research in Intonational Phonology (e.g. Bruce, 1977; Pierrehumbert, 1980; Pierrehumbert and Beckman, 1988; Ladd, 1996) has been in the vanguard of work attempting to bridge categorical and gradient aspects of speech sounds, and as such stood at the basis of Laboratory Phonology. The main concern of Intonational Phonology was to bring gradient and varying physical F0 events into a systematically organized pattern by tying them with phonological representations of underlying tones (or tonal targets) through a mapping of categorical tonal representations to continuous F0 events, that is, the structure of the tune of a given utterance is assumed to be generated by phonetic interpolation between tonal targets.

The fundamentals of Intonational Phonology have thus been adopted in defining 'intonationally based' prosodic structure, which consists of structural entities such as prominence-lending locations (e.g. stressed or accented syllables) and prosodic junctures of different levels (see Section 2.1 for further discussion of prosodic structure; and Jun, 1998, for discussion on intonational versus syntactic approaches to defining prosodic structure). How is such phonologically defined prosodic structure manifested in fine-grained phonetic details? And how are sounds and sound systems (including application of phonetic and phonological rules) constrained and informed by prosodic structure? These questions have played a key role in research on the phonetics-prosody interface. The general consensus in the literature is that sounds and sound patterns of human languages can never be fully understood without understanding how they interact with prosodic structure in the grammatical system of the language (cf. Shattuck-Hufnagel and Turk, 1996). In the remainder of this chapter I will consider the phonetics-prosody interface in more detail. I will start with an introduction to the structural view of prosody. This is followed by an in-depth discussion on the manifestation of prosodic structure in speech production and its role in speech comprehension.

2.1 A Structural View of Prosody

A fundamental property of utterances is that they are prosodic in nature. An utterance is prosodic in that its segmental make-up is superimposed by prosodic (or suprasegmental) features such as F0, duration and amplitude; it is also prosodic in that such prosodic features are employed to build a prosodic structure of the utterance being spoken. Prosodic structure can be defined as 'a hierarchically organized structure of phonologically defined constituents and heads' (Beckman, 1996: 19). On this structural view, prosody reflects both constituent- and prominence-based hierarchies.

The constituent-based hierarchy is built with prosodic constituents. In English, these include Syllable, Foot, Prosodic (or Phonological) Word, Intermediate (Phonological) Phrase and Intonational Phrase (cf. Shattuck-Hufgnagel and Turk, 1996). The size of prosodic constituents becomes progressively larger with smaller constituents combining to form immediately larger ones in a hierarchical fashion – that is, one or more syllables are grouped into a prosodic word; one or more words combine to form an immediately larger prosodic constituent, the Intermediate Phrase (=ip); and finally one or more Intermediate Phrases combine to form the Intonational Phrase, which is often assumed to be the largest prosodic unit (e.g. Beckman and Pierrehumbert, 1986; but see Nespor and Vogel, 1986). These prosodic constituents are also regarded as prosodic domains in that they often serve as domains for certain intonational patterns and certain phonological rules (cf. Selkirk, 1984b, 1995b; Jun, 1998).

The prominence-based prosodic hierarchy is constituted with increasing degree of prominence from null stress, secondary lexical stress (if it exists), to primary lexical stress and to sentence stress (nuclear pitch accent) (cf. Liberman and Prince, 1977; Beckman, 1986; Hayes, 1989b; Beckman and Edwards, 1994).

This prominence-based hierarchy is closely intertwined with the constituent-based hierarchy in that the lexically stressed syllable serves as the head of the prosodic word, and the syllable with nuclear pitch accent as the head of the Intermediate Phrase (cf. Beckman and Edwards).

Prosodic structure thus serves a dual function in speech production – that is, prosodic boundary marking by which the hierarchical grouping of prosodic constituents is determined, and prominence marking by which relative prominence among prosodic constituents is determined. A single sentence (with the same lexical content and syntactic structure), however, may be produced with different prosodic structures, conditioned by a complex interaction of various factors, for example, 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 prosodic domain; and how fast the utterance is produced (e.g. Nespor and Vogel, 1986; Pierrehumbert and Hirshberg, 1990; Jun, 1993; Keating and Shattuck-Hufnagel, 2002). Put differently, different prosodic structures of the same sentence may be constructed online in speech production, conveying a great deal of linguistic and extralinguistic information. From the production perspective, prosodic structure is considered an essential element of speech production, modulating phonetic encoding which fine-tunes phonetic outputs so as to signal the prosodic structure (e.g. Keating and Shuttuck-Hufnagel, 2002). From the perception perspective, phonetic signatures of a particular prosodic structure are thought to be exploited by listeners in that they help them to process lexical, syntactic and discourse information (e.g. Cho et al., 2007; Carlson et al., 2001; Couper-Kuhlen and Selting, 1996). Thus, the structural view of prosody assumes that prosodic structure is an important grammatical entity in its own right (Beckman, 1996), and therefore a complete picture of speech production and comprehension cannot be obtained without understanding the interplay between phonetics and prosody at various levels of the production and comprehension process.

2.2 Phonetic Manifestations of Prosodic Structure

As an initial step to understand the phonetics-prosody interplay, an increasing number of researchers have endeavoured to explore how prosodic structure may be phonetically manifested. In particular, phonetic manifestations of prosodic structure have been interpreted in terms of 'prosodic strengthening', a phenomenon associated with prosodic landmark locations such as prosodic domain edges and prominent syllables. (Here, 'prosodic strengthening' is used as a cover term for spatial and/or temporal expansion that is the result of boundary and prominence marking (e.g. Cho, 2005, 2008; Cho and McQueen, 2005)).

Below, I will discuss how prosodic structure can be phonetically expressed in terms of boundary marking and prominence marking.

2.3 Prosodic Strengthening as Boundary Marking

2.3.1 Domain-Final Prosodic Strengthening (At the Right Edge of Prosodic Domains)

One of the Most consistent phonetic correlates of prosodic structure is temporal expansion of domain-final segments. The degree of this is closely correlated with the prosodic level or the prosodic boundary strength at the prosodic juncture (e.g. Edwards et al., 1991; Wightman et al., 1992; Gussenhoven and Rietveld, 1992; Berkovits, 1993; Byrd, 2000; Cambier-Langeveld, 2000; Byrd et al., 2006; Cho, 2006). Temporal expansion at the right edge of prosodic domains is usually accompanied by intonational marking, referred to as boundary tones (Beckman and Pierrehumbert, 1986; Pierrehumbert, 1980). Thus, a general consensus is that domain-final articulation is characterized primarily by temporal expansion, which together with boundary tones demarcates prosodic boundaries.

More recently, however, domain-final elements have also been shown to be accompanied by spatial expansion in some cases. For instance, the amount of linguopalatal contact for a domain-final (pre-boundary) vowel (as measured by Electropalatography, EPG) decreases as the boundary level increases (Fougeron and Keating, 1997). A decreased EPG contact indicates more vocalic opening at the end of larger prosodic domains. Subsequent studies with a magnetometer (Electromagnetic Midsagittal Articulography (EMA)) have revealed more evidence for spatial expansion of articulation in domain-final positions, for example, larger C-to-V displacement in French (Tabain, 2003), higher tongue position for /i/ but lower for /a/ in English (Cho, 2005), larger lip aperture for /ɑ/ and /i/ in English (Cho, 2006). Cho (2004) has also shown coarticulatory resistance of domain-final /ɑ, i/ in English, which has been interpreted as another type of articulatory strengthening. These studies thus suggest that domain-final elements may involve spatial expansion, though not as robustly as temporal expansion, the former often being inconsistent across speakers (cf. Byrd et al., 2006) or non-observable (Edwards et al., 1991; Beckman et al., 1992).

2.3.2 Domain-Initial Prosodic Strengthening (At the Left Edge of Prosodic Domains)

Another Prosodic landmark location is the other side of the prosodic domain, domain-initial position. This has been considered another locus for spatio-temporal expansions which mark prosodic boundary strength. Fougeron and Keating (1997) and others (Cho and Keating, 2001; Fougeron, 2001; Keating et al., 2003), for example, used electropalatography (EPG) to measure the degree of linguopalatal contact of domain-initial consonants, and showed that the strength of the consonant articulation, as reflected in degree of oral constriction and seal (closure) duration, increases cumulatively for each higher level in the prosodic hierarchy. Subsequent work has demonstrated similar domain-initial strengthening in various acoustic and articulatory dimensions across languages (Byrd and Saltzman, 1998; Byrd et al., 2000; Lavoie, 2001; Kim, 2003; Cho, 2005, 2006; Cho and McQueen, 2005; Byrd et al., 2006; Kuzla et al., 2007; Cho and Keating, 2009, among others). For example, it has been found cross-linguistically that aspirated stops are produced with longer VOTs in domain-initial than in domain-medial position (for English, see Cole et al., 2007; Cho and Keating, 2009; for Korean, see Jun, 1993, 1995; Cho and Jun, 2000; Cho and Keating, 2001; for Japanese, see Onaka, 2003; Onaka et al., 2003; for Taiwanese, see Hsu and Jun, 1998; Hayashi et al., 1999; for French, see Fougeron, 2001). The longer, that is, temporally expanded, VOTs can be interpreted as a consequence of strengthening of the glottal abduction gesture (Pierrehumbert and Talkin, 1992; cf. Cooper, 1991). In a fiberscopic study, Jun et al. (1998) indeed found larger glottal apertures in AP-initial position than in AP-medial position. (AP is short for Accentual Phrase, which is assumed to be an intermediate prosodic level in Korean; see Jun, 1993, 1995.)

A number of proposals have been advanced to account for the articulatory nature of domain-initial strengthening. Given the close relationship between temporal and spatial expansions associated with domain-initial position, Cho and Keating (2001) proposed an Articulatory Undershoot Hypothesis: in domain-initial position there is enough time to execute the articulatory action, which results in the full attainment of the assumed articulatory target; in domain-medial position, on the other hand, there is some articulatory undershoot due to the insufficient durations associated with these positions. Fougeron (1999) suggested that domain-initial strengthening is ascribable to 'articulatory force' (Straka, 1963), which can be defined as 'the amount of energy necessary to the realization of all the muscular effort involved in the production of a consonant' (Delattre, 1940, translated). In domain-initial position consonants are produced with greater articulatory force, which causes contraction of the muscles involved in the articulation. Based on this assumption, domain-initial strengthening has often been referred to as domain-initial 'articulatory' strengthening (Fougeron and Keating, 1997; Cho and Keating, 2001; Keating et al., 2003). Greater articulatory force in domain-initial position is supposed to apply to the following vowel in such a way as to enhance the aperture contrast between C and V. (See Section 3.1.1 for further discussion on CV enhancement.) However, it remains as yet unclear whether domain-initial strengthening is strictly confined to initial consonants and, if it is not, how it relates to sonority expansion in those cases where the domain-initial segment is a vowel.

2.4 Prosodic Strengthening as Prominence Marking

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 and Fuchs, 2002; Cho, 2006, among others). Prominence marking by accent (sentence stress or nuclear pitch accent) is acoustically manifested with greater F0 movement, longer duration, greater amplitude and unreduced vowel quality (cf. Lehiste, 1970; Beckman, 1986). Articulatorily, it is accompanied by an increased respiratory effort. Based on EMG (Electromyogram) studies, Ladefoged and his colleagues (e.g. Ladefoged, 1967; Ladefoged and Loeb, 2002) showed that syllables with stress, both lexical and sentential, are produced with increased respiratory power due to additional activity of the internal intercostals, which may be closely linked to some characteristics of suprasegmentals under accent. For example, increased subglottal pressure due to heightened respiratory power would result in increased amplitude (see Lehiste, 1970, for discussion).

Prominence is also marked by supralarygeal articulation, which as Cho (2006) puts it, is 'simply bigger in all ways – in distance, time, and speed.' For example, the jaw opening gesture under accent is associated with an increase in duration and displacement, sometimes with a faster movement speed (Fowler, 1995) and sometimes without it (Beckman et al., 1992). The C-to-V lip opening gesture under accent has also been found to be associated with an increase in spatio-temporal expansion along with a faster movement speed (Cho, 2006).

2.5 Boundary Versus Prominence Marking

One of the important questions that has been explored by researchers in the past few decades is whether speakers differentiate between boundary and prominence information in speech production, and if so, how. This question is motivated by the assumption that prosodic strengthening serves a dual function (i.e. boundary marking and prominence marking). If it does, then speakers may feasibly differentiate between the two kinds of strengthening, given that they signal different aspects of prosodic structure. As discussed in the previous sections, phonetic signatures of boundary and prominence markings are similar in some aspects and different in others. To recap, the most obvious similarity has been found in temporal expansion, in that both boundary and prominence markings are accompanied by lengthening. Both also show some spatial dimension, but differ in that prominence marking is more likely to come with expansion of vocalic displacement whereas boundary marking tends to be characterized primarily by heightened consonantal constriction domain-initially and sometimes expanded vocalic displacement domain-finally. Further, speed of articulatory movement is consistently higher under accent, whereas articulatory movements are not necessarily faster when marking prosodic boundary, neither domain-initially nor finally. These findings, as they are now understood, clearly suggest that the two kinds of prosodic strengthening are indeed differentially manifested in speech production, which supports the view that they are separately encoded in speech planning (cf. Keating, 2006; Cho and Keating, 2009); see Section 3.3 for further discussion on this.

3. Issues in Prosodic Strengthening in Speech Production

Thus far, I have discussed some basic phonetic correlates of prosodic structure which can be broadly characterized as prosodic strengthening as a function of boundary and prominence markings. In the following subsections I will discuss some important issues regarding the production aspects of prosodic strengthening that have been considered in speech prosody literature.

3.1 Prosodic Strengthening and Linguistic Contrast Maximization

From a phonological point of view, a fundamental question is how prosodic strengthening relates to linguistic contrast. An important assumption is that if segments are articulatorily strengthened in prosodic landmark locations, this could result in a linguistically significant heightening of phonetic 'clarity', and hence be associated with enhancement of linguistic contrast. Given the observed phonetic differences between boundary and prominence markings, researchers have proposed hypotheses about different types of linguistic contrast maximization, which can be said to fall under the rubrics of *syntagmatic* and *paradigmatic* contrast enhancement (Beckman, 1996; Hsu and Jun, 1998; Cho and Jun, 2000; see Fougeron, 1999, for a review).

3.1.1 Syntagmatic Contrast Enhancement

Boundary marking has often been thought to be syntagmatically, or structurally, motivated, in that it results in enhancement of contrast between neighbouring segments at prosodic junctures (see Fougeron, 1999, for a review). There is a clear connection between this type of enhancement and domain-initial strengthening. Previous studies suggest that what is strengthened domain-initially is the 'consonantality' of the segment. An increased oral constriction degree augments #CV displacement domain-initially and V#C displacement across a prosodic boundary. Longer closure duration and aspiration augment the voicelessness of oral consonants (making them less sonorous), and therefore enhance the contrast with the sonority of neighbouring vowels. Similarly, reduced nasal airflow (Fougeron, 2001) and nasal energy (Cho and Keating, 2009) of /n/ in domain-initial position make this consonant less sonorous and therefore more consonantal. Furthermore, vocalic opening expansion, which is often observed in both pre-boundary and post-boundary positions, can be interpreted as sonority expansion of vowels, which, together with augmented consonantality, contributes to the enhancement of #CV and V#C contrasts.

3.1.2 Paradigmatic Contrast Enhancement

Contrary to boundary-induced syntagmatic enhancement, prominence marking is thought to give rise to paradigmatic contrast enhancement, that is, maximization of the phonemic distinction of contrastive sounds. De Jong (1995) refers to this as 'localized hyperarticulation'. He observed that the English vowel /ʊ/ is produced with a lowered jaw and tongue when accented. This is consistent with an earlier account of prominence marking – the sonority expansion hypothesis – which predicts that the amount of mouth opening, which is related with amplitude, is greater under accent (Beckman et al., 1992, Edwards and Beckman, 1988). However, de Jong also observed that the tongue body is more retracted, enhancing the backness feature of the vowel, which cannot be entirely ascribable to sonority expansion. He therefore proposed that certain accented segments are produced with more extreme articulatory movements in a direction that results in an enhancement of the distinctive features of segments. (See also Cho, 2005, who showed that accent induces fronting of the tongue for English /i/, enhancing the advancement feature.) These, then, are examples of localized hyperarticulation, which is based on Lindblom's (1990) notion of hyperarticulation, but is assumed to apply not just to extended discourses, but also locally to individual syllables. Thus, it can be used to characterize the extreme articulations in stressed/accented syllables, which ultimately lead to maximization of lexical distinctions. In a similar vein, Fowler (1995) also suggested that, based on the assumption that stress consists of a global increase in production effort (e.g. Öhman, 1966; Lehiste, 1970), accent-induced prosodic strengthening is driven by the prominence maximization principle, reflecting speakers' 'global effort' to increase the perceptual saliency of accented elements for the benefit of listeners.

3.1.3 Language Specificity and Contrast Enhancement

De Jong's localized hyperarticulation hypothesis is based on data from English, a stress-timed language in which lexical stress interplays with sentence stress. However, there are also languages that do not have any word-level prosody (e.g. lexical stress with nuclear pitch accent superimposed on it) and which therefore have a different prominence hierarchy than English has. A question that arises, then, is whether such languages (Korean is an example) differentiate linguistic contrast enhancements in a similar way as English does. This is an open question, which requires further research.

However, recent studies on Korean show some evidence that boundary marking may also be characterized by paradigmatic contrast enhancement (in line with localized hyperarticulation), at least as far as consonantal articulation is concerned. For example, in an acoustic-aerodynamic study Cho and Jun (2000) examined variations in VOT and airflow for domain-initial Korean stops, which display a three-way contrast between lenis, fortis and aspirated (cf. Cho et al., 2002). Cho and Jun found that for lenis and aspirated stops both VOT and the amount of airflow were greater domain-initially than domain-medially, that is, were more consonantal, in line with syntagmatic CV contrast enhancement. However, domain-initial fortis stops, which were produced with reduced VOT and airflow, showed the opposite pattern. The asymmetric domain-initial effects on laryngeal articulation (as reflected in VOT and airflow data) among different stops were interpreted as evidence for enhancement of laryngeal features. Following Lombardi (1991a and b), it was assumed that the three-way stop contrast in Korean is specified in terms of two privative laryngeal features, [spread glottis] and [constricted glottis]: aspirated and fortis stops are specified by [spread glottis] and [constricted glottis], respectively, whereas lenis stops are unspecified for either feature. The increased VOT and airflow for domain-initial aspirated stops were then viewed as enhancement of [spread glottis] whereas the reduced VOT and airflow for domain-initial fortis stops as an enhancement of [constricted glottis]. Cho and Jun further proposed that the unspecified lenis stop was still strengthened, but this time driven by syntagmatic CV contrast, explaining the increased VOT and airflow associated with domain-initial lenis stops. Importantly, the range of VOT and airflow variation for lenis stops rarely overlapped with that of aspirated stops, suggesting that the lenis stops' syntagmatic contrast enhancement was constrained in such a way that it does not blur the paradigmatic contrast between aspirated and lenis stops. (See Hsu and Jun, 1998 for a similar discussion on VOT variation in voiced, unaspirated and aspirated stops in Taiwanese.)

Cho and McQueen (2005) further asked whether domain-initial strengthening is constrained by language-specific phonological factors. They compared domain-initial strengthening in Dutch and English. Both languages have a two-way phonological contrast in stops (voiced versus voiceless), which can be specified with phonological features [±voice] (e.g. Kingston and Diehl, 1994). However, they differ in how the two stops are mapped onto phonetic categories. As Keating (1984) proposed, only three phonetic categories are needed to account for the phonetic realizations of a two-way phonological voicing contrast between stops in the world's languages: voiced, voiceless unaspirated and voiceless aspirated. The Dutch stops map onto the voiced and voiceless unaspirated categories, whereas English stops generally map onto as voiceless unaspirated and voiceless aspirated categories. Crucially, Cho and McQueen found shorter VOTs for Dutch /t/ in domain-initial position, as opposed to longer VOTs in that position in English. Based on the asymmetrical VOT patterns between the two languages, they proposed 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 shortened VOT in Dutch was also reportedly associated with /t/ in accented syllables, reflecting enhancement of [–spread glottis] as prominence marking under accent, which is again different from the lengthened VOT under accent in English.

Taken together, these results indicate that language-specific phonetic features are enhanced not only as boundary marking but also as prominence marking. Cho and McQueen (2005) proposed that there are cross-linguistic differences in the prosodic modulation of segment realization: the language-specific phonetic component of the grammar (cf. Keating, 1984) modulates the phonetics-prosody interplay. They concluded that prosodic strengthening is not simply a low-level phonetic event but a complex linguistic phenomenon which gives rise not only to enhancement of phonological/phonetic features, but also expresses positional strength, which may license phonological contrasts.

Some phonologists have indeed considered prosodically strong positions as 'privileged' or 'licensing' positions. In these positions, phonological contrasts are most often maintained and segments act as triggers of phonological modification of neighbouring segments (e.g. vowel harmony) but they themselves resist such a modification (e.g. Beckman, 1998; Steriade, 1999c; Barnes, 2002). V-to-V coarticulatory resistance found in prosodically strong locations may also be viewed as a result of contrast maintenance (Cho, 2004). However, as Cho (2005) points out, these approaches do not make clear whether such a positional privilege is phonetically grounded and attributable to the richness of the phonetic cues associated with that position (Steriade, 1999) or structurally driven and purely attributable to the position itself (Beckman, 1998). Nonetheless, what is clear is that prosodic strengthening is closely linked to maintenance or maximization of phonological contrasts, and is modulated by the language-specific sound system.

3.2 Models of Prosodic Strengthening

In the previous section I discussed how prosodic strengthening can be related to linguistic contrast enhancement. Such enhancements, however, can be achieved only through fine-tuning of articulation of particular segments. In this section I will discuss some possible mechanisms which might further illuminate the nature of prosodic strengthening, with special reference to domain-initial strengthening. An important question is how prosodic structure influences the detailed operation of the articulators involved in strengthening.

3.2.1 Prosodic Strengthening as a Result of Dynamical Parameter Settings in a Mass-Spring Gestural Model

In Articulatory Phonology (e.g. Browman and Goldstein, 1990, 1992), which is based on a mass-spring task dynamic model (e.g. Saltzman and Munhall, 1989; see Hawkins, 1992, for an overview for non-specialists), linguistically significant vocal tract constrictions, dubbed 'gestures,' are viewed as the primitives of phonological representation. In the model, phonological contrasts are maintained in terms of what kind of gestures are involved and how they are temporally coordinated. From the point of view of task dynamics, gestures are described in terms of the behaviour of the abstract 'mass,' which is connected to a 'spring,' and a 'damper' in a critically damped mass-spring system. As Hawkins describes it, it is as if one end of the spring were fixed at the mass, and the other end to the gestural point attractor. As the point attractor moves to a different target location, the spring is stretched, and the mass is pulled towards the target location. The damper makes the mass-spring system critically damped, such that the gesture is generally realized as a one-directional movement towards the target: The mass does not oscillate due to the damping (i.e. it never reaches the target location, and is not pulled back to its original location), but it stays in the target region, continuously and slowly reaching the equilibrium position of the spring (therefore the mass asymptotes towards the equilibrium position).

In the model gestures are specified with a set of dynamical parameter values. Some of the relevant dynamical parameters include target (underlying amplitude), stiffness (or natural frequency) and intergestural timing. Execution of gestures is characterized by articulatory movements, which vary depending on values of these parameters specified for a given gesture. The stiffness parameter reflects the stiffness of the spring, such that the stiffer the string, the faster the articulatory movement; the target parameter determines the target location, such that the larger the target, the more displacement of the articulatory movement; the intergestural timing determines how early or late the following gesture is timed with the preceding gesture, such that the earlier the following gesture comes, the more overlap between the adjacent gestures and thus the more truncation of the preceding gesture. (For more recent development of Articulatory Phonology, see Goldstein et al., 2006; Goldstein, Pouplier et al., 2007; Goldstein, Chitoran et al., 2007; Saltzman et al., 2008.)

Since its introduction in the late 1980s, researchers working in Articulatory Phonology have attempted to account for prosodically conditioned speech variation in terms of gestural dynamics (e.g. Edwards et al., 1991; de Jong, 1991; Beckman et al., 1992; Harrington et al., 1995; Saltzman, 1995; Byrd and Saltzman, 1998, 2003; Byrd et al., 2000; Byrd, 2000; Byrd et al., 2006; Cho, 2006, 2008), and suggested that prosodically conditioned articulatory variation may be controlled by dynamical parameter settings.

As to accent-induced articulatory variations, some researchers found that articulation under accent is accompanied by an increase in duration and displacement without a concomitant change in peak velocity (see, for example, Edwards et al., 1991; Beckman et al., 1992). As no change in peak velocity in kinematics can be interpreted as a result of no change in stiffness, they suggested that accent-induced kinematic variations are better accounted for by a change in intergestural timing. However, later kinematic studies indicated that while the intergestural timing parameter may be involved with articulatory movements under accent (Harrington et al., 1995; de Jong, 1991), it cannot be the sole controlling parameter as increased movement displacement under accent is often associated with an increase in movement velocity, which can be better accounted for by a change in the target parameter – that is, the larger the target, the faster the movement (cf. de Jong, 1991; Fowler, 1995; Cho, 2006). Recall that Cho (2006) characterizes accent-induced articulatory strengthening as 'big in all ways' in that it is accompanied by larger, longer and faster articulatory movements, which suggests that rather than a particular dynamical parameter setting, multiple parameters must be involved in accent-induced articulatory variation.

Turning to boundary-induced strengthening, one question that has been raised is whether the same dynamical account of accent-induced kinematic variations may apply to boundary-induced articulation. Results of some previous kinematic studies (Edwards et al., 1991; Byrd and Saltzman, 1998; Byrd et al., 2000; Byrd, 2000) have converged on the conclusion that the dynamical mechanism governing boundary-induced strengthening is indeed different from that of accent-induced strengthening. A change in the stiffness parameter has been considered as a plausible explanation for boundary-induced strengthening (as opposed to either intergestural timing or target changes, which in some cases may better account for accent-induced strengthening). For example, the lip opening and closing gestures for English bilabial /m/ at edges of prosodic domains showed that movement duration and time-to-peak velocity were highly correlated (Byrd and Saltzman, 1998), suggesting that a local stiffness change may be the source of variation in boundary-adjacent lengthening: the higher the boundary, the less the gestural stiffness (see also Edwards et al., 1991, for the domain-final jaw closing gesture and Byrd (2000) for the transboundary tongue movement from /a/ to /i/). But again the stiffness account cannot be the complete story since it fails to explain the often observed boundary-induced spatial expansion. Nevertheless, taken together, these studies suggest that speakers must exploit differential dynamical mechanisms, which distinctively govern prominence versus boundary markings. However, it is still unclear whether and exactly how these are controlled differentially by particular sets of dynamical parameters.

3.2.2 π-Gesture as a Device Modulating Boundary-Adjacent Articulation

As previous studies have not been entirely successful in pinpointing the exact dynamical parameter setting underlying boundary-adjacent articulation, Byrd and her colleagues (Saltzman, 1995; Byrd et al., 2000; Byrd, 2000, 2006; Byrd and Saltzman, 2003; Byrd et al., 2006) have proposed that boundary-induced articulation can be better understood as a result of the influence of a so-called π-gesture that is governed by prosodic constituency in the task dynamics model. The π-gesture is an abstract and non-tract variable 'prosodic' gesture, which determines articulatory movement speed, modulating the rate of the clock that controls articulatory activation of constriction gestures. (A 'non-tract variable' is a gesture that is not actually realized in terms of vocal tract constrictions.) As the clock, controlled by the π-gesture, is slowed down at a prosodic juncture, the articulatory movement at the juncture becomes slower, and possibly spatially larger. In temporal domains the π- gesture is anchored at a prosodic boundary, such that its clock-slowing effect is stronger at the juncture, dwindling farther from the edge. In line with this, some recent kinematic studies have shown that boundary-induced lengthening is most robustly manifested in the 'transboundary' articulatory movements (movements that start domain-finally and end domain-initially spanning the intervening juncture, as in /i/-to-/a/ movement in /i#Ca/ context), which are assumed to be under a direct influence of the π-gesture (see also Byrd et al., 2006 and Cho, 2006, 2008, for English data; Tabain, 2003 and Tabain and Perrier, 2005, for French data).

However, Cho (2008) observes that the π-gesture approach raises a number of issues. One concerns the question exactly how far the boundary-adjacent temporal influence can be extended, in particular with respect to the right of the prosodic boundary. Put differently, is the boundary effect on duration strictly local to domain-initial position? Previous studies have in fact shown mixed effects on domain-initial temporal expansion, suggesting that domain-initial strengthening is not strictly local to 'initial' consonantal articulation. On the one hand, quite a few studies (Fougeron and Keating, 1997; Cho and Keating, 2001;

Keating et al., 2003; Cho and McQueen, 2005; Cho and Keating, 2007, 2009) demonstrated that boundary effects are mainly local to C in domain-initial CV. On the other hand, some kinematic studies have indicated that articulatory movement lengthening effects can be pervasive into the vocalic articulation after the initial consonant (Byrd, 2000; Cho, 2006, 2008; Byrd et al., 2006). For example, Byrd et al. found lengthening of the articulatory opening movement from C to V after a boundary (though this was not observed for all speakers) and suggested that articulatory lengthening effects are roughly equivalent in both pre-boundary (domain-final) and post-boundary (domain-initial) articulations, especially as far as articulations immediately adjacent to the juncture are concerned. (Note, however, that post-boundary compensatory shortening was found, especially for consonantal gestures in onsets, in the second and the third syllables; cf. Krivokapić, 2007.) Cho (2006, 2008) also showed lengthening of vocalic gestures in domain-initial #CV lip opening movement and V-to-V tongue movement across a boundary (V#CV). These mixed results have been interpreted as suggesting that domain-initial effects on duration are not strictly local to the edge but may be gradient in nature, as a function of the distance from the boundary (Cho, 2008; Cho and Keating, 2009).

This gradience is indeed what the π-gesture predicts, that is, its influence wanes as it gets farther away from the boundary. Nevertheless, the mixed results do prompt further questions: exactly how far can the effects of the π-gesture be extended around prosodic junctures, and what factors influence the determination of its scope? Although answering theses questions requires substantial follow-up studies, a possible determining factor is the language-specific prosodic system. Barnes (2001, 2002) suggested that, in English at least, the vowel in CV syllables is not subject to domain-initial acoustic lengthening because vowel duration in this language is a major cue for lexical stress. (See also Keating et al., 2003 for discussion of how languages with different prosodic systems may show differential domain-initial strengthening effects.)

Another topic of debate concerns the question of how the π-gesture captures possible spatial expansion at prosodic boundaries. While a body of experimental work has characterized the nature of boundary-adjacent articulations in primarily temporal dimension, other studies (as they are now understood) have shown that boundary-induced spatial expansion affects not only initial consonants, but also often the following vocalic articulation. A simulation of a clock-slowing of the π-gesture (Byrd and Saltzman, 2003) showed that clock-slowing can reduce articulatory overlap between domain-initial consonantal gesture and the following vocalic gesture. The resulting non-truncation of the articulatory target of the consonant gesture may explain consonantal strengthening effects (which is reminiscent of the articulatory undershoot account; cf. Cho and Keating, 2001). However, it remains as yet unclear how this theory can explain vocalic expansion after the consonant; for this, further development of the model is needed.

As has been discussed so far, the π-gesture theory 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. However, further research is required to account for the gradient nature of both temporal and spatial variation, in a way that is both cognitively plausible and computationally implementable.

3.2.3 Bonding Strength

A possible alternative account that is available in a mass-spring dynamical system involves the notion of 'bonding strength', that is, the degree of cohesion among gestures (see Cho, 2001, based on the work of Browman and Goldstein, 1992). An important assumption of this account is that every gestural phase relation is associated with a differential degree of bonding strength. The original motivation for this was to take into account the variability in context-sensitive phasing relationships between consonantal gestures. Greater bonding strength is assumed to give rise to stronger gestural cohesion and more co-articulation between adjacent gestures. Given that this could be extended more generally to other phasing relationships, Cho et al. (2007) discussed the possibility that some types of boundary-induced phonetic variation can be explained on the assumption that bonding strength is inversely correlated with prosodic boundary strength (or the size of the prosodic domain). For example, cross-boundary V-to-V coarticulatory resistance (Cho, 2004) could be due to reduced bonding strength between vocalic gestures across a prosodic boundary. Similarly, reduced bonding strength would result in less coarticulation between a domain-initial consonant and its neighbouring gestures, leading to a greater duration of the former. If, as I propose, bonding strength is inversely co-indexed with boundary strength, then this information could be fed to the phonetic implementation stage, where prosodically driven fine-tuning of articulation may occur.

3.2.4 The Window Model

Prosodically conditioned phonetic variation can also be understood through Keating's (1990) window model (see also Byrd, 1996; Cho, 2004; Cohn, 1990; Docherty, 1992; Keating, 1996; see, for example, Guenther, 1995; Guenther et al., 1998, for an independent development of a window model). In the window model, articulatory movement targets are assumed to vary within specified ranges (i.e. 'windows') as opposed to fixed values. The model has been proposed to capture boundary-induced phonetic variation in such a way that the range of articulations adjacent to a prosodic boundary can be expanded or shrunk (Keating and Shattuck-Hufnagel, 2002). An alternative formulation, however, was proposed by Cho (2004): the possible range of variation of a given articulation remains fixed, but prosodic factors such as boundaries could specify an operating target region within the fixed window. For example, the target for an articulator could be modulated according to the prosodic structure generated on line, such that the same gesture is specified for a more extreme and narrower operating range in domain-initial than in domain-medial articulation. Cho (2004) suggested that such modulation of the window range can also be applied to prominence-induced articulatory variation. The Window Model thus captures prosodically driven phonetic variation in a linguistically intuitive way: the phonological category of a segment is expressed by an invariable window size, but the prosodically driven variability is allowed within that window. It still remains to be seen, however, how this interpretation of windows can be further developed in a computationally implementable way.

3.2.5 An Exemplar-Based Model

A relatively recent type of production model is the exemplar-based approach (e.g. Pierrehumbert, 2001, 2002, 2003). The main tenet of the model is that speech perception involves the storage of exemplars of specific speech events in a multidimensional phonetic space, and that emerging phonetic categories form 'clouds' in different regions of this space. Each member of a cloud is a remembered instance of a given category (or label), such that each category is associated with a frequency-weighted distribution of phonetic events. Crucially, perceptually cumulated exemplars are used in speech production. When a phonetic category label is chosen, the motor commands for that label are executed based on a random sampling from the distribution of exemplars associated with that label. As Cho et al. (2007) suggest, the exemplar clouds could be co-indexed by prosodic domain. For example, the cloud for a consonant could be subdivided into clouds according to the level of prosodic domain (e.g. Intonational Phrase, Intermediate Phrase, Prosodic Word), so that once the category label and its prosodic position have been selected in production, a candidate exemplar could be selected at random from the prosodically appropriate sub-distribution of exemplars associated with that label (though Cho et al., also discuss a number of potential drawbacks of an exemplar-based approach).

3.3 The Prosody Generator in Speech Production

The mechanisms discussed in the previous section have some implications for psycholinguistics theory of speech production. In the theory of production developed by Levelt and colleagues (Levelt, 1989; Levelt et al., 1999), the influence of the prosodic structure of an utterance on articulatory planning occurs after lexical access, operated by the 'Prosody Generator' device. (See Keating and Shattuck-Hufnagel, 2002, for further discussion on where in the speech planning process the prosodic structure of an utterance should be built.) This device receives abstract phonological input about words along with information about their prosodic structure, and adjusts phonological specifications as a function of the context – a process called 'prosodification', which occurs before the phonetic encoding stage. (Phonetic encoding involves the specification of phonological representations on the phonetic surface, while phonological encoding refers to the phonological specification of lexical representations or 'lemmas'.) The Prosody Generator determines, for example, resyllabification and stress shift. However, in the current model it does not take into account fine-grained phonetic details that correlate with prosodic structure, such as domain-initial strengthening. The output of prosodification is fed into the phonetic encoding stage, where 'preprogrammed' gestural scores (in the sense of Articulatory Phonology) are retrieved from the 'mental syllabary'. Crucially, the model assumes that the retrieval of the gestural scores is the final step in speech planning, after which motor execution takes place as an automatic and biomechanic phonetic process. What has not been considered in the model is how fine-tuning of these 'preprogrammed' gestural scores occurs as a function of prosodic structure. Cho et al. (2007) proposed that such fine-tuning could still be done by the Prosody Generator, but that it must occur at the phonetic encoding stage, given that it must reflect prosodically driven subphonemic variation of speech. (See also Keating, 2006, for discussion of this issue.) To do this, the Prosody Generator could be harnessed with some of the mechanisms discussed in the previous section. Given that the notion of gestural scores adopted by the model is based on a mass-spring gestural dynamical system, it may be theoretically coherent for the Prosody Generator to be further developed by incorporating the mechanisms of the π-gesture model.

4. Issues in Prosodic Strengthening in Speech Perception

As we now understand it, prosodic strengthening is not simply a physical articulatory event beyond the domain of linguistic control, but is realized in a linguistically meaningful way, conveying information about boundaries and prominence. A question that naturally arises, then, is in what ways such structural information is useful in speech comprehension. In this section I will discuss some possible roles of prosody in speech comprehension, especially in terms of lexical processing, and propose that prosodic information is processed in parallel with segmental information, by a special device called the 'Prosody Analyser'.

4.1 Phonetic Manifestations of Prosodic Structure in Speech Comprehension

Some of the earlier studies on roles of prosody in spoken word recognition have focused on how word-level prosodic information about lexical stress is exploited in lexical segmentation. For example, Cutler and her colleagues (Cutler and Norris, 1988; Cutler and Butterfield, 1992) showed that English listeners tend to use the frequent strong-weak (trochaic) stress pattern to detect the beginning of a word, which of course is cued by suprasegmental features such as F0, duration and amplitude (Lehiste, 1970). Fragment priming experiments in Spanish (Soto et al., 2001) also suggested that when the stress of the spoken fragment is matched with the stress of the visual target, recognition of the target word is facilitated, but inhibition occurs when stress is mismatched. Other studies have examined effects of cues to prosodic word boundaries in lexical access (e.g. Gow and Gordon, 1995; Davis et al., 2002; Salverda et al., 2003). One particular line of enquiry has examined the case of shorter words embedded in longer words (e.g. *ham* embedded in *hamster*). In eye-tracking experiments with cross-spliced Dutch materials, for example, where participants were instructed to click on the object picture (e.g. *ham* or *hamster*) mentioned in the auditory sentence, Salverda et al. showed more transitory fixations to pictures of monosyllabic words (e.g. *ham*) when the first syllable of the target word (e.g. *hamster*) was cross-spliced from the original monosyllabic word (*ham*) than when it was from the longer word (*hamster*). They then argued that the acoustic durational cues were a consequence of a difference in prosodic structure, and that they modulated lexical activation.

The studies discussed above demonstrated word-level prosodic effects on spoken word recognition. Our understanding of how phrase-level prosodic information is used in spoken word recognition is still limited. However, more recently research in lexical processing has begun to investigate possible roles of prosodic information in higher-level prosodic structure (e.g. Christophe et al., 2004; Kim, 2004; Cho et al., 2007; Shukla et al., 2007; Welby, 2007; Kim and Cho, 2009; Warner et al., in press). For instance, in word monitoring and phoneme detection tasks in French, Christophe et al. showed that lexical access for monosyllabic words (e.g. *chat* 'cat') in two-word sequences (e.g. *chat grincheux* 'cat grumpy') was faster when Phonological Phrase (an intermediate level in the prosodic hierarchy) boundaries intervene than when the sequences occur within such a phrase. They proposed that listeners terminate pending lexical searches when they encounter Phonological Phrase boundaries, and therefore that lexical access operates within the domain of the Phonological Phrase. However, Christophe et al.'s study did not examine exactly what kind of phonetic information about prosodic structure was used in lexical processing, though they speculated that phrase-final lengthening would be a major cue.

In an effort to explore how specific prosodic cues to high-level prosodic structure may be used, Kim (2004) employed an artificial language learning paradigm (for this, see Saffran et al. 1996). Kim created a small artificial lexicon and tested how novel words were learned by Korean listeners. In the learning phase, listeners were exposed to continuous sequences of the novel words for a certain period of time, and in the test phase they had to identify whether a given string of sounds is a likely word that might occur in the artificial language that they had been exposed to. Crucially, in creating spoken novel words, various prosodic cues were manipulated, such as duration (word-finally, in line with phrase-final lengthening), amplitude (word-initially, in line with domain-initial strengthening), and F0 (rising word-finally, in line with the Korean intonational structure, where an Accentual Phrase ends with a rising tone). When novel words contained these cues in conformity with the specific prosodic characteristics of Korean, learning the novel words was improved as compared to the control case, where no prosodic cues were available. In particular, the presence of AP-final durational cue was found to improve learning the most. On the other hand, when a rising tone, an AP-final intonational marker in Korean, was aligned with word-initial position, learning performance became poorer than in the control case. Kim's study therefore suggested that each of prosodic cues to high-level prosodic structure contributes to lexical segmentation, but only if the cue conforms to the language-specific prosodic structure (cf. See Tyler and Cutler, 2009, for a cross-linguistic study where phrase-final lengthening served as a cross-linguistic cue).

Subsequently, in word-spotting experiments in Korean, Kim and Cho (2009) further tested how such multiple prosodic cues to high-level prosodic structure contribute cumulatively to lexical segmentation. They found that when both phrase-final duration and intonational cues are available, listeners did not necessarily benefit from them both. Instead, the results showed partial cumulative effects: the addition of phrase-final durational cues was most efficiently exploited when intonational information was not optimal (i.e. with infrequent intonational patterns that do not conform to the intonational phonology of Korean; cf. Jun, 1993, 1995, 2000). Taken together, these studies therefore demonstrate that suprasegmental cues which mark prosodic boundaries are indeed exploited by listeners in lexical segmentation, but the way they are used is constrained by language-specific prosodic structure.

Let us now consider some possible roles of domain-initial strengthening whose characteristics are not entirely suprasegmental, that is, those which are accompanied by extreme articulation. Domain-initial position has often been considered to be an informationally rich locus in speech processing (see Gow et al., 1996). A basic assumption has been that the speaker signals prosodic structure via articulatory domain-initial strengthening, and that the listener makes use of its acoustic consequences in comprehension (cf. Cho et al., 2007). As Keating (2006) observes, domain-initial strengthening is presumably motivated by the fact that initial segments are less determined by prior context, and the contextual information gap is compensated for by domain-initial strengthening. Fougeron and Keating (1997) also speculated that since domain-initial strengthening entails increased articulatory contrast between segments straddling a prosodic boundary, this contrast could contribute to marking that boundary, thereby helping listeners to parse the continuous incoming speech signal into words and thus facilitating lexical segmentation.

Cho et al. (2007) moved beyond speculations, and directly tested the role of domain-initial strengthening in lexical segmentation. In cross-modal identity-priming experiments, listeners made lexical decisions to visual targets (e.g. *mill*) as they heard the sentence with two-word sequences containing lexical ambiguities (e.g. *mill company*, with the competitor *milk*). The two-word sequences (e.g. *mill # company*) contained Intonational Phrase or Prosodic Word boundaries (IP-boundary context, for example, *To learn about wood products, they visited a MILL COMPANY in Alabama last summer* versus Word-boundary context, for example, *When I was thinking about buying a coffee MILL, COMPANY names were the most important things I considered*). The second word's onset (e.g. [kɑ] in *company*) was then spliced from another token of the sequence in IP-initial position (strengthening condition) or Wd-initial position (non-strengthening condition). When related targets were identical to pre-boundary words (e.g. *mill*) in Word-boundary context, a stronger priming effect was observed in the strengthening condition (when the post-boundary onsets were spliced from IP-initial than from Wd-initial position). The strengthened [k] was viewed to provide a better match to the onset of *company* than the coda of *milk*, its competitor in the sequence *mill company*. Domain-initial strengthening present in the onset of the following word was, therefore, interpreted as serving as a cue to lexical segmentation via resolving lexical ambiguity. A general conclusion is that phonetic correlates of domain-initial strengthening are used as acoustic cues in the segmentation of continuous speech, and, more broadly, that speakers signal prosodic structure in systematic, fine-grained phonetic detail and that listeners make use of it in speech comprehension.

4.2 The Prosody Analyser in Speech Comprehension

Successful spoken word recognition necessitates successful lexical segmentation of continuous speech input that is, finding the word boundaries as intended by the speaker. A well-known approach to lexical segmentation is to consider the process as a consequence of lexical competition (e.g. Marslen-Wilson and Welsh ,1978; McClelland and Elman, 1986; Norris, 1994). In lexical competition, a set of competitors whose acoustic beginnings are consistent with the speech input is initially activated. Competitors are then inhibited as they become mismatched with the input as it unfolds over time. Eventually, a single candidate remains as the winner in the competition. This also ends the search for the word boundary, and hence lexical segmentation. The fact that fine-grained phonetic (subphonemic) details influence lexical segmentation (including prosodically driven phonetic variation) has challenged traditional models of lexical competition which rely on phonemic representations only (cf. Gow and Gordon, 1995; Norris et al., 1997; Gow, 2002).

To account for how the fine-grained phonetic information about prosodic structure influences lexical competition, Cho et al. (2007) have proposed the 'Prosody Analyser' account (see also Salverda et al., 2003), as a mirror image of the Prosody Generator account in speech production (see Section 3.3). Their hypothesis was that the Prosody Analyser uses information which specifies suprasegmental (and possibly segmental) aspects of the speech signal to compute the prosodic structure of the current utterance. Crucially, the Prosody Analyser is not completely separated from the process of segmental analysis (which is necessary to retrieve phonemic representations), but representations of prosodic structure are taken to be extracted in parallel to segmental representations. On the one hand, the segmental analysis determines what words are considered (i.e. the content) in the current input, possibly in terms of phonemic representations in the Shortlist model (Norris, 1994) and the TRACE model (McClelland and Elman, 1986), or possibly in terms of fine-grained phonetic details as in the Feature Parsing model (Gow, 2002). On the other hand, the Prosody Analyser guides the process which determines where words are likely to begin and end, for which it uses the location information of boundaries to modulate the lexical competition process. Potential lexical boundaries, computed by their segmental match with the speech input, are then checked against prosodic boundaries computed by the Prosody Analyser, so that the alignment between the two hypothetical boundaries plays an important role in determining the 'winner' in the lexical competition. The Prosody Analyser account is therefore in line with the assumption that segmental and prosodic information have different roles to play in the word recognition process. See Cho et al. (2007) for further discussion on how an episodic account of word recognition (e.g. Goldinger, 1996, 1998; Johnson, 1997) could make use of fine-grained phonetic details in lexical processing and on possible advantages of the Prosody Analyser account over the episodic account.

5. A New Challenge

This chapter has considered the phonetics-prosody interface in the realm of Laboratory Phonology, focusing on how prosodic structure is manifested in fine-grained phonetic details in speech production and what roles they may play in speech comprehension. We have seen ample evidence that prosodically driven fine-grained phonetic details do not arise simply as low-level physical phenomena, but instead play linguistically significant roles, serving dual functions of boundary marking and prominence marking. We have also learned that phonetic manifestations of prosodic structure are closely linked to contrast enhancement, both paradigmatically (i.e. phonemic) and syntagmatically (i.e. structurally), and that the way linguistic contrasts are enhanced is modulated by language-specific phonological systems. Moreover, we have seen that phonetic consequences of prosodic strengthening are indeed used in speech comprehension, especially in terms of lexical segmentation and resolution of temporal lexical ambiguities. It was proposed that prosodic structure should be computed (e.g. by the Prosody Analyser) in parallel with segmental analysis, and that the match between computations of prosodic structure and lexical content facilitates lexical processing.

Over the past two decades, we have clearly obtained better insights into the nature of sound systems in which phonetics and prosody are intertwined in a linguistically systematic way. Understanding the interplay between phonetics and prosody must now be seen as an essential prerequisite for understanding the whole linguistic communication system of a given language. We are now left with a new challenge: how do we reinforce our knowledge and extend it to build models that capture the complex interplay of phonetics and prosody, in a way that is both descriptively and explanatorily adequate?

In building such a model, possible mechanisms underlying phonetic manifestations of prosodic structure may include the π-gesture model, Bonding Strength, the Window Model and Exemplar-based approaches, which can be combined with a psycholinguistic production model such as Levelt's, about which I proposed that prosodically driven fine-tuning of articulation must operate at the level of phonetic encoding, possibly by the Prosody Generator. The challenge that we are facing is how to put these insights together and integrate them into a more general model of speech production. As Beckman (1996) puts it, prosody is now considered to be a grammatical entity in its own right. In building such an integrated model, the most important issue will be where in the general architecture of grammar prosody needs to be placed and how the prosody component interacts with other components in the grammar. All in all, since phonetic manifestations of prosodic structure convey information about various strata of speech production (from low-level phonetic realizations to the discourse structure), and since prosodic information is indeed exploited in speech comprehension, the Laboratory Phonology community needs to develop a more full-fledged model. This model should provide an adequate explanation of the interactions between prosody and other linguistic components in order to achieve a better understanding of global linguistic functions of speech prosody in the architecture of grammar.

References

Barnes, J. A. (2001). Domain-initial strengthening and the phonetics and phonology of positional neutralization. Proceedings of the Northeast Linguistic Society (NELS), 32, vol. 1, pp.1-20.

Barnes, J. A. (2002). Positional Neutralization: A Phonologization Approach to Typological Patterns. Unpublished Ph.D. dissertation. University of California, Berkeley.

Beckman, J.N. (1998). Positional Faithfulness. Ph.D. dissertation, University of Massachusetts, Amherst.

Beckman, M. E. (1986). Stress and Non-stress Accent. Dordrecht: Foris Publications.

Beckman, M. E. (1996). The parsing of prosody. Language and Cognitive Processes, 11, 17-67.

Beckman, M.E., Edwards, J. and 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. .

Beckman, M. E. and Edwards, J. (1994). Articulatory evidence for differentiating stress categories. In Papers in Laboratory Phonology III: Phonological Structure and Phonetic Form (P. A. Keating, editor), pp. 7-33. Cambridge: Cambridge University Press.

Beckman, M. E. and Kingston, J. (1990). Introduction. In J. Kingston and M. Beckman (Eds.), Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech, pp. 1-16. Cambridge: Cambridge University Press.

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

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

Boersma, P. (1998). Functional Phonology. LOT International Series 11. The Hauge: Holland Academic Graphics.

Browman, C. and 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. and Goldstein, L. (1992). Articulatory phonology: An overview. Phonetica, 49, 155-180.

Bruce, G. (1977). Swedish Word Accents in Sentence Perspective. Travaux de L'institut de Linguistique de Lund XII. Gleerup, Lund.

Byrd, D. (1996). A phase window framework for articulatory timing. Phonology, 13, 139-169.

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. and Saltzman, E. (1998). Intragestural dynamics of multiple prosodic boundaries. Journal of Phoentics, 26, 173-199.

Byrd, D., Kaun, A., Narayanan, S. and 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, and 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. and Saltzman, E. (2003). The elastic phrase: modeling the dynamics of boundary-adjacent lengthening. Journal of Phonetics, 31, 149-180.

Carlson, K., Clifton, C., Jr., and Frazier, L. (2001). Prosodic boundaries in adjunct attachment. Journal of Memory & Language, 45, 58-81.

Cambier-Langeveld, T. and Turk, A. (1999). A cross-linguistic study of accentual lengthening: Dutch vs. English. Journal of Phonetics, 27, 255-280.

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

Cho, T. (2001) Effects of morpheme boundaries on intergestural timing: evidence from Korean. Phonetica, 58, 129-162.

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 Laboratory phonology 8: Varieties of phonological competence (L. M. Goldstein, D. H. Whalen, C. T. Best, editors), pp.519-548. Berlin/New York: Mouton de Gruyter.

Cho, T. (2008). Prosodic strengthening in transboundary V-to-V lingual movement in American English. Phonetica, 65, 45-61.

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

Cho, T., Jun, S.-A. and Ladefoged, P. (2002). Acoustic and aerodynamic correlates of Korean stops and fricatives. Journal of Phonetics, 30, 193-228.

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

Cho, T. and Keating, P. (2009). Effects of initial position versus prominence in English. Journal of Phonetics, 37, 466-485.

Cho, T. and Ladefoged, L. (1999). Variation and universals in VOT: evidence from 18 languages. Journal of Phonetics, 27, 207-229.

Cho, T. and 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. and Cox, E. (2007). Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English. Journal of Phonetics, 35, 210-243.

Chomsky, N. & Halle, M. (1968). The Sound Pattern of English. New York: Harper and Row. (Reprinted, 1991, Boston: MIT Press.)

Christophe, A., Peperkamp, S., Pallier, C., Block, E., and Mehler, J. (2004). Phonological phrase boundaries constrain lexical access: I. Adult data. Journal of Memory and Language, 51, 523-547.

Cohn, A. (1990). Phonetics and Phonological Rules of Nasalization. Ph.D. disseration. UCLA (Also published as UCLA Working Papers in Phonetics, No. 76.)

Cohn, A. (1998). The phonetics-phonology interface revisited: Where's phonetics? Texas Linguistics Society 1998 Conference Proceedings, pp.25-40.

Cohn, A. (2006) Is there gradient phonology? In G. Fanselow, C. Fery, R. Vogel and M. Schlesewsky (Eds.), Gradience in Grammar: Generative Perspectives, pp.25-44. Oxford: Oxford University Press.

Cohn, A. (2007). Phonetics in phonology and phonology in phonetics. Working Papers of the Cornell Phonetics Laboratory, 16, 1-13.

Cohn, A. (in press). Laboratory phonology: Past successes and current questions, challenges, and goals. In Laboratory Phonology 10, Berlin: Mouton de Gruyter.

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.

Cooper, A. (1991). Glottal Gestures and Aspiration in English. Ph.D. Dissertation, Yale University.

Couper-Kuhlen, E. and Selting, M. (eds.) (1996). Prosody in Conversation: Interactional Studies. Cambridge: Cambridge University Press.

Cutler, A., and Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. Journal of Memory and Language, 31, 218-236.

Cutler, A., and Norris, D. (1988). The role of strong syllables in segmentation for lexical access. Journal of Experimental Psychology: Human, 14, 113-121.

Davis, M. H., Marslen-Wilson, W. D., and Gaskell, M. G. (2002). Leading up the lexical garden-path: Segmentation and ambiguity in spoken word recognition. Journal of Experimental Psychology: Human Perception and Performance, 28, 218–244.

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

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.

Delattre, P. (1940). La force d'articulation consonantique en français. The French Review, 14, 220-232.

Docherty, G. J. (1992). The timing of British English obstruents. Berlin: Foris.

Edwards, J. E. and Beckman, M. E. (1998). Articulatory timing and the prosodic intepretation of syllable duration. Phonetica, 45, 156-174.

Edwards, J. E., Beckman, M. E., and 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.

Flemming, E. (1995). Auditory Representations in Optimality Theory. Ph.D. disseration. UCLA. (Published by Routledge, 2002).

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. and Keating, P. A. (1997). Articulatory strengthening at edges of prosodic domains. Journal of the Acoustical Society of America, 106, 3728-3740.

Fourakis, M. and Port, R. (1986). Stop epenthesis in English. Journal of Phonetics, 14, 197–221.

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

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. Journal of Experimental Psychology: Learning, Memory, and Cognition, 22, 1166–1183.

Goldinger, S. D. (1998). Echoes of echoes?: An episodic theory of lexical access. Psychological Review, 105, 251–279.

Goldstein, L., Byrd, D., & Saltzman, E. 2006. The role of vocal tract gestural action units in understanding the evolution of phonology. In M. Arbib, (Ed.). Action to Language via the Mirror Neuron System. New York: Cambridge University Press, Pp. 215-249

Goldstein, L, Pouplier, M., Chen, L., Saltzman, E., and Byrd, D. (2007). Dynamic action units slip in speech production errors. Cognition, 103, 386-412

Goldstein, L., Chitoran, I, & Selkirk, E. (2007). Syllable structure as coupled oscillator modes: Evidence from Georgian and Tashlhiyt Berber. Trouvain, W. and Barry, W. Proceedings of the XVI International Congress of Phonetic Sciences, pp. 241-244.

Gow, D. W. (2002). Does English coronal assimilation create lexical ambiguity? Journal of Experimental Psycholology: Human Perception and Performance, 28, 163-179.

Gow, D. W., and Gordon, P. (1995). Lexical and prelexical influences on word segmentation: evidence from priming. Journal of Experimental Psycholology: Human Perception and Performance, 21, 344-359.

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

Guenther, F. H. (1995). Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. Psychological Review, 102, 594–621.

Guenther, F. H., Hampson, M., and Johnson, D. (1998). A theoretical investigation of reference frames for the planning of speech movements. Psychological Review, 105, 611–633.

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

Harrington, J., Fletcher, J. and 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., and Keating, P. (1999). Domain-initial strengthening in Taiwanese: A follow-up study. UCLA Working Papers in Phonetics, 97, 152-156.

Hayes, B. (1989). The prosodic hierarchy in meter. In Phonetics and Phonology, Vol.I: Rhythm and Meter (P. Kiparsky and G. Youmans, editors), pp. 47-96. New York: Academic Press.

Hayes, B., Kirchner, R. and Steriade, D. (2004). Phonetically Based Phonology. Cambridge: Cambridge University Press.

Hume, E. and Johnson (eds.) (2001). The Role of Speech Perception in Phonology. San Diego: Academic Press.

Hsu, C.-S. and Jun, S.-A. (1998). Prosodic strengthening in Taiwanese: Syntagmatic or paradigmatic? UCLA Working Papers in Phonetics, 96, 69-89.

Johnson, K. (1997). Speech perception without speaker normaliztion. In K. Johnson & J.W. Mullennix (Eds.), Talker Variability in Speech Processing, pp.145-166. San Diego: Academic Press.

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

Jun, S.-A. (1995). Asymmetrical prosodic effects on the laryngeal gesture in Korean, in Papers in Laboratory Phonology IV: Phonology and phonetic evidence, edited by B. Connell and A. Arvaniti (Cambridge University Press, Cambridge), pp. 235-253.

Jun, S.-A. (1998). The Accentual Phrase in the Korean prosodic hierarchy. Phonology, 15(2), 189-226.

Jun, S.-A. (2000). K-ToBI (Korean ToBI) Labelling Conventions (Version 3.1). (http://www.linguistics.ucla.edu/people/jun/ktobi/K-tobi.html).

Jun, S.-A., Beckman, M. and Lee, H.-J. (1998). Fiberscopic evidence for the influence on vowel devoicing of the glottal configurations for Korean obstruents. UCLA Working Papers in Phonetics, 96, 43-68.

Keating, P. A. (1984). Phonetic and phonological representation of stop consonant voicing. Language, 60,286–319.

Keating, P. A. (1985). Universal phonetics and the organization of grammars. In V.A. Fromkin (ed.), Phonetic Linguistics: Essays in Honor of Peter Ladefoged, pp. 115–132. Orlando FL: Academic Press.

Keating, P. A. (1990). The window model of coarticulation: Articulatory evidence. In J. Kingston, and M. Beckman (Eds.), Papers in laboratory phonology I: Between the grammar and the physics of speech (pp. 451–470). Cambridge: Cambridge University Press.

Keating, P. A. (1996). The phonology–phonetics interface. In U. Kleinhenz (Ed.), Interfaces in phonology (pp. 262-278). Berlin: Studia grammatica 41, Akademie Verlag.

Keating, P. A. and Shattuck-Hufnagel, S. (2002). A prosodic view of word form encoding for speech production. UCLA Working Papers in Phonetics, 101, 112-156.

Keating, P. A. (2006). Phonetic encoding of prosodic structure. In J. Harrington and M. Tabain (Eds), Speech Production: Models, Phonetic Processes, and Techniques (pp.167-186). New York and Hove: Psychology Press.

Keating, P. A., Cho, T., Fougeron, C. and Hsu, C. (2003). Domain-initial strengthening in four languages. In Papers in laboratory phonology 6: Phonetic interpretations (J. Local, R. Ogden, and R. Temple, editors), 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. (2004). The Role of Prosodic Phrasing in Korean Word Segmentation. Ph.D. disseration. UCLA.

Kim, S. and Cho, T. (2009). The use of phrase-level prosodic information in lexical segmentaiton: Evidence from word-spotting experiments in Korean. Journal of the Acoustical Society of America, 125(5), 3373-3386.

Kingston, J. and Diehl, R.L. (1994). Phonetic knowledge. Langauge, 70, 419-454.

Kirchner, R. (1998). An effort-based approach to consonant lenition. Ph.D. dissertation, UCLA. (Published by Routledge, 2001).

Krivokapić, J. (2007). Prosodic planning: Effects of phrasal length and complexity on pause duration. Journal of Phonetics, 35, 162-179.

Kuzla, C., Cho, T. and Ernestus, M. (2007). Prosodic strengthening of German fricatives in duration and assimilatory devoicing. Journal of Phonetics 35(3), 301-320

Ladd, D. R. (1996). Intonational Phonology. Cambridge, UK: Cambridge University Press.

Ladefoged, P. (1967). Three areas of experimental phonetics. London: Oxford University Press.

Ladefoged, P. and Loeb, G. (2002). Preliminary studies on respiratory activity in speech. UCLA Working Papers in Phonetics, 101. 50-60.

Lavoie, L. (2001). Consonant Strength: Phonological Patterns and Phonetic Manifestations. New York: Garland Publications.

Lee, S., Byrd, D. and Krivokapić, J. (2006). Functional data analysis of prosodic effects on articulatory timing. Journal of the Acoustical Society of America, 119, 1666-1671.

Levelt, W. J. M. (1989). Speaking: From intention to articulation. Cambridge: MITPress.

Levelt, W. J. M., Roelofs, A., and Meyer, A. S. (1999). A theory of lexical access in speech production. Brain and Behavioral Sciences, 22,1-38

Lehiste, I. (1970). Suprasegmentals (MIT Press, Cambridge, MA).

Liberman, M. and Prince, A. (1977). On stress and linguistic rhythm. Linguistic Inquiry, 8, 29-336.

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.

Lombardi, L. (1991a). Laryngeal features and privativity. Unpublished manuscript, University of Massachusetts, Amherst.

Lombardi, L. (1991b). Laryngeal Features and Laryngeal Neutralization. Ph.D. dissertation. University of Massachusetts, Amherst.

Marslen-Wilson, W., and Welsh, A. (1978). Processing interactions during word-recognition in continuous speech. Cognitive Psychology, 10, 29–63.

McClelland, J. L., and Elman, J. L. (1986). The TRACE model of speech perception. Cognitive Psychology, 18, 1-86.

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

Nam, H. (2007) A competitive, coupled oscillator model of moraic structure: Split-gesture dynamics focusing on positional asymmetry. In J. Cole & J. Hualde (Eds.), Laboratory Phonology 9, pp. 483-506. Berlin and New York: Moiuton de Gruyter.

Nespor, M., and Vogel, I. (1986). Prosodic Phonology (Dordrecht: Foris Publications).

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. Cognition, 52, 189-234.

Norris, D., and McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. Psychological Review, 115 (2), 357-395.

Norris, D., McQueen, J. M., Cutler, A., and Butterfield, S. (1997). The possible-word constraint in the segmentation of continuous speech. Cognitive Psychology, 34, 191-243.

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.

Onaka, A., Watson, C., Palethorpe, S. and Harrington, J. (2003). An acoustic analysis of domain-initial strengthening effect in Japanese. In Proceedings of the 6th International Seminar on Speech Production (S. Palethorpe and M. Tabain, editors), pp. 201-206. Sydney.

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

Pierrehumbert, J. & Beckman, M. (1988). Japanese Tone Structure. Cambridge, MA: MIT Press.

Pierrehumbert, J. (2001). Exemplar dynamics: Word frequency, lenition, and contrast. In J. Bybee, and P. Hopper (Eds.), Frequency effects and the emergence of linguistic structure (pp. 137–157). Amsterdam: John Benjamins.

Pierrehumbert, J. (2002). Word-specific phonetics. In C. Gussenhoven, and N. Warner (Eds.), Laboratory phonology VII (pp. 101–140). Berlin: Mouton de Gruyter

Pierrehumbert, J. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. Language and Speech, 46(2–3), 115–154.

Pierrehumbert, J. and 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.

Pierrehumbert, J. and Talkin, D. (1992). Lenition of /h/ and glottal stop. In Papers in Laboratory Phonology II: Gesture, Segment, Prosody (G. Docherty and D. R. Ladd, editors), pp. 90-117. Cambridge, UK: Cambridge University Press.

Saffran, J. R., Newport, E. L., and Aslin, R. N. (1996). Word segmentation: The role of distributional cues. Journal of Memory and Language, 35 (4), 606-621.

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. and Munhall, K.G. (1989). A dynamical approach to gestural patterning in speech production. Ecological Psychology, 1, 333-82.

Saltzman, E., Nam, H., Krivokapic, J., & Goldstein, L. (2008). A task-dynamic toolkit for modeling the effects of prosodic structure on articulation. In P. A. Barbosa, S. Madureira, & C. Reis, (Eds.), Proceedings of the Speech Prosody 2008 Conference, Campinas, Brazil.

Salverda, A.P., Dahan, D., and McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. Cognition, 90, 51-89.

Selkirk, E. (1984). Phonology and syntax: The relation between sound and structure (MIT Press, Cambridge).

Selkirk, E. (1995). Sentence Prosody: Intonation, Stress, and Phrasing. In The Handbook of Phonological Theory (J. A. Goldsmith, editor), pp.550-569. Cambridge, Mass & Oxford: Blackwell Publishers.

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

Shukla, M., Nespor, M. and Mehler, J. (2007). An interaction between prosody and statistics in the segmenation of fluent speech. Cognitive Psychology, 54, 1-32.

Soto, S., Sebastián-Gallés, N., and Cutler, A. (2001). Segmental and suprasegmental mismatch in lexical access. Journal of Memory and Language, 45, 412-432.

Steriade, D. 1993. Positional neutralization. A talk presented at NELS 23. University of Massachusetts, Amherst.

Steriade, D. 1999. Phonetics in phonology: The case of laryngeal neutralization. UCLA Working Papers in Linguistics, No. 2, Papers in Phonology 3, edited by M. Mordon, pp.25-146.

Straka, G. (1963). La divison de sons du language en voyelles et consonnes peut-elle être justifiée? Travaux de Ling. et de Littérature, U. de Strasbourg, 1, 17-99.

Tabain, M. and Perrier, P. (2005). Articulation and acoustics of /i/ in preboundary position in French. Journal of Phoentics, 33, 77-100.

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

Turk, A. E. and Shattuck-Hufnagel, S. (2000). Word-boundary-related durational patterns in English. Journal of Phonetics, 28, 397-440.

Tyler, M.D. and Culter, A. (2009). Cross-language differences in cue use for segmentation. Journal of the Acoustical Society of America, 126, 367-376.

Warner, N., Otake, T. and Takayuki, A. (to appear). Intonational Structure as a Word Boundary Cue in Japanese. Language and Speech.

Welby, P. (2007). The role of early fundamental frequency rises and elbows in French word segmentation. Speech Commununication, 49, 28-48.

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

Zsiga, E. C. (1995). An acoustic and electropalatographic study of lexical and postlexical palatalization in American English. In Connell, B., & A. Arvaniti (Eds.), Papers in Laboratory Phonology IV: Phonology and Phonetic Evidence, pp. 282–302. Cambridge: Cambridge University Press.