Journal of Phonetics 44 (2014) 1–7

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/phonetics

DYNAMICS of ARTICULATION

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

and PROSODIC STRUCTURE

More than a magic moment – Paving the way for dynamics of articulation and prosodic structure[†]

Doris Mücke[a,*], Martine Grice[a], Taehong Cho[b]

[a] IfL Phonetics, University of Cologne, Herbert-Lewin-Strasse 6, 50931 Cologne, Germany
[b] Hanyang Phonetics & Psycholinguistics Lab, Department of English Language & Literature, Hanyang University, Seoul 133-791, Republic of Korea

ARTICLE INFO

Article history:
Received 21 February 2014
Received in revised form
28 February 2014
Accepted 4 March 2014
Available online 3 April 2014

Keywords:
Prosody
Articulation
Dynamical systems
Dynamics of speech

ABSTRACT

Research into human communication through the spoken language is full of dichotomies that have often stood in the way of progress in the past, notably the distinction between phonetics and phonology, and more recently, and somewhat orthogonally, between prosody and articulation. The papers collected here make considerable advances in overcoming these restrictions, providing valuable contributions towards the integration of these fields. The increasing evidence for dependencies across the different levels of linguistic structure, and the complexity of the interplay between them, has led to the application of dynamical approaches to spoken language description. With these approaches, coordination and variation within and across systems have begun to play a central role. This paper identifies a common thread through the papers in this issue, in which variation is a consequence of dynamically time-varying behavior that cannot be captured by static snapshots (magic moments).

1. Introduction

Research into human communication through the spoken language is full of dichotomies that have often stood in the way of progress in the past, notably the distinction between (1) *phonetics and phonology*, and more recently, and somewhat orthogonally, between (2) *prosody and articulation*. A promising way of framing their integration into a unified system is provided by (3) *dynamical systems* approaches. We deal with each of these three aspects in turn below.

1.1. Phonetics and phonology

As early as 1990 Ohala claimed that “[t]here is no interface between phonetics and phonology”, indicating that there was no need for an interface if there was no clear division between the two in the first place (Ohala, 1990). In recent decades much has been done to break down the artificial boundaries constructed to delineate the two fields. Even if we regard them as two poles on a continuum, any attempt to divide up this continuum leads to ambiguities in terminology (Pierrehumbert, 1990; Rischel, 1990). These ambiguities can be found, not only across different theories, but also within a single framework (Keating, 1990). The term *phonetics–phonology interface* is now often used in a broader sense, employed to characterize the *interplay* rather than interface as such (e.g., Cohn, 2007; Hayes, Kirchner, & Steriade, 2004; Hume & Johnson, 2001; Keating, 1996; Kingston, 2007). The discussion here centres around a broader, overarching question: to what extent does abstract phonological structure inform, or is informed by, detailed phonetic patterns. In other words, it is concerned with the question as to “how autonomous phonetics and phonology are from one another and whether one can be largely reduced to the other” (Kingston, 2007:401). What has now emerged is an integrated view of phonetics and phonology, where the interplay between the two is not merely a problem to be solved but their very essence.

* Corresponding author.
E-mail address: doris.muecke@uni-koeln.de (D. Mücke).
[†] All three authors contributed equally to this work as co-authors.

0095-4470/$ - see front matter
http://dx.doi.org/10.1016/j.wocn.2014.03.001

1.2. Prosody and articulation

In autosegmental metrical models of intonation, there is a fundamental split between what is referred to as the tune (tonal aspects at lexical and postlexical levels) and the text (segmental aspects). Furthermore, the text involves not only the segments in terms of time slots and features filling them, but also the way these segments are organized into prosodic structures, such as syllables, feet, words and higher phrases and, depending on the language, prominence relations (Beckman, 1996; Grice, 2006; Gussenhoven, 2004; Shattuck-Hufnagel & Turk, 1996). Prosodic structure is generally assumed to reflect both constituent-based and prominence-based prosodic hierarchies of an utterance on the one hand, and tonal structure on the other, determining the distribution of the tune over the text.

While the tune and text levels – or tiers on which they are separately represented – may be linked to each other through phonological association (Goldsmith, 1976; Ladd, 2008; Leben, 1973), the interplay between these tiers has come to light in recent years under the rubric of the *phonetics–prosody interface* (see Cho, 2011 for a review). This is concerned with how abstract prosodic structure influences the phonetic implementation of sound categories, and how fine-grained phonetic detail, in turn, informs higher-level prosodic structure (e.g., Beckman, 1996; Byrd & Choi, 2010; Cho, 2006; Cho, McQueen, & Cox, 2007; Keating, Cho, Fougeron, & Hsu 2003; Krivokapić & Byrd, 2012). One important aspect of the phonetics–prosody interface is domain initial strengthening: a given segment is produced with stronger articulation when at the initial edge of higher levels of prosodic structure than at the initial edge of lower ones (e.g., Cho & Keating, 2009; Fougeron & Keating, 1997). More recently, however, there has been a growing awareness that the different levels in the network of speech production interact in complex ways. The increasing evidence for dependencies across the different levels, and the complexity of the interplay between them, has led to the application of dynamical approaches to spoken language description. With these approaches, *coordination* and *variation* within and across systems have begun to play a central role.

1.3. Dynamical approaches

The lowest common denominator across different dynamical approaches is the notion that systems are not made up of static entities such as symbols and rules, but that their behavior is best studied in terms of change over time. This makes variation and context-dependency an important part of a description. Context-dependent variation in speech production often stems from interactions between various subcomponents of linguistic structure in the grammar. In this regard, a dynamical approach deals with the interdependency between levels of linguistic structure that gives rise to systematic variation in speech production. This perspective can be traced back to Firth's Prosodic Phonology (Firth, 1948), an approach that has informed various frameworks, in particular Articulatory Phonology and Autosegmental Phonology.

In a specific sense, *dynamical systems* describe the evolution of complex behavior of a system. Dynamical systems use fixed rules to describe the evolution of a system over time using the language of mathematics. These rules are formulated by differential equations that generate trajectories, which, in turn, can be understood as a way of modeling continuous human behavior and cognition (see Kelso, 1995; Spivey, 2007). For instance, we know from movement research that inter-limb coordination can be understood as a dynamical system, changing its state in time in a lawful manner. These rules are invariant mathematical laws. Take, for instance, the coordinated movements of the index fingers when oscillating rhythmically (finger wiggling). Subjects are able to wiggle their fingers in two intrinsic modes: either both fingers move up and down in synchrony (simultaneous activation of homologous muscle groups) or they alternate (homologous muscle groups contract in an alternating fashion; Kelso, 1981, 1984; Turvey, 1990).

In the framework of coupled oscillators, these modes involve synchronous (in-phase) or sequential (anti-phase) activation. When the cycling frequency is gradually increased, there is an abrupt coordinative shift at a critical frequency, a phase transition from a less stable, anti-phase mode to a more stable, in-phase mode. The fact that a gradual increase in cycling frequency determines an abrupt coordinative shift at a critical stage is evidence that we are dealing with a non-linear system, non-linearity being a common property of human behavior and cognition. Relative phase transitions in rhythmic bimanual or inter-limb coordination have been formalized in a model developed by Haken, Kelso, and Bunz (1985) known as the HKB-model.

Fig. 1(a–c) illustrates potential functions in the HKB-model, as described in Nam, Goldstein, and Saltzman (2010). The *x*-axes of these figures represent the collective variable "relative phase," where 0° corresponds to an in-phase coordinative pattern and ±180° to anti-phase coordinative patterns. The *y*-axis represents the "potential" associated with different values of relative phase. The figure employs the metaphor of a marble, where the motion of the marble describes the motion of a system in time. If a marble rolls into a smooth round bowl (the attractor basin), it eventually comes to rest at the bottom of the bowl (the attractor), where the "potential" or energy is at a minimum.

In Fig. 1(a) the larger (deeper and broader) basin is associated with the more stable in-phase attractor (potential minimum at 0°) and the two smaller basins with the less stable anti-phase attractors (potential minima at 180°). The system can stabilize either in an in-phase or an anti-phase mode.

> "There are two potential minima at 0 and 180 degrees and, depending on initial conditions, relative phasing can stabilize at either of the two minima, making them attractors. However, the valley associated with the in-phase minimum is both deeper and broader. Thus, it technically has a larger basin because there is a larger range of initial values for ψ that will eventually settle into that minimum."(Nam et al., 2010: 304)

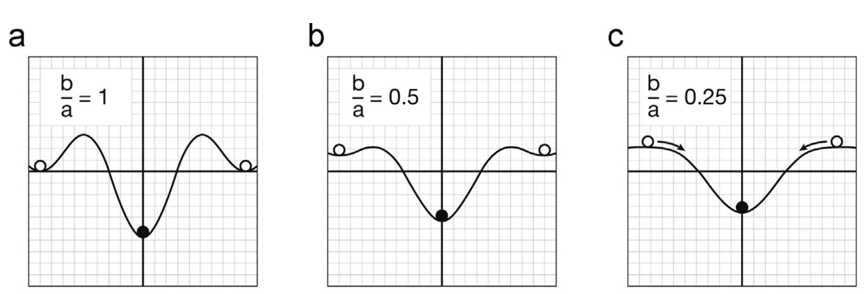

Fig. 1. Marble metaphor for the HKB-model, simplified according to Nam et al. (2010: 304) for $V(\psi) = -a\cos(\psi) - b\cos(2\psi)$; ($\psi = \phi_2 - \phi_1$).

The depth and the breadth of the valley are determined by the control parameter (b/a), which is the cycling frequency of the limb movements – determined externally by the experimenter (e.g., via a metronome). As the control parameter decreases, cycling frequency increases, and the anti-phase attractor becomes less stable, which is shown by the anti-phase attractor basin becoming more shallow (Fig. 1b). At a certain critical frequency of the control parameter, there is an abrupt change observable in the system from anti-phase to in-phase. To return to the metaphor, all marbles will eventually roll into the large bowl (as indicated by the arrows in Fig. 1c). At high cycling frequency, the attractor at 0° is, in the long run, the only stable state the system can be in – and, in line with this, people performing inter-limb coordination tasks will invariably switch to an in-phase coordination pattern.

Speech can also be understood as a non-linear dynamical system with invariant laws prescribing parameters related to the variation in coordination of the articulators (e.g. Task Dynamics and its implementation in Articulatory Phonology, Browman & Goldstein, 1986, 1988, 1991; Fowler, 1977; Fowler, Rubin, Remez, & Turvey, 1980; Saltzman, 1986; Saltzman & Kelso, 1987; Saltzman & Munhall, 1989, or Embodied Task Dynamics, Šimko & Cummins, 2010). These models capture both *contextual variation* and *task-specific invariance* (i.e. they generate continuous time-functions – or movement trajectories – for discrete, linguistic tasks). Crucially they use one formal language to do so:

> "(…) cognition is best understood using a single formal language that can express both discrete and continuous aspects of complex systems, the mathematics of nonlinear dynamics. In this view, the key constructs are not symbol strings (representations) and algorithms for their manipulation (discrete computation), but rather laws stated in the form of differential equations. These laws prescribe how some behavior's essential parameters (e.g., perceptual response or relative phase in interlimb coordination) change as contextual parameters are modified (e.g., stimulus properties, oscillation frequency)." Gafos and Beňuš (2006:906)

The fact that non-linear dynamical systems are able to integrate discrete, qualitative and continuous, quantitative aspects of speech means that they are optimally suited for accommodating the complexities of multifaceted interactions across different levels of linguistic description. This makes them optimal for capturing the integration of phonetics and phonology.

However, a challenge for such non-linear dynamical models is how to integrate prosody into the system. Although some efforts have been made to provide a mechanism in dynamical terms for how temporal variation is modulated at prosodic junctures (e.g. by defining prosodic gestures or π-gestures, Byrd & Saltzman, 2003, see also Katsika et al., this issue), prosody is far from being fully integrated into existing dynamical models. The coordination between different prosodic constituents is still uncharted, but treating the relation between them in terms of the dynamical notion of coordination is becoming increasingly important (Cummins & Port, 1998; Goldstein, Nam, Saltzman, & Chitoran, 2009; Krivokapić, 2013; Mücke, Nam, Hermes, & Goldstein, 2012; Nam & Saltzman, 2003; Nam, Saltzman, Krivokapić, & Goldstein, 2009; O'Dell & Nieminen, 1999; Saltzman & Byrd, 2000; Saltzman, Nam, Krivokapić, & Goldstein, 2008; Shaw, Gafos, Hoole, & Zeroual, 2011).

The title of this introduction "More than a magic moment" was inspired by the contribution of Vatikiotis-Bateson, Barbosa & Best in the present volume that emphasizes the fact that complex linguistic behavior is best understood dynamically, as reflected in the way it unfolds over time. The central point here is that variation in speech and gesture arises as an inevitable consequence of dynamically time-varying behavior. It is therefore an integral part of the system rather than noise. Indeed, it encodes information about the system itself.

> "The inevitable and even desirable presence of fluctuations has several important implications for research on spatiotemporal behavior. Importantly, it means that we cannot simply disregard measured variability as irrelevant noise, as has been done so often in psychological and linguistic research, because variance conflates notions of noise and error with mandatory, healthy fluctuations in patterned behavior. Implicitly, then, the behavior of the system must be examined dynamically as it unfolds through time – certainly, snap shot, magic moment measures will not suffice." (Vatikiotis-Bateson, Barbosa & Best, this issue)

This means that it is not the perfect snapshot or the best dataset, but the variation within and across datasets that is important for the description of speech. And, indeed, if speech is intrinsically dynamic, that is, characterized by changes over time, static snapshots might not suffice to adequately characterize the system. The present special issue takes up these points in a number of different domains, as summarized in the following section.

2. Structure of the special issue

The papers are organized into four sections: Section 1, Dynamics of articulation within the syllable; Section 2, Dynamics of articulation and prominence; Section 3, Dynamics of articulation and domain edges; and Section 4, Dynamical modeling of coordination.

2.1. Section 1: Dynamics of articulation within the syllable

All three papers in this section deal with the organization of consonantal and vocalic gestures and the relations between them, with a particular focus on subsyllabic constituents. Hoole and Bombien capture glottal abduction gestures associated with voiceless obstruents. They explore how these glottal gestures are coordinated with oral gestures in complex onsets containing obstruent and sonorant sequences in German. Marin and Pouplier on Romanian, and Tilsen on American English both investigate intergestural timing in the oral system. Marin and Pouplier show that segmental make-up in terms of clear and dark liquids systematically influences articulatory timing patterns in complex codas, but not in onsets. Tilsen uses a cued insertion task to explore articulatory planning, supporting a split-control model of intragestural and intergestural timing.

Philip Hoole and *Lasse Bombien* use videofiberendoscopy and transillumination to investigate the coordination between laryngeal and oral constriction gestures in mixed-voicing syllable onsets in German. As speech material, they used simple onsets containing voiceless obstruents, and complex onsets with voiceless obstruents followed by voiced sonorants, /l/ or /r/. Furthermore, all target words were tested in different prosodic conditions involving accent placement (contrastive nuclear accents vs. background) and boundary marking (phrase initial vs. medial) in terms of domain initial strengthening. On the acoustic surface, voice onset time was longer in complex onsets compared to simple onsets. In the articulatory dimension, there were strong speaker specific strategies to account for the longer VOTs: For instance, some speakers produced longer VOTs either by lengthening the glottal gesture or by shortening the oral gesture of the initial consonant. Although some variations were simply due to passive aerodynamic effects, there was systematic variation due to the segmental structure, e.g. obstruent-lateral clusters and obstruent-rhotic clusters differed: the latter showed a clear shift in their coordination patterns to ensure a substantial period of voicelessness (longer VOT) after the release of the initial obstruent. When looking at the results of prosodic strengthening, an interesting picture arises. As expected, oral gestures underwent lengthening in prosodically strong conditions, but the duration of the devoicing (abduction) gesture appeared to be stable. In stops, but not in fricatives, this abduction gesture was delayed. Furthermore, there were also spatial modifications in the laryngeal kinematics in terms of a larger glottal aperture.

Stefania Marin and *Marianne Pouplier* use EMA data to investigate the articulatory timing of liquid-obstruent clusters in Romanian onsets and codas. In Articulatory Phonology it has been assumed that complex onsets are timed differently from complex codas. Complex onsets exhibit a c-center organization, where adding a consonant to a cluster increases CV overlap. Complex codas are sequentially organized, where adding a consonant does not influence VC timing. However, it has been shown for German and American English, that different segment types, such as clear and dark /l/, might influence the local timing in codas. In German (which has clear /l/), VC timing remains unaffected when adding a consonant to the coda. In American English (which has dark /l/, exhibiting a higher degree of tongue body constriction), VC overlap increases, leading to vowel compression on the acoustic surface. In the present study, the authors test the effect of segmental make-up on onset and coda clusters for different liquid types in Romanian. They used /l/ (non-velarized, i.e. clear, in onset and coda position) and /r/ (sharing the articulatory properties of dark /l/ in terms of tongue body constriction). In onset clusters, the articulatory timing patterns were unaffected by liquid type, and both exhibited c-center organization. However, different timing patterns were found in coda clusters. In /l/-coda clusters, V and C were timed sequentially. In /r/-coda clusters there was an increase in VC overlap and vowel compression. The authors conclude that variation in segmental make-up plays a crucial role when investigating the timing patterns of onset-nucleus and nucleus-coda relations. This has to be taken into account when using CV and VC “lag” measures in the framework of Articulatory Phonology.

Sam Tilsen investigates how temporal constraints on production influence articulatory timing in onsets and codas. To do this he develops a new methodology, a cued insertion task, in which subjects repeat a CV syllable at regular intervals and have the task of inserting a further consonant into the syllable upon seeing a visual cue. The aim was to test a split-control model of syllable articulation, as laid out in Tilsen (2013). In this model there are different control mechanisms for onset and coda consonants. Whereas onset consonants are typically co-selected and coordinated with the vowel, coda consonants are usually competitively selected and thus not coordinated with the vowel. One advantage of this model is that it provides a principled account of coda timing in articulatory phonology, in particular why the coda consonantal gesture is timed relative to the attainment of the vocalic target. This is predicted by the competitive selection account, as the attainment of the target induces deselection of the vocalic gesture, allowing for the coda gesture to be selected. As predicted, these different control mechanisms for onset and coda gestures lead to different responses to time pressure: The timing of the insertion cue influenced whether the consonantal gesture was inserted as a syllable onset or a coda. Moreover, onset-syllabified insertion led to greater temporal compression than coda-syllabified insertion. There will no doubt be further testing of the model on less constrained articulatory tasks. In particular, it will be interesting to see how this model deals with extrametricality.

2.2. Section 2: Dynamics of articulation and prominence

This section consists of two papers that address issues related to effects of prominence on speech production. Mücke and Grice examine prosodic strengthening under focal prominence in German, which is mediated, but not directly modulated, by a tonal event for a pitch accent, whereas Katsika et al. explore the coordination of a tonal event for boundary marking with other gestures in Greek as conditioned by lexical stress.

Doris Mücke and Martine Grice question whether strengthening of articulation under ‘focal’ prominence – usually involving a pitch accent – stems from accentuation itself or whether it makes direct reference to a higher linguistic structure – i.e., the informational structure of the utterance. The authors compare the effects of four types of focal prominence (contrastive, narrow and broad focus, and out of focus) on disyllabic test words (e.g., ‘Bahber’), looking at acoustic durations of the initial syllable and the foot, and at the lip opening kinematics of the initial /b/. Crucially, the three different focus types (contrastive, narrow, broad), which are all marked by a pitch accent, are further differentiated in both acoustic and articulatory dimensions. More strikingly, no difference arises between the broad focus and the out of focus (background) conditions, even though they are differentiated by the presence or absence of a pitch accent. The authors note that, while a larger, faster and longer articulation tends to characterize the strongest focus condition (contrastive focus), its effect appears to be attenuated in a gradient fashion from contrastive, through narrow to broad focus. Moreover, no single dynamical parameter-setting can be pinpointed that captures prosodic strengthening of lip opening under prominence, indicating that different dynamical strategies may be employed to express the same type of prosodic strengthening. The authors underscore that accentuation (expressed by a pitch accent) may induce differential acoustic/kinematic strengthening patterns in accordance with the degree of the prominence that is reflected in the informational structure, so that in some cases (e.g., under broad focus) the presence of a pitch accent alone does not necessarily give rise to prosodic strengthening of articulation. Based on these observations, the authors conclude that “prosodic strengthening is not simply a concomitant of accentuation, but it is a means of highlighting a word in its own right,” and is thus directly modulated by the informational structure. This study paves the way for further study, in particular regarding the extent to which the gradient nature of prosodic strengthening under prominence is applicable to other languages, especially languages that do not employ pitch accents in their prominence system.

The paper by *Argyro Katsika*, *Jelena Krivokapic*, *Christine Mooshammer*, *Mark Tiede* and *Louis Goldstein* explores whether laryngeal gestures responsible for tonal events are systematically coordinated with oral gestures, and, if so, how their coordination patterns can be accounted for in dynamical terms. This study disentangles the effects of lexical and phrasal prominence on the coordination of boundary tones with the oral (vocalic) gesture within a phrase-final syllable. The trisyllabic phrase-final test words are manipulated in terms of lexical stress (e.g., *MAmima*, *maMIma*, and *mamiMA*), which occurs with or without a pitch accent. The authors show that the onset of the boundary tone (BT) is roughly timed with the target of the vocalic (V) gesture across stress/accent conditions. This is interpreted as implying an anti-phase coordination between the BT and V gestures at the boundary, although they fail to show reliable temporal stability between the gestures. They further show that the onset of BT occurs relatively earlier in the final syllable of words with non-final stress than with final stress across accent conditions, which implies that lexical stress modulates the (phrase-level) boundary tone. Another intriguing point is that the effect of stress on boundary tones parallels the earlier initiation of preboundary lengthening in Greek. The authors then provide a sketch of a unified gestural account in which both BT gestures and the oral constriction gestures are governed interactively by the π-gesture (that governs the temporal realization of gestures at the prosodic boundary) and the μ-gesture (that governs the articulatory realization of gestures under the influence of stress). The steps taken here to integrate tonal gestures with oral and prosodic gestures in the framework of Articulatory Phonology will most certainly move this line of research forward.

2.3. Section 3: Dynamics of articulation and domain edges

Three papers in this special issue are concerned with prosodic strengthening at domain edges. Georgeton and Fougeron report results of an articulatory-acoustic study on effects of domain-initial strengthening in a wide range of French oral vowels; Cho et al. concentrate on CV intergestural timing and its temporal stability in 'tautolexical' vs. 'heterolexical' conditions in Korean; and Beňuš and Šimko investigate intergestural phasing relations in Slovak arising from low-level variation in tempo and articulatory precision.

The paper by *Laurianne Georgeton* and *Cécile Fougeron* is a welcome addition to the literature on boundary-induced prosodic strengthening, examining domain-initial strengthening (DIS) effects on vowels, rather than consonants, that have so far been the focus of attention. They look at a wide range of vowels (/i, e, ɛ, a, y, ø, oe, u, o, ɔ/) in French, and investigate how effects over the phonological vowel space can be understood in terms of the enhancement of phonological contrasts. The basic finding of the study is that all vowels are produced with an increase in both lip aperture and width, although the effect is shown to be smaller for rounded than for unrounded vowels. The lesser degree of lip opening and spreading on rounded vowels, along with some additional acoustic evidence (e.g., lower values in *F*3 and *F*3–*F*2), is interpreted to indicate an enhancement of the rounding feature. The authors also discuss DIS effects in terms of expansion of the vowel space. Most vowels (i.e., /e, ɛ, ø, a, ɔ, o, u/) become more peripheral in the *F*1–*F*2 vowel space. This is closely linked to the enhancement of front/back and height contrasts. There is a tendency towards a higher *F*2 for front vowels and a lower *F*2 (and *F*2–*F*1) for back vowels (possibly enhancing the front/back contrast). There is a robust effect showing a higher *F*1 only for open and mid-open vowels, which may be seen as enhancing the vowel height, even though the enhancement was not clear across all aperture levels. The increase in *F*1 for open and mid-open vowels is seen as sonority expansion – i.e., enhancing syntagmatic contrast (with a demarcative function), but the effect is constrained by vowel height, such that high vowels resist sonority expansion, as expansion would otherwise run counter to their paradigmatic contrast (in height) with other vowels. The authors underscore that the boundary-induced enhancement of phonological contrast in a dense vowel system such as in French is not straightforwardly understood in terms of enhancement of all the phonological (distinctive) features, but rather it appears to be "conditioned by physiological constraints governing the degree of freedom of the articulators, and by linguistic constraints related to the density of the system." It is hoped that further cross-linguistic studies will follow in exploring the relationship between prosodic strengthening on the vowels and the vowel density of the language.

Taehong Cho, *Yeomin Yoon*, and *Sahyang Kim's* articulatory study is another welcome contribution to the literature on prosodic strengthening at domain edges, shedding light on how prosodic boundary and syllable structure interactively influence the temporal realization of consonantal and vocalic gestures (/m/ and /a/ in C#V and #CV). The authors, like Georgeton and Fougeron, continue to explore the locality issue of domain-initial strengthening – i.e., to what extent the V-gesture undergoes prosodic strengthening when it is strictly domain-initial (C#V) as compared to when it is non-initial (#CV). An interesting hypothesis is that the prosodic boundary effect on the V-gesture (as modulated by the π-gesture) would remain the same across #CV and C#V conditions, given the overlapping nature of CV gestures, and given the theoretically assumed independent realization of V-gesture in a separate functional articulatory tier. The V-gesture, however, is found to undergo robust temporal expansion only when V is 'strictly' initial in C#V, and only in an attenuated fashion in #CV. The authors suggest that "the influence of π-gesture may have to be fine-tuned as a function of whether the vocalic gesture is constellated with or without an onset consonantal gesture." Another important finding is that CV timing patterns are virtually identical between C#V and #CV phrase-internally, showing a potentially complete 'resyllabification' of C with V in C#V. This temporal neutralization implies that the underlyingly heterosyllabic CV gestures (belonging to different lexical items in C#V) 'reorganize' in phrase-medial position, so that they are (post-lexically determined to be) phased in an in-phase coupling mode in much the same way as the CV gestures in #CV. The authors also report that intergestural timing in #CV remains invariant regardless of boundary strength, and strikingly, CV timing is shown to be most stable after an IP boundary. This leads the authors to suggest that temporal stability is an important characteristic of domain-initial strengthening. In other words, prosodic strengthening reinforces the in-phase CV coordination. The question as to what extent the observed temporal patterns in Korean are language-specific or generalizable as characteristics of the human speech motor system will be hopefully pursued in future studies.

Štefan Beňuš and *Juraj Šimko* explore the temporal organization of articulatory gestures at prosodic boundaries by employing a novel 'continuous' ('bottom-up') elicitation paradigm that differs from 'discrete' ('top-down') elicitation methodologies typically adopted by previous studies. Prosodic boundaries with varying degrees of strength are obtained for the target sequences (/m#abi/ and /m#iba/) in Slovak by manipulating speech tempo (from 'extremely slow' to 'extremely fast') and phonetic precision (from 'extremely precise' to 'extremely relaxed'). This way, as the authors propose, "the linguistic structure provided only an affordance for the [prosodic] break to optionally, and unintentionally, emerge as a potential strategy for resolving low-level continuous prosody demands of rate and precision." An important finding is that at a weaker prosodic boundary, the onset of the /b/-closing gesture (/m#abi/ and /m#iba/) is "squeezed" between surrounding gestures (especially the preceding /m/-opening gesture): For example, the expected earlier onset for /b/ in /abi/ than in /iba/ disappears (or the pattern gets reversed – i.e., an earlier /b/-onset in /iba/ vs. /abi/), while at a stronger boundary, /b/-closing gesture moves further away from the surrounding gestures. Based on these findings, the authors propose that "[a]s the localized temporal pressure decreases, the temporal coordination of the surrounding gestures get rearranged," so that "[t]hey become less crowded." The authors model the emergence of prosodic boundaries using an embodied optimization approach, the Embodied Task Dynamics (ETD) model (see further comments below for Šimko et al's contribution). Crucially, the ETD model successfully predicts qualitatively equivalent patterns, especially for phasing relations between the /b/-onset and the following V-gesture by simulating the local temporal pressure through gradually varying demands of temporal cohesion. Phasing relations between two adjacent homorganic consonantal gestures (across a prosodic boundary as in /m#Vb/) can also be explained as the possible emergence of gradual relaxation of phasing between the opening (for /m#/) and the closing (for /#Vb/) bilabial gesture. The authors argue that modulation of inter-gestural phasing as a function of boundary strength might arise "as a result of efficient resolution of tradeoffs among articulatory effort, perceptual clarity, and localized adjustments of temporal cohesion." It will be interesting to see how far the model can be extended to account for a wider range of phasing relations under different kinds of prosodic strengthening within and across languages.

2.4. Section 4: Dynamical modeling of coordination

This section comprises three papers that combine experimental articulatory data with dynamical modeling. The models employed are Embodied Tasks Dynamics (applied to geminates and singletons in Finnish; Šimko et al., this issue), Stochastic Temporal Analysis (applied to syllable organization of simple onset parsers in Moroccan Arabic; Gafos et al., this issue) and Correlation Map Analysis (applied to assess fluctuation of coordinated articulatory movements within and between speakers; Vatikiotis-Bateson et al., this issue). All papers conclude that variation in speech processing should be treated as an important and informative component of the speech system's behavior.

The study by *Juraj Šimko*, *Michael O'Dell* and *Martti Vainio* is framed in Embodied Task Dynamics, a variant of the task dynamics approach in Articulatory Phonology. The system uses cost functions to generate gestural coordination patterns that are optimized for minimized articulatory effort and maximized speech output clarity. In this model, tasks are not abstract or context-free, but embodied. The focus in this paper is on the phonological quantity contrast in bilabial intervocalic singletons and geminates in Finnish. In a first step, the computational model predicts optimized articulatory patterns for CV**C(C)**V sequences. In these sequences they systematically varied vowel height (<pipa> vs. <pippa>, <papi> vs. <pappi>) and the place of articulation of the initial consonant of the word, (<pipa> <tipa> etc.). In an optimization task, the model generated distinct coordination patterns that account for both the consonantal quantity contrast and contextual variation. For the quantity contrast, two discrete, locally optimal patterns emerge though continuous adjustments of a parameter representing perceptual properties of the consonant in question. These corresponded to singletons and geminates. These patterns primarily differ in the coordination of the consonant and the coproduced vowel gestures. That is, the temporal lag between the onset of the intervocalic consonantal gesture (e.g. pi**p(p)**a) and the onset of the vocalic gesture (e.g. pip(p)**a**) is greater in geminates than in singletons. In terms of variation, the generated pattern reflected physiology and synergies between the different oral articulators such as jaw, tongue and lips, e.g. the intervocalic lip gesture was systematically smaller, when the preceding consonantal gesture was labial than when it was coronal (e.g. smaller in pap**(p)**i than in ta**p(p)**i). To evaluate the model prediction, the authors recorded four Finnish speakers with an electromagnetic articulograph (EMA). The articulatory results confirm the model prediction, underlining the importance of gestural *coordination* to capture variation and distinctiveness.

Adamantios Gafos, *Simon Charlow*, *Jason Shaw* and *Philip Hoole* provide an extensive and analytical discussion of the relation between qualitative syllable parses and their quantitative phonetic consequences. By using stochastic statistics, they model the variation of continuous phonetic parameters associated with distinct phonological syllable organization such as the simple onset parse. These continuous parameters use stability indices to quantify the temporal behavior of consonants and vowels that are incorporated into a given syllable (see also Marin and Pouplier). In this study, the authors focus on simple onset parses in Moroccan Arabic, a language which allows only one consonant to be in the onset, unlike English, which allows for complex onsets. In a first step, the authors critically provide a stochastic backdrop of syllable organization. In a next step, the authors evaluate predictions for a simple onset parse in Moroccan Arabic by discussing results of a previous EMA study and a corresponding simulation task, where the kinematic data served as input for the simulation task. In the simulation task, noise was systematically introduced to the system by increasing variability in the temporal intervals within the syllabic constituents. At a certain point, the system changed its behavior by showing cross-over points in the temporal stability patterns, i.e. it changed from values associated with simple onset coordination to those associated with complex onset coordination. In a next section, the authors demonstrate analytically that this behavior is in fact a prediction of the speech system's output when organized according to the demands of the simple onset. Furthermore, they systematically increase the size of the lexical sample as well as diversifying phonetic contexts to predict the influence of variation on the phonetic indices. They conclude that variation is a useful source of information on phonological structure. Thus, cases in which the phonological organization fails to provide the expected phonetic consequences also contribute to our understanding of the relation between symbolic structure and continuous manifestations of this structure.

Eric Vatikiotis Bateson, *Adriano Barbosa* and *Catherine Best* use correlation map analysis to investigate the coordination of articulatory movements within and between two speakers. Correlation map analysis is designed to capture coordination between behavioral events. It computes time-varying, instantaneous correlations between pairs of signals. Therefore, the correlation between each value in one signal and all the values in the other signal are computed and plotted in a two-dimensional correlation map that includes a time scale and a scale for signal offsets. In contrast to other methods, correlation map analysis avoids averaging over large windows. Moreover, it captures *fluctuation* of coordination in a given system. This fluctuation is assumed to occur naturally in biological systems: even though correlation values between signals might vary over time, coordination itself remains. In this study, the authors applied the correlation map analysis to kinematic recordings of two speakers talking simultaneously in a face-to-face situation, using two EMA systems. The speakers produced simultaneous repetitive productions of words (e.g. one speaker produced <top>, the other <cop>) to assess inter- and intra-speaker coordination during intentional synchronization. Subsequently, they produced spontaneous speech. Results reveal for intentional synchronization, as well as spontaneous linguistic interactions, that fluctuation in the coordination of articulatory movements, both within and between speakers, is an essential part of the speech system's behavior, while pure synchrony cannot be expected to be natural.

Despite their organization into four sections, the papers in this special issue are interconnected, both in terms of the subject matter and the methods of analysis, reflecting the workshop that sowed the seeds for this issue. To place the papers in the context in which they were originally conceived, we briefly summarize the aims of the workshop below.

3. DyMo workshop in Cologne

The workshop "Dynamic Modeling of Articulation and Prosodic Structure", held in Cologne on May 7 and 8, 2012 (http://dymo.uni-koeln.de/), was jointly organized by Doris Mücke (University of Cologne), Phil Hoole (University of Munich, LMU) and Martine Grice (University of Cologne), and funded by the German Research Foundation Priority Programme, SPP-1234 "Phonological and phonetic competence: between grammar, signal processing, and neural activity". Its aim was to bring together people doing research on articulation and prosody from a dynamical perspective, and to exchange ideas in an interdisciplinary context. Some of the areas we were looking to cover were (a) kinematics, acoustics and dynamical modeling of units in the prosodic hierarchy (e.g. syllables, feet, prosodic word), (b) head marking and edge marking (prominence and delimitation of prosodic units), and (c) prosody and Task Dynamics (speech and/or manual gestures).

Invited guests were Štefan Beňuš (Constantine the Philosopher University, Slovak Academy of Sciences), Taehong Cho (Hanyang University, Seoul), Cécile Fougeron (Laboratoire de Phonétique et Phonologie, CNRS/Sorbonne Nouvelle), Susanne Fuchs (ZAS Berlin), Adamantios Gafos (University of Potsdam & Haskins Laboratories), Laurianne Georgeton (Laboratoire de Phonétique et Phonologie, CNRS/Sorbonne Nouvelle), Louis Goldstein (University of Southern California & Haskins Laboratories), Argyro Katsika (Haskins Laboratories, New Haven), Stefan Kopp (University of Bielefeld), Tine Mooshammer (Haskins Laboratories, New Haven), Marianne Pouplier (IPS University of München), Juraj Šimko (University of Bielefeld), Adrian Simpson (University of Jena) and Eric Vatikiotis-Bateson (University of British Columbia, Vancouver).

The following paragraph from Juraj Šimko (pers com) sums up the workshop succinctly: "A unifying theme that ran through the Workshop was treating and describing speech as a communicative action, continuous change, motion. We heard of breathing cycles, pitch movement, articulation, etc. And, of course, most work has dealt with the connections between various manifestations of speech dynamics: links among perception, articulation, prosody, rhythm (embodied in suprasegmental structure of syllables, phrases), and accompanying action in other modalities (manual gestures, facial expressions)."

The workshop made considerable advances in overcoming restrictions resulting from the traditional separation between the communities working on prosody and intonation on the one hand, and articulatory phonology on the other. The talks and ensuing discussions were fruitful and constructive, stimulating cross-fertilization across the different domains. The decision to collect the work presented in the current volume[1] is seen as a first step in a process of integration of these fields that will certainly continue over the years to come.

[1] The paper by Sam Tilsen was not presented at the workshop, but is included due to its relevance.

Acknowledgments

We are grateful to Bodo Winter and Juraj Šimko, who gave us numerous insightful comments on an earlier draft of this paper.

References

Beckman, M. E. (1996). The parsing of prosody. *Language and Cognitive Processes*, *11*, 17–67.
Browman, C. P., & Goldstein, L. (1986). Towards an articulatory phonology. In: C. Ewen, & J. Anderson (Eds.), *Phonology yearbook*, *3* (pp. 219–252). Cambridge: Cambridge University Press.
Browman, C. P., & Goldstein, L. (1988). Some notes on syllable structure in articulatory phonology. *Phonetica*, *45*, 140–155.
Browman, C. P., & Goldstein, L. (1991). Tiers in articulatory phonology, with some implications for casual speech. In: J. Kingston, & M. E. Beckman (Eds.), *Papers in laboratory phonology I: Between the grammar and the physics of speech* (pp. 341–376). Cambridge: Cambridge University Press.
Byrd, D., & Choi, S. (2010). At the juncture of prosody, phonology, and phonetics – The interaction of phrasal and syllable structure in shaping the timing of consonant gestures. In: C. Fougeron, B. Kühnert, M. D'Imperio, & N. Vallée (Eds.), *Laboratory Phonology*, *Vol. 10* (pp. 31–59). Berlin, New York: Mouton de Gruyter.
Byrd, D., & Saltzman, E. (2003). The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics*, *31*, 149–180.
Cho, T. (2006). Manifestation of prosodic structure in articulation: Evidence from lip kinematics in English. *Laboratory Phonology*, *8*, 519–548.
Cho, T. (2011). Laboratory phonology. In: N. C. Kula, B. Botma, & K. Nasukawa (Eds.), *The Continuum companion to phonology* (pp. 343–368). London, New York: Continuum.
Cho, T., & Keating, P. (2009). Effects of initial position versus prominence in English. *Journal of Phonetics*, *37*, 466–485.
Cho, T., McQueen, J., & Cox, E. (2007). Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English. *Journal of Phonetics*, *35*, 210–243.
Cohn, A. (2007). Phonetics in phonology and phonology in phonetics. *Working Papers of the Cornell Phonetics Laboratory*, *16*, 1–13.
Cummins, F., & Port, R. (1998). Rhythmic constraints on stress timing in English. *Journal of Phonetics*, *26*, 145–171.
Firth, J. (1948). Sounds and prosodies. *Transaction of Philological Society*, *47*(1), 127–152.
Fougeron, C., & Keating, P. A. (1997). Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America*, *101*(6), 3728–3740.
Fowler, C. A. (1977). *Timing control in speech production*. Bloomington: Indiana University Linguistics Club.
Fowler, C. A., Rubin, P. E., Remez, R. E., & Turvey, M. T. (1980). Implications for speech production of a general theory of action. In: B. Butterworth (Ed.), *Language production, Vol. I: Speech and talk* (pp. 373–420). New York: Academic Press.
Gafos, A. I., & Beňuš, Š. (2006). Dynamics of phonological cognition. *Cognitive Science [2005 Rumelhart Prize Special Issue Honoring Paul Smolensky: Optimization and Grammar in the Cognitive Science of Language]*, *30*(5), 905–943.
Goldsmith, J. (1976). *Autosegmental phonology* (Ph.D. thesis), MIT. Distributed by IULC and published 1979 by Garland Press, New York.
Goldstein, L., Nam, H., Saltzman, E., & Chitoran, I. (2009). Coupled oscillator planning model of speech timing and syllable structure. In: G. Fang, H. Fujisaki, & J. Shen (Eds.), *Frontiers in phonetics and speech science* (pp. 239–250). Beijing: The Commercial Press.
Grice, M. (2006). Intonation. In: *2nd ed.* K. Brown (Ed.), *Encyclopedia of language and linguistics*, *Vol. 5*.
Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge: Cambridge University Press.
Haken, H., Kelso, J. A. S., & Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. *Biological Cybernetics*, *51*, 347–356.
Hayes, B., Kirchner, R., & Steriade, D. (2004). *Phonetically based phonology*. Cambridge: Cambridge University Press.
Hume, E., & Johnson, K. (Eds.). (2001). *The role of speech perception in phonology*. San Diego: Academic Press.
Katsika, A., Krivokapić, J, Mooshammer, C., Tiede, M., & Goldstein, L. (this issue). The coordination of boundary tones and its interaction with prominence. Journal of Phonetics.
Keating, P. A. (1990). Phonetic representations in a generative grammar. *Journal of Phonetics*, *18*, 321–334.
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., Cho, T., Fougeron, C., & Hsu, C. (2003). Domain-initial strengthening in four languages. In: J. Local, R. Ogden, & R. Temple (Eds.), *Laboratory phonology*, *Vol.* 6 (pp. 145–163). Cambridge: Cambridge University Press.
Kelso, J. A. S. (1981). On the oscillatory basis of movement. *Bulletin of the Psychonomic Society*, *18*, 63.
Kelso, J. A. S. (1984). Phase transitions and critical behavior in human bimanual coordination. *American Journal of Physiology: Regulatory, Integrative and Comparative*, *15*, R1000–R1004.
Kelso, J. A. S. (1995). *Dynamic patterns: The self-organization of brain and behavior*. Cambridge, MA: MIT Press.
Kingston, J. (2007). The phonetics–phonology interface. In: P. de Lacy (Ed.), *The Cambridge Handbook of Phonology* (pp. 435–456). Cambridge, UK: Cambridge University Press.
Krivokapić, J., & Byrd, D. (2012). Prosodic boundary strength: An articulatory and perceptual study. *Journal of Phonetics*, *40*, 430–442.
Krivokapić, J. (2013). Rhythm and convergence between Speakers of American and Indian English. In M. Grice & D. Mücke (Eds.), *Rhythm*, *speech timing*, *and perceptual processing (special issue)*. *Laboratory Phonology* 4 (1), 39–65.
Ladd, D. R. (2008). *Intonational phonology*. Cambridge, UK: Cambridge University Press.
Leben, W. (1973). Suprasegmental phonology (PhD thesis), MIT. Published 1980 by Garland Press, New York.
Mücke, D., Nam, H., Hermes, A., & Goldstein, L. (2012). Coupling of tone and constriction gestures in pitch accents. *Consonant clusters and structural complexity*. Munich, Germany: de Gruyter.
Nam, H., Goldstein, L., & Saltzman, E. (2010). Self-organization of syllable structure: a coupled oscillator model. In: F. Pellegrino, E. Marisco, & I. Chitoran (Eds.), *Approaches to phonological complexity* (pp. 299–328). Berlin: de Gruyter.
Nam, H. & Saltzman, E. (2003). A competitive, coupled oscillator of syllable structure. In: *Proceedings of the 12th international congress of phonetic sciences*, Barcelona, Spain. pp. 2253–2256.
Nam, H., Saltzman, E., Krivokapić, J., & Goldstein, L. (2009). Modeling the durational difference of stressed vs. unstressed syllables. In: G. Fant, H. Fujisaki, & J. Shen (Eds.), *Frontiers in phonetics and speech science*. Beijing: The Commercial Press.
O'Dell, M. L., & Nieminen, T. (1999). Coupled oscillator model of speech rhythm. In: J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, & A. C. Bailey (Eds.), *Proceedings of the XIVth international congress of phonetic sciences*, vol. 2 (pp. 1075–1078). New York: American Institute of Physics.
Ohala, J. J. (1990). There is no interface between phonetics and phonology. A personal view. *Journal of Phonetics*, *18*, 153–171.
Pierrehumbert, J. B. (1990). Phonological and phonetic representation. *Journal of Phonetics*, *18*, 375–394.
Rischel, J. (1990). What is phonetic representation? *Journal of Phonetics*, *18*, 395–410.
Saltzman, E. (1986). Task dynamic coordination of the speech articulators: A preliminary model. In: H. Heuer, & C. Fromm (Eds.), *Generation and modulation of action patterns* (pp. 129–144). Berlin: Springer-Verlag.
Saltzman, E., & Kelso, J. A. S. (1987). Skilled actions: A task-dynamic approach. *Psychological Review*, *94*, 84–106.
Saltzman, E., & Munhall, K. (1989). A dynamic approach to gestural patterning in speech production. *Ecological Psychology*, *1*, 333–382.
Saltzman, E., & Byrd, D. (2000). Task-dynamics of gestural timing: Phase windows and multifrequency rhythms. *Human Movement Science*, *19*, 499–526.
Saltzman, E., Nam, H., Krivokapić, J., & Goldstein, L. (2008). *A task-dynamic toolkit for modeling the effects of prosodic structure on articulation*. *Proceedings of the 4th international conference on speech prosody* (pp. 175–184).
Shattuck-Hufnagel, S., & Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research*, *25*, 193–247.
Shaw, J. A., Gafos, A., Hoole, P., & Zeroual, C. (2011). Dynamic invariance in the phonetic expression of syllable structure: A case study of Moroccan Arabic consonant clusters. *Phonology*, *28*(3), 455–490.
Šimko, J., & Cummins, F. (2010). Embodied task dynamics. *Psychological Review*, *17*(4), 1229–1246.
Šimko, J., O'Dell, M., & Vainio, M. (this issue). Emergent consonantal quantity contrast and context-dependence of gestural phasing. Journal of Phonetics.
Spivey, M. (2007). *The continuity of mind*. Oxford: Oxford University Press.
Tilsen, S. (2013). A dynamical model of hierarchical selection and coordination in speechplanning. *PLoS ONE*, *8*(4), e62800.
Turvey, M. I. T. (1990). Coordination. *American Psychologist*, *45*, 938–953.