Dynamics of Phonological Cognition

Adamantios I. Gafos[a], Stefan Benus[b]

[a]*Department of Linguistics, New York University and Haskins Laboratories, New Haven, Connecticut*
[b]*Department of Computer Science, Columbia University*

Abstract

A fundamental problem in spoken language is the duality between the continuous aspects of phonetic performance and the discrete aspects of phonological competence. We study 2 instances of this problem from the phenomenon of voicing neutralization and vowel harmony. In each case, we present a model where the experimentally observed continuous distinctions are linked to the discreteness of phonological form using the mathematics of nonlinear dynamics.

Keywords: Grammar; Subsymbolic paradigm; Phonology; Phonetics; Nonlinear dynamics

1. Introduction

A primary aim in the cognitive science of language is to discover the computational principles underlying our ability to speak. A major problem is how to relate the discrete aspects of our speaking competence to their continuous manifestation in terms of vocal tract action. Traditionally, the study of these two aspects of speech has been pursued under separate domains, the discrete being the domain of *phonology* and the continuous being the domain of *phonetics.*

How is the discreteness of phonological knowledge related to the continuity of their phonetic substance? This question defines the problem of the phonology–phonetics *relation,* the central theme of the laboratory phonology research community (Beckman & Kingston, 1990; Cohn, 1990; Ladefoged, 1988; Ohala, 1990). It is also an instance of a broader question in cognitive science, namely, the question of how to relate the low-dimensional, discrete aspects of cognition to the high-dimensional, detailed aspects of performance.

There is a major dichotomy between two broad approaches to this question, the symbolic and the dynamical approach. In the symbolic approach, the essential claim is that cognition is to be understood as an abstract computation involving discrete representations and rewrite rules manipulating these. The formalization of linguistic knowledge within generative grammar has provided a leading paradigm for this view of cognition. Specifically, in spoken lan-

---

Correspondence should be addressed to Adamantios I. Gafos, Department of Linguistics, New York University, 719 Broadway, New York, NY 10003. E-mail: adamantios.gafos@nyu.edu

guage this view was firmly established with the development of generative phonology (Chomsky & Halle, 1968/1991). Building on developments in the theory of computation during the first half of the 20th century, Chomsky and Halle (1968/1991) proposed to express phonological knowledge as a sequence of rewrite rules that transform an input string of symbols to an output string. What a speaker knows, then, about the phonological component of his or her language is an abstract system of discrete computation. This knowledge is manifested in the continuous substance of the speaking event, the language's phonetics. The relation between discrete and continuous aspects of phonology–phonetics is to be couched in terms of a translation between two different formal languages, discrete mathematics and continuous mathematics. This is the view in the background of most current work on language and cognitive science in general (Fodor, 1975; Fodor & Pylyshyn, 1981; Harnad, 1990; Haugeland, 1985; Newell & Simon, 1976).

An alternative view maintains that 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). The dynamical approach has been pursued for various aspects of cognition under different schools such as synergetics (Haken, 1977; Kelso, 1995), connectionism (McClelland, Rumelhart, & the PDP Research Group, 1986; Rumelhart, McClelland, & the PDP Research Group, 1986), ecological psychology (Kugler & Turvey, 1987), and morphodynamics (Petitot-Cocorda, 1985). Port and van Gelder (1995) made a recent attempt at extracting common themes and compiling representative work across such different schools.

On the one hand, the symbolic view maintains that a rigorous understanding of cognition must be couched in terms of discrete computation. This view recognizes that some, usually peripheral, aspects of cognition are best described using continuous mathematics, but tends to suppress the role of that level of analysis in understanding cognition. On the other hand, one can identify a version of the dynamical view that denies the importance of symbolic representations and rules, emphasizing instead formalization at the level of artificial neuron-like nodes and their connection strengths. At root, the debate lies in the proper treatment of a fundamental issue, the right level for the formal analysis of cognition. This is the issue uniquely taken up in Smolensky's (1988) contribution on the formal foundations of cognitive science. The key proposal in Smolensky's work is that the right level for the rigorous analysis of cognition is to be found somewhere between the two levels emphasized by the two preceding views. This intermediate subconceptual level transcends in abstraction the level of neurons and their connection strengths. Crucially, it is also lower than the level described by symbolic computation: computation at the subconceptual level is fundamentally continuous, highly context dependent, and takes the form of a nonlinear dynamical system. Smolensky's argument is explicitly engaged with the issue of relating this fundamental, continuous lower level of computation to the discrete aspects of cognition. The crucial work in this task is to be done by using and extending concepts from the mathematics of nonlinear dynamics that enable this reaching up to the symbolic level. Equally crucial is the development of new empirical methodologies that probe cognition at this level of analysis.

Work that has followed Smolensky's subsymbolic paradigm has provided a novel, rigorous basis for conceptualizing symbol-like properties of cognition such as schemas and sequential thought processes as emergent from a fundamentally continuous level of computation (Rumelhart, Smolensky, McClelland, & Hinton, 1986; Smolensky, 1986). Specifically, in the domain of linguistic cognition this paradigm has given rise to an alternative constraint-based model for linguistic grammar, Prince and Smolensky's (1993/2004, 1997) optimality theory.

The goal in this article is to extend the argument for the dynamical view by focusing on the relation between discrete and continuous aspects of phonology–phonetics. Specifically, we propose models of the relation between discreteness and continuity for two language-particular but nevertheless generalizable phenomena. The main claim is that the essential constructs of phonological cognition are dynamical in nature. This means that phonological representations and the grammar constraints governing their organization must be fleshed out in the form of continuous nonlinear dynamical systems. A key idea is that symbol-like aspects of phonological cognition emerge from the interplay of lower level continuous nonlinear dynamical systems. In effect, as envisioned in Smolensky (1988), the symbolic description of the phenomena we investigate can be seen as an approximation to the linguistic behavior captured by the subsymbolic model. The main consequence of our approach for the relation between phonology and phonetics is that, unlike in the translational view outlined earlier, in our approach there is no translation. Nonlinear dynamics enables us to integrate the discreteness and continuity of the integrated phonology–phonetics system within the same formal description.

The article is organized as follows. Section 2 covers background notions from dynamics. Section 3 discusses dynamics in the domain of phonological form. The next two sections extend dynamics to organizational principles of linguistic grammars. Section 4 addresses the interaction of environmental variables with the phonological grammar in the phenomenon of voicing neutralization. The model proposed therein maintains the discrete aspects of grammar but also accounts for the continuous phonetic variation due to changes in context parameters. Section 5 addresses the problem of how low-level spatial phonetic properties of vowels relate to the high-level phonological behavior of suffix choice in the phenomenon of vowel harmony. We conclude in section 6 with a summary of the main points.

2. Background notions from dynamics

To express theoretical ideas precisely and model experimental results, we employ concepts and tools from the mathematics of nonlinear dynamics. This section is a brief review of the basic notions.

We begin with a general formulation for the simplest class of dynamical systems, namely, *first-order* dynamical systems (Percival & Richards, 1982). These are described by a differential equation, $\dot{x} = f(x)$ where x is the state of the system and $f(x)$ is the so-called force function. This is a first-order differential equation because no derivates higher than the first derivative of x are involved. In general, the force function depends on x and the time variable t, but we limit attention here to autonomous dynamical systems, where the force does not depend on t. For such systems, also known as *gradient systems,* the force can be expressed as a function of the

derivative of a potential $V(x)$, $\dot{x} = f(x) = -dV(x)/dx$. An intuitive grasp of the dynamics of our state variable x can be inferred by examining geometric properties of $f(x)$ or its related potential $V(x)$. For example, consider the behavior of a particle placed in the potential of Fig. 1, and assume that the position of this particle, its x coordinate, corresponds to the state of the system. The points x_k where $f(x_k) = 0$ represent states of equilibrium—if our particle is placed initially at x_k it remains there for all time. Such points are called *fixed points*. There are two types of fixed points, stable and unstable. *Stable fixed points* correspond to the minima of the potential $V(x)$; see x_1, x_3 in Fig. 1. Around these points the force function $f(x)$ is a decreasing function of x, or intuitively, the arrows on the x axis in Fig. 1, which show the *flow,* point toward that point. *Unstable fixed points* correspond to the maxima of the potential $V(x)$; see x_2 in Fig. 1. Around such points, $f(x)$ is an increasing function of x, and the arrows of the flow point away from that point. Stable fixed points are also called *attractors* and unstable fixed points are called *repellers*.

Fig. 1 represents the assumption that x draws values from two recognizably distinct parts of its state space (the state space is the entire x axis). It thus describes qualitatively distinct states of the system indexed with x, or in other words, it describes a dimension of macroscopic order. For this reason, it is called an *order parameter* (Haken, 1977). In behavioral patterns, order parameters have the quality of *dynamic stability,* a term borrowed from Goodwin (1970). This means that attractive states exhibit small fluctuations around their mean values (x_1, x_3 earlier). Fluctuations are due to noise. Noise is present due to the organizational complexity of behavior, or the fact that behavior involves parallel activity of distinct faculties at different hierarchical levels. At a high level, any behavior can be described by a few parameters with dynamics that are coupled to lower level subsystems controlling the more specific components. For phonological cognition, this level corresponds to the macroscopic phonological parameters (e.g., place and degree of oral constriction, larynx state, velum position). The lower levels correspond to the neuronal, aerodynamic, and myodynamic subsystems controlling vocal tract action. This coupling between the different levels necessarily introduces noise in the dynamics of the high-level parameters. Following Haken (1977), we describe noise as a small, random perturbation force pushing the high-level parameter x back and forth randomly. Mathematically, noise enters the dynamics of high-level parameters as in Equation 1. The additional factor rep-
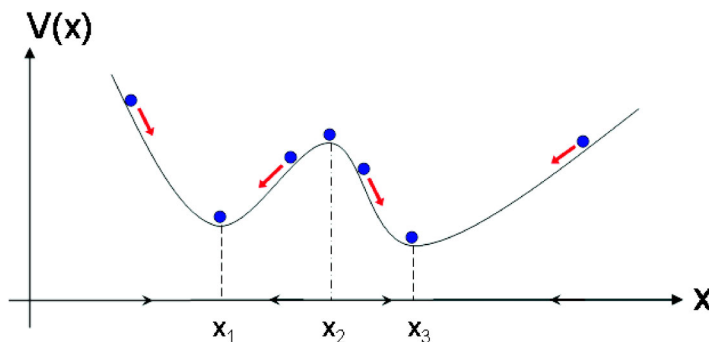

Fig. 1. Potential $V(x)$, attractors x_1, x_3, repeller x_2 of a simple dynamical system.

resents Gaussian white noise of strength q. This is the general form of a nondeterministic or stochastic dynamical system.

$$\dot{x} = f(x) + Noise = -dV(x)/dx + q\,\xi_t \tag{1}$$

The presence of noise introduces indeterminacy in the behavior of the order parameter x. Consequently, we can only compute the probability of finding x within a given region of values. This probability is described by the probability density function $p(x)$ multiplied by the length of the region. For any nondeterministic, first-order dynamical system, there exist analytical methods allowing us to compute the probability density function by finding a stationary solution to the Fokker–Planck equation (Freidlin & Wentzell, 1984; Haken, 1977). An example of a probability density function corresponding to a bistable potential is shown in Fig. 2. It can be seen that the probability of finding the state of the system, indexed by the x value, around the states of the two attractors is quite high. As we move away from the mean states, the probability of finding the system at some other region decreases quickly but it may not be zero.

Another way to estimate a probability density function is by a histogram. Assuming a system described by $\dot{x} = f(x) + Noise$, we can use the computer to numerically simulate the asymptotic behavior of parameter x and thus approximate the solutions to our equation (Higham, 2001). We then partition the state space of x (the x axis) into a number of bins, and we count the number of solutions falling in each bin. An estimate of the density measure of each bin is n^i / N, where n^i is the number of solutions falling in the bin and N is the total number of points. A histogram estimation of a density function is shown in Fig. 3. The left panel shows the potential $V(x)$ with two attractors. This is a symmetric potential in that the wells of the attractors are
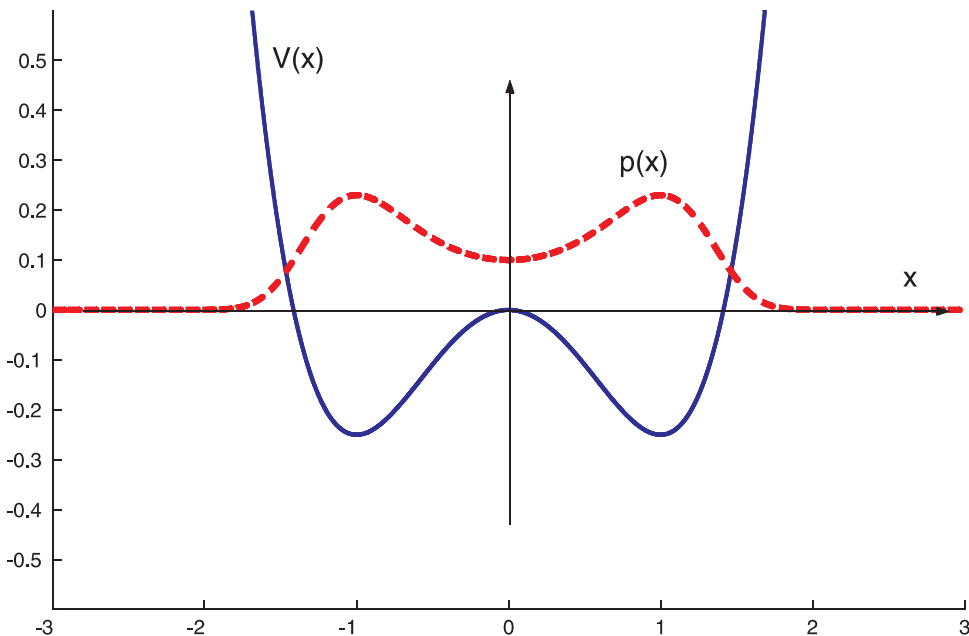

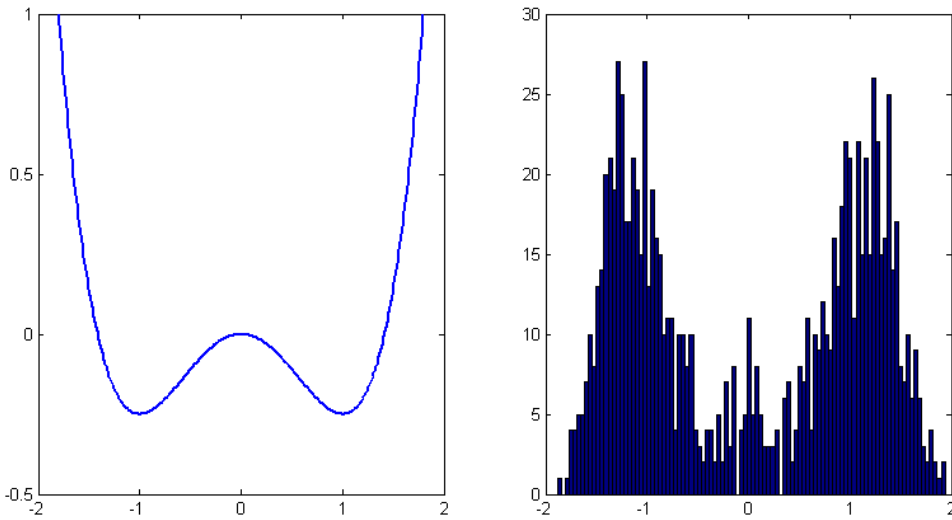Fig. 2. *V(x)* and its corresponding probability density function *p(x)*.

Fig. 3. Symmetric $V(x)$ and its density function estimate using a histogram.

equally deep. We see that the probability density estimate on the right is also symmetric with roughly equally populated regions around the attractors.

The simulation in Fig. 4 estimates the probability density function for an asymmetric potential. This allows us to illustrate the notion of the *strength* of an attractor, which depends on the depth of the minimum in the potential and the steepness of the slope toward the minimum value. Because the attractors are not equally strong, the density estimate shows two unequal
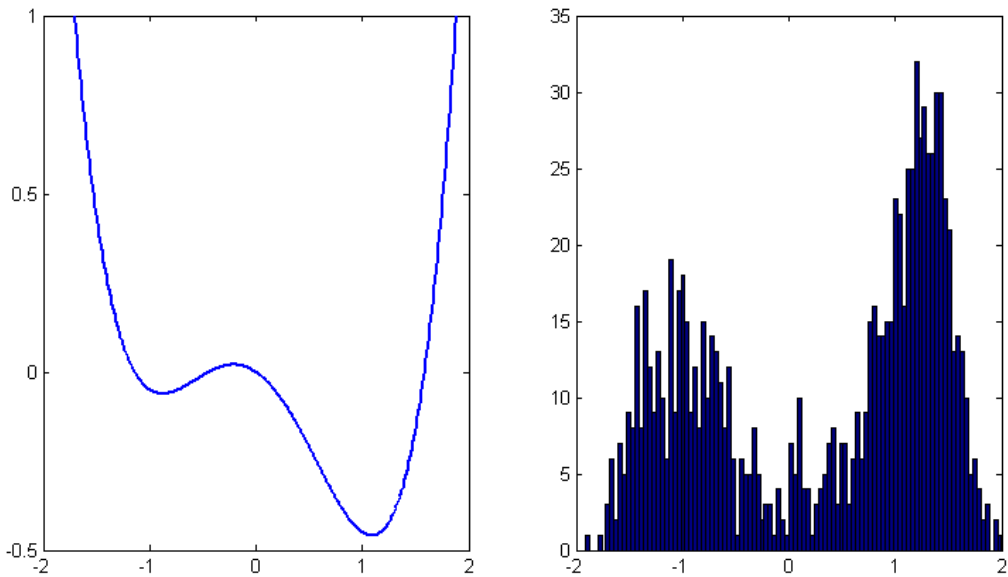

Fig. 4. Asymmetric $V(x)$ and its density function estimate using a histogram.

peaks. The histograms in these figures are generated from a total of $N = 1,000$ solution points of $\dot{x} = f(x) + Noise$, where noise is Gaussian and the initial position of the particle is a random number drawn from the interval $x \in [-1, 1]$.

The preceding theoretical results and numerical simulations show that the preferred regions of order parameters, the attractors, are resistant to noise in a probabilistic sense. It is also true, however, that in behavioral systems this stability in the presence of noise coexists with the flexibility to change. Such qualitative changes in a system's order parameter induced by scaling of some control parameter provide a particularly informative entry point into the construction of dynamical models. At a formal level, the ability to change in qualitative ways requires that we relax the property of dynamic stability. Specifically, order parameters must be resistant to noise relative to certain ranges of control parameter values. As the control parameter is scaled beyond some region of values, the order parameter may change abruptly. We are thus led to the essential notion of nonlinearity.

A system exhibits *nonlinearity* when large or discontinuous changes can be observed in the behavior of that system as some control parameter varies smoothly. In a prototypical example of this situation from speech, Stevens (1972, 1989) argued that the relation between articulatory parameters and their acoustic and auditory output is quantal in the following sense. There are certain ranges of articulatory parameter variation within which the acoustic output remains relatively stable. In other ranges, however, small variations in the articulatory parameter cause large (nonlinear) changes in the quality of the acoustic output. Put differently, gradual changes in some articulatory parameters lead to qualitatively distinct acoustic outputs. In another example from biological coordination, Kelso (1984) observed that when adults are asked to move their index fingers in an antiphase pattern (both fingers move to the left or to the right at the same time) they can perform this task over a wide range of cycling frequencies. However, as frequency is increased, participants show a spontaneous shift to an in-phase pattern; that is, to a pattern where the fingers move toward each other or away from each other at the same time. The important point in these examples is that scaling of a continuous parameter results in qualitative changes, the shift from one stable mode to another. Such qualitative changes are commonly referred to as *bifurcations* by mathematicians or *phase transitions* by physicists.

To express nonlinear relations between order and control parameters, we augment the general form of a dynamical system $\dot{x} = f(x) + Noise = -dV(x)/dx + Noise$, with a control parameter P. This gives us the equation $\dot{x} = f(x, P) + Noise = -dV(x, P)/dx + Noise$. In general, as P changes continuously, the corresponding solutions to our equation also change continuously. However, when P crosses a critical value the system may change qualitatively or discontinuously. We can illustrate this fundamental property with a simple mathematical example. Consider a force function parameterized by a control parameter k and specified by $f(x, k) = -kx - x^3$. We are interested in what happens to x—ultimately the solutions to our equation—as the control parameter k is varied. The potential corresponding to our force function $V(x, k) = kx^2/2 + x^4/4 + C$ is plotted in Fig. 5 for various values of k. For $k < 0$, as shown in the top row of Fig. 5, the control parameter changes but the system retains a qualitative sameness of form. The two minima in the valleys of the potential represent the stable fixed points of x. These are the attractors, the preferred regions within the continuum of x where the particle ends up. However, as k passes through zero, suddenly a qualitative change occurs. The system changes to a monostable
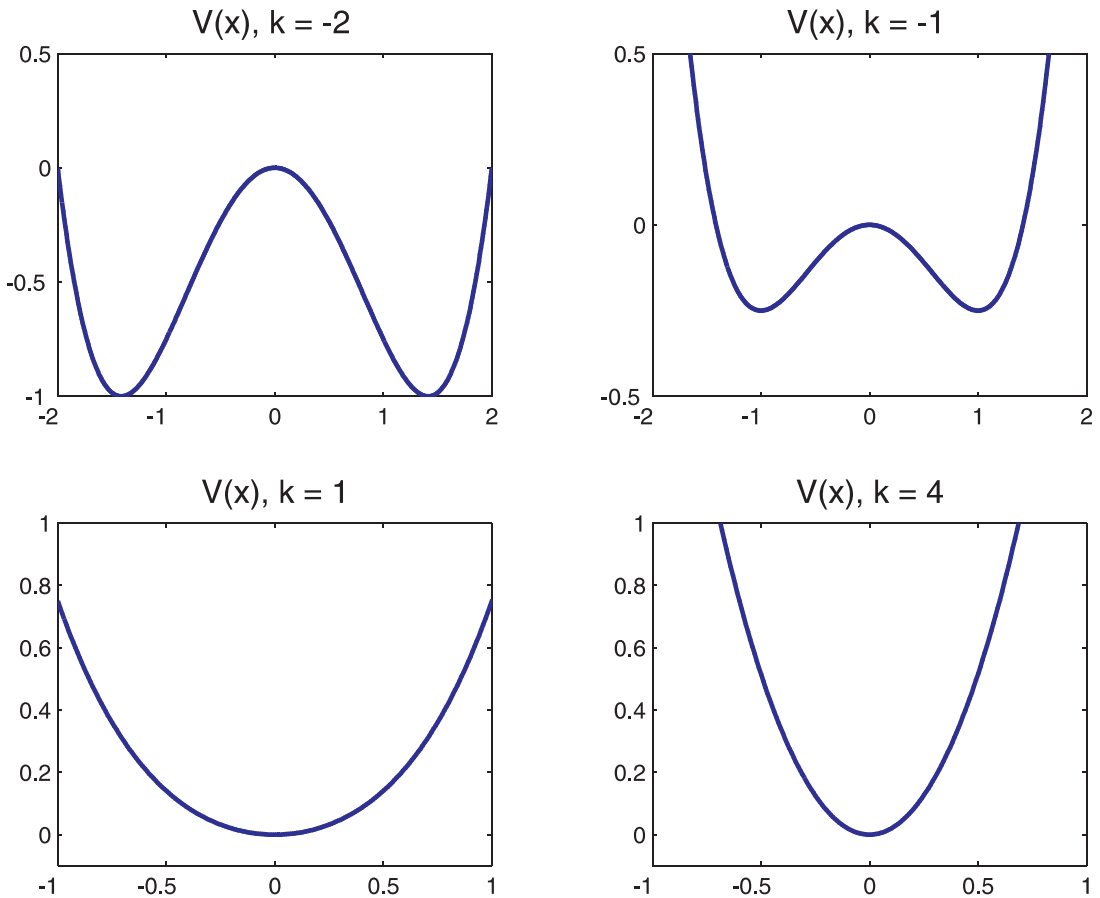
Fig. 5. Potential as a function of a control variable *k*.

regime, showing a single attractor, and retains this form for different values of the control parameter.

Within the ranges $k > 0$ or $k < 0$, variation in k does affect smoothly the attractor landscape. This is known as scaling. For example, from $k = 4$ to $k = 1$ there is a change in the stability of the attractor, as is evident from the flattening of the walls in the attractor's basin and hence the flattening of the corresponding probability density function. However, as long as k does not pass the critical value of 0, that is, within the ranges $k > 0$ or $k < 0$, the macroscopic form of the system remains unaltered. It is only when k passes a certain critical value (here, $k = 0$) that a qualitative change takes place.

Next, we consider how the dynamical concepts introduced here can be applied to modeling speech patterns. Section 3 discusses dynamics in the domain of linguistic form. Section 4 addresses the incompleteness of phonological neutralization, and section 5 focuses on the relation between continuous and discrete dimensions of speech patterns in vowel harmony.

3. Dynamics and the nature of phonological form

The main claim of this article is that the units of phonological structure and their organizational principles are dynamical in nature. Smolensky's (1988) proposal can be seen as a precursor to this claim for cognition in general. In the specific domain of the relation between phonology and phonetics, our claim finds another important precursor in the dynamically based theory of phonological form developed by Browman and Goldstein (1986). To highlight the connections between our claim and previous work, we provide a synopsis of the main aspects of that theory.

A fundamental starting point of phonological theory is that utterances can be decomposed into sequences of discrete units. All competing models of phonology embody this assumption. Beyond this, there is considerable disagreement about the nature of these units. In phonological theory, the central issue is the nature of phonological representations. A major theoretical dichotomy has emerged here between two views, the symbolic and the dynamical view.

The symbolic view originates in the foundational work of generative phonology, Chomsky and Halle's (1968/1991) *Sound Pattern of English* (*SPE*). In this view, utterances take the form of a two-dimensional phonological matrix. The columns are placeholders for the elements in a symbol string of phonological segments (e.g., /m-a-p/). The rows are the distinctive features, such as [±Nasal], [±Consonantal], [±Continuant], and [±Voiced], each classifying a segment as a member of a pair of opposed categories. For example, the nasality of the first segment in *map* is specified with a + in the row for [Nasal]; a – would declare membership in the complementary category as is correct for the final segment of *map*. The cell for the feature [Consonantal] takes the value + for segments p, m and – for the *a* vowel, and so on (see Anderson, 1974, on features). To anticipate a point of contrast with the dynamical view, we note that these representations are static or atemporal. There is no notion of time except for the trivial left-to-right ordering of segments in the sequence (e.g., /m^1-a^2-p^3/). For example, two segments cannot overlap with one another. Rather, one can only follow or precede another.

In the years following *SPE,* the two-dimensional phonological matrix was adopted as the standard for phonological analyses. As the body of work using this representation quickly expanded, it became clear that this system had to be refined in important ways to incorporate tonal contrasts, multiply articulated or complex segments, and phenomena where features extend beyond a single segment as in vowel harmony (Anderson, 1976; Clements, 1985; Goldsmith, 1976; Sagey, 1986/1991). This led to the theory of *autosegmental representations* where features continue to be the basic unit of phonological representation, but more insightful description of sound patterns is gained by adding further structure to the feature set.

Both *SPE* and autosegmental representations share a key working assumption about how their discrete units relate to the continuous aspects of the speech event. Phonological representations are discrete, but phonetics is continuous. Because of this disparity, phonology and phonetics are assumed to be fundamentally separate but related through a process of translation from discrete symbols to continuous properties of an articulatory and acoustic nature. Features and further symbolic structure in the phonological matrix serve as control commands for setting the vocal tract in action with time-varying spatial and acoustic characteristics. For this view, see Chomsky and Halle (1968/1991) on the notion of physical scales and the notion of implementation in Keating (1990), Cohn (1990), Pierrehumbert (1990), and Coleman (1992).

The symbolic and fundamentally translational view just outlined contrasts with a more recent and less widely explored view fleshed out in Browman and Goldstein's theory of articulatory phonology. There are three interlocking hypotheses surrounding this theory. The first is that the variability and context-specificity of speech movements can be captured by dynamical units called gestures; gestures are the primitives of speech production. The second hypothesis holds that these same primitives can also provide a cognitively plausible basis for a theory of phonological representations; gestures are the primitives in the phonological component of linguistic grammars. The third hypothesis is that the laws governing gestural action are central also to the process of understanding speech; gestures are the primitives of speech perception (see the motor theory of Liberman & Mattingly, 1985, and the direct realist view of Fowler, 1986).

Before addressing the dynamical aspects of gestures, we consider the spatial dimensions over which gestures are partly specified. A speech gesture is an intentional event consisting of the formation and release of a constriction at some place in the vocal tract. The constrictors consist of a small set of anatomically separate but linked organs: the lips, the tongue tip, the tongue body, the larynx, and the velum. The distinction between the organ forming a constriction and the individual articulators contributing to this movement is important. As a rule, vocal tract action corresponding to a speech gesture involves multiple articulators. For example, the consonants [p], [b], and [m] share a lip-closing gesture effected by the synergistic activity of a *coordinative structure* (Fowler, Rubin, Remez, & Turvey, 1980), in this case, a set of three individual articulators: upper lip, lower lip, and jaw. Coordinative structures are functional; that is, each of the articulators comprising the coordinative structure contributes flexibly in attaining the constriction goal or task of the gesture. Thus, in attaining the task of lip closing, the upper lip, lower lip, and jaw contribute to different degrees as a function of the vocalic context; for example, in [aba] the jaw shows a greater extent of movement than in [ibi] due to the differences in the jaw position of the surrounding vowel contexts. Hence, the spatial dimensions for gestures, the *vocal tract variables,* are specified not at the level of individual articulator positions but at the functional or task level of constriction goals. At that level, the lip closing gesture for [p], [b], and [m] has a target specification for the lip aperture (LA) tract variable, Target-LA = {closure}. Gestures can differ on the basis of the organ used to form the constriction and thus on the basis of the tract variables being controlled; for example, [b]'s constrictor organ is the lips and the relevant tract variable is lip aperture, whereas [d]'s organ is the tongue tip and the relevant tract variables are tongue tip constriction location and tongue tip constriction degree. Gestures can also differ on the basis of tract variable target values for shared tract variables: thus, both [d] and [s] involve the tongue tip organ but [d]'s Target-Constriction Degree is {closure} whereas [s]'s is {critical}, the target Constriction Degree value of fricative consonants, and so on (labels like {closure}, {critical} are cover symbols for numerical value ranges corresponding to language-particular degrees of stricture or points of articulation along the longitudinal axis of the vocal tract).

The most important aspect of gestures is that they are dynamical units. There are two senses of dynamical here. One sense implies motion. As usually described, gestures involve motion. However, to claim that motion is a definitional property of gestures is confusing: How can entities defined in terms of physical movement serve as mental units of representation in phonological cognition? To eliminate confusion we must turn to the other sense of dynamical, which

is central to the point we want to emphasize here. To preview, gestures are dynamical in that their model employs the mathematics of linear and nonlinear dynamics. The model takes the form of an invariant mathematical law. This law is hypothesized to give rise to the continuous movement trajectories of gestures. It is at the level of such laws where the power of any dynamical theory inheres. Once this point is understood, it becomes clear that it is misleading to equate kinematics with gesture or to take motion as the defining property of gestures.

We now make this point in more explicit terms. The continuous movement of gestures unfolds from the task dynamics model of Saltzman (1986), and Saltzman and Kelso (1987). Part of the task dynamic characterization of a gesture is a mathematical model of the continuous trajectory involved in attaining the task of the gesture; that is, the formation of a constriction at some point in the vocal tract. As a gesture unfolds, the tract variables change in value over time, and can thus be modeled as dynamical systems. Consider, for example, the constriction degree (CD) and constriction location (CL) tract variables of the alveolar stop [d]. These have target values Target-CD = {closure}, Target-CL = {alveolar}. The CD tract variable, independent of its initial value, will approach its target value, which is {closure}, complete contact between tongue tip and palate. It is this continuous change in the value of CD that is modeled as a dynamical system. The same applies to modeling the continuous change in CL. The specific mathematical model employs the dynamics of damped mass-spring systems (a *point-attractor;* Abraham & Shaw, 1982). These are governed by the second-order differential equation in Equation 2.

$$\text{Dynamical model of gestures: } d^2x / d^2t + bdx / dt + k(x - x_0) = 0 \qquad (2)$$

Like any differential equation, Equation 2 describes how x, a cover symbol for tract variables like LA or CD, changes over time. Different evolution paths of x values can be obtained by tuning the system parameters, the initial position of x, the stiffness k, and the equilibrium position x_0 (we turn to this point later). The equilibrium position x_0 is the target value of the tract variable being modeled by the equation. Thus, for any given gesture, the tract variables represented with x in Equation 2 change in time, but x_0 is a constant corresponding to some fixed value of Target-CL or Target-CD.

It is useful to consider how this model provides a basis for capturing the variability and context specificity of speech movements. One property of the task dynamic model is that when gestures overlap in time their influences on common articulators are blended. This captures one of the most fundamental facts about speech, the context dependency observed in spatial characteristics of speech movements. A well-known case comes from Öhman's (1966) study of coarticulation. In X-ray movies of velar stop production, Öhman observed that the location of linguo-palatal contact of [g] varies systematically with the surrounding vowel context: In [ugu] versus [igi] the respective regions of linguo-palatal contact during the closure of [g] do not overlap. This posed problems for Öhman's coarticulation model, which had to be complicated in ad hoc ways to describe this dependency between the consonant's contact location and its vowel context (Öhman, 1966, p. 319). In the gestural model, this dependency falls out from the temporal dimension of gestural units. Utterances, in this model, are not atemporal linear sequences of static units. Rather, gestures have internal temporal structure and therefore can overlap in time. When a vowel gesture overlaps with the gesture of a velar stop, the two gestures impose different demands on the same tongue mass and thus on the same tract variables.

Hence, the tract variable motions will be affected by the target values of both gestures. This gives rise to the context-specific effects on the locus of linguo-palatal contact for [g] as the vocalic environment is varied. At the underlying specification level, however, the gestures of the velar stop and the surrounding vowels have their respective invariant target values for tongue body CL.

It is not only in spatial but also in temporal aspects of speech movements where variability and context specificity are found. Temporal changes in phonological form are a salient feature of speech prosody. One of the well-established prosodic phenomena is the lengthening of gestural durations at word and phrase boundaries. It has been proposed that prosodic lengthening effects can be derived by changing the stiffness parameter k of the boundary adjacent gestures (Beckman, Edwards, & Fletcher, 1992; Byrd & Saltzman, 1998). In the dynamical model, lowering the stiffness k has specific kinematic consequences, resulting in longer movement duration and smaller peak velocity/movement amplitude ratios. There is also evidence that phrasal boundaries change the relative timing of their local gestures. An account of this has been proposed by locally tuning the time course of gestural activation (Byrd & Saltzman, 2003). This is done using a scaling analysis of the time parameter t in the task dynamic model, so that effectively the same gestures can be slowed down or speeded up at the level of the central clock. In each case, simulations of the proposed dynamical model are in good agreement with the experimentally observed patterns.

One central idea about the role of dynamics emerges from this discussion. The dynamical model states an intimate and immediate link between the abstract invariance of gestural units and the continuity of their phonetic substance. Gestures are abstract and invariant in the following respects. First, gestures are specified by context-independent target values for the tract variables—the x_0 of Equation 2. These targets do not change during the lifetime of the dynamic event that is a gesture. For example, [b] has a different target constriction degree from [w], [d] has a different target CL from [g], and so on. Gestures are invariant also in that their model does not change during their lifetime. Equation 2 expresses an invariant law. This is the law that gives rise to a gesture's kinematic patterns. As we have seen, the kinematics is continuous and context specific. However, the underlying law governing this variability is an abstract statement in the form of a differential equation. Dynamical laws then stand at the same level of abstraction as the invariant context-independent units of representation in the symbolic view. Crucially, however, it would be misleading to view the kinematics as implementing these (dynamical) symbols. This is because no additional formal construct is needed to go from the dynamical law defining gestures to the continuity and context specificity of their kinematic patterns. This is in contrast to the symbolic view. As emphasized earlier, this view necessitates a translation from phonology to phonetics, embodied in the construct of implementation rules.[1]

We have so far seen that gestural units provide a basis for a rigorous study of the context-dependent details of speech (gestures are the primitives of speech production). The second essential hypothesis in Browman and Goldstein's view is that gestures also serve as abstract combinatorial units of phonological computation in linguistic grammar. Several parallel research lines are devoted to this hypothesis.

One important domain of work has been to elucidate the nature of higher phonological units like segments and syllables in terms of the primitives of the model, namely, gestures and their temporal relations. Saltzman and Munhall (1989), Browman and Goldstein (1991), and Byrd

(1996) proposed that the phonological unit of segment may correspond to a characteristic temporal overlap relation among the segments' constituent gestures. Saltzman and Byrd (2000) offered an explicit computational model of this idea using the notion of relative phase in coupled oscillators. A very similar idea has led to a particularly promising characterization of the phonological unit of the syllable. Previous proposals to define the syllable in terms of articulatory and aerodynamic parameters have proved ineffective. The most explicit of them, proposed by Stetson (1951), hypothesized that syllables correspond to the pulses created by contractions of the intercostal muscles, which control lung volume during speech. Later studies of pulmonary air pressure during speech showed that this hypothesis could not be maintained (Draper, Ladefoged, & Whitteridge, 1960). These studies revealed that lung pressure is kept relatively steady over the course of the production of a sentence and that the slight variations in pressure do not correspond perfectly to Stetson's syllable pulses. Research within the articulatory phonology view, in contrast, has developed empirically progressive methods and hypotheses centered on the idea of defining syllables as characteristic patterns of gestural coordination. Honorof and Browman (1995) found differences between onset and coda consonants in their timing relation to a nucleic vowel. Browman and Goldstein (2000), in particular, provided evidence for the hypothesis that such patterns can be seen to emerge as the result of numerical optimization under multiple, competing coordination relations among gestures. In refining this idea and modeling it explicitly, Nam and Saltzman (2003) showed how asymmetries in variability of timing relations in syllable onset versus coda position can be derived from the interaction of competitive coupling relations among gestures comprising syllables.

Another domain of work is devoted to reanalysis of phenomena that seem to fall squarely within the symbolic view of phonological form. A central concern of phonological theory has been the proper characterization of cases where the same phoneme (or feature thereof) varies as a function of other sounds in its context, what is known as *allophonic variation.* An example is the difference between clear and dark allophones of English /l/, as in *lip, late, lie* versus *pill, feel,* and *cool* ([l] versus [ɫ]). In the symbolic view, this difference is expressed by saying that the basic allophone is the clear /l/ and in syllable-final position this changes to the dark or velarized version by a feature-change rule adding the feature [+back]. Looking at movements using the X-ray microbeam system, Sproat and Fujimura (1993) found that English /l/ is composed of two gestures, a tongue tip consonantal gesture and a tongue dorsum vocalic gesture, and that the relative timing of these varies as a function of syllable position and adjoining prosodic boundary. In syllable-initial position, the two gestures show a synchronous pattern of relative timing, with tongue tip and tongue dorsum attaining their goals at the same time. In syllable-final position, the tongue dorsum gesture significantly precedes the tongue tip gesture, with the tongue dorsum attaining its target at the onset of the tongue tip gesture. In syllable-final position, then, the acoustic portion of the syllable corresponding to the vowel is significantly more overlapped with the tongue dorsum gesture. The acoustic consequences of this difference in overlap are what give rise to the distinction between clear and dark /l/s. This finding is of considerable theoretical interest as it shows that what appears to be from a distributional point of view a single /l/ segment is in fact decomposed into two separate elements, a vowel-like gesture and a consonant-like gesture, and that the timing of these two gestures is mainly responsible for the distinction between the clear and dark variants of /l/ (Browman & Goldstein, 1995b; Sproat & Fujimura, 1993). Krakow (1989, 1993), on English nasals, found a

similar pattern of timing between the component gestures of velic lowering and oral closing, and showed how this pattern can explain the allophonic variation between oral and nasalized vowels as in *meat* versus *team*. Similarly, Browman and Goldstein (1990) and Zsiga (1995) showed how fast speech phenomena such as apparent assimilations and deletions of segments can be insightfully seen as consequences of increased temporal overlap among gestures with varied acoustic consequences. In effect, this research provides evidence for the claim that gestures are the right combinatorial units of phonological computation.

In a study of the phonological system of Moroccan Colloquial Arabic, Gafos (2002) argued that phonological knowledge more abstract than allophonic variation can make reference to the temporal dimension of linguistic form. This proposal makes contact with optimality theory (Prince & Smolensky, 1993/2004, 1997) by expressing language-particular patterns as the result of optimization under a set of violable constraints, some of which must crucially refer to temporal relations among gestures. Benus, Smorodinsky, and Gafos (2004), Davidson (2003), and Hall (2003) also pursued a model of grammar based on gestural representations and optimality theory in analyzing independent phenomena in other languages. Other areas of research extending the articulatory phonology view include rhythm (Smith, 1991), the nature of speech errors (Pouplier & Goldstein, 2005), and the emergence of discrete gestural specifications from a process of attunement between computational agents (Goldstein, 2003).

In sum, several parallel lines of work seem to provide converging evidence for a gesturally based theory of phonological form. The leading idea is that the central token of exchange in linguistic communication is the gesture, a dynamically defined primitive shared across the three building blocks of linguistic cognition, grammar, production, and perception. For our purposes, the distinctive property of the dynamical view of phonological form is its use of a mathematical language in which discreteness and continuity coexist. Our aim in the next two sections is to extend the argument for the dynamical basis of phonological cognition from representations to organizational principles of linguistic grammars.

4. Interaction between grammar and communicative context

The translational view of the relation between phonology and phonetics expresses a valid intuition, namely, that there is a distinction to be made between continuous and discrete aspects of phonological cognition. For certain phenomena, however, the specific way of drawing that distinction within the translational approach proves to be too rigid. This section takes up one class of such phenomena whose defining property is the interaction between discrete grammatical requirements and continuous communicative or environmental variables. As we will see, the translational view precludes this kind of interaction. In turn, this interaction underscores the need for a formal language integrating discrete and continuous aspects of the phonology–phonetics system. Using nonlinear dynamics, our aim is to provide a way of reconciling the valid intuition of the translational view with cases of interactionism.

A basic property of phonological systems is that the phonetic properties of sounds are not equally distributed across all syllabic positions. For example, in the phonological system of German, consonants that can exhibit a contrast in voicing (e.g., voiceless [t] vs. voiced [d]) must be voiceless when they occupy a syllable-final position. Thus, *Rad* 'wheel' is produced as [ʁat]

'wheel-Nominative', with a voiceless [t], when that consonant is at the end of a syllable, but with a voiced [d] in all other contexts (e.g., [ʁadəs] 'wheel-Genitive'). Moreover, whereas the produced form of 'wheel' alternates between a voiced and a voiceless final consonant, other words do not show this alternation. For example, *Rat* 'advice' is always produced with a [t], [ʁat] 'advice-Nominative', [ʁatəs] 'advice-Genitive'. Because the contrast between 'wheel' and 'advice' is neutralized in some positions, in that both are produced with a final [t], this phenomenon is referred to as (*contrast*) *neutralization* (Bloomfield, 1933, p. 218).

When the performance of German speakers is closely observed, however, it deviates slightly from the simple preceding description of neutralization. The basic result is that the [t] in [ʁat] 'wheel-Nominative' is pronounced quantitatively differently from the final consonant of [ʁat] 'advice-Nominative'. Specifically, in [ʁat] 'wheel-Nominative', the [t] shows traces of voicing that make it distinguishable from the completely voiceless [t] of [ʁat] 'advice-Nominative'. Moreover, this difference depends on the communicative context. Port and Crawford (1989) showed this by setting up different experimental conditions in an attempt to study speakers' behavior in different communicative contexts. In one condition, speakers read a given word list. In another condition, speakers are asked to read sentences like *Ich habe Rat(Rad) gesagt; nicht Rad(Rat)* "I said Rat(/Rad) not Rad(/Rat)" while a German assistant, present in the experimental setting, is assigned the task of writing down the order of the test words in such sentences. In the latter task, speakers are encouraged by the context to convey the difference between *Rad* 'wheel' and *Rat* 'advice' more than in the word list reading task. The result is that the neutralized final consonant of *Rad* in [ʁat] 'wheel-Nominative' shifts more toward a voiced [d] than in the word list reading task.[2]

We now turn to see how neutralization can be analyzed in phonological theory. We first consider an analysis based on the conception of linguistic grammar as a set of feature-based representations and hard rules operating on these (Chomsky & Halle, 1968/1991). In such a theory, the standard way of expressing the German neutralization phenomenon is to say that German speakers have internalized a rule changing voiced consonants to their voiceless counterparts when they are in a syllable-final position. This is the rule of final devoicing shown in (3).[3] There are three parts to this rule. The focus part '[+Voiced, −Sonorant]' lists the features specifying the class of segments undergoing the rule (here, the voiced obstruents), the change part '→ [−Voiced]', specifies the featural change effected by the rule, and the environment part '/ __]$^\sigma$' specifies the context in which the rule applies (here, at the end of a syllable). In addition to this rule, German speakers internalize distinct representations for the basic forms of 'wheel' and 'advice' in their mental lexicon, /ʁad/ and /ʁat/, respectively. Phonological rules, like final devoicing, mediate between these input or underlying representations and their surface realization. Following standard practice, we use / / for the underlying representations in the mental lexicon, [ ] for their surface representation, and . for syllable breaks. Thus, final devoicing applies to the mental lexicon entry for 'wheel', /ʁad/, to produce [ʁat] but fails to apply to /ʁat/ or /ʁad + əs/ as shown in (3).

(3) Final Devoicing (*FD*): [+Voiced, −Sonorant] → [−Voiced] / __]$^\sigma$

     Input   *Output*

  a. /ʁad/   → [ʁat] (*FD* applies)

  b. /ʁat/   → [ʁat] (*FD* cannot apply as [t] is not [+Voiced])

c. /ʁad + əs/   →   [ʁa.dəs] (*FD* cannot apply as [d] is not syllable-final)

There are two properties of this analysis that are difficult to reconcile with the phonetic facts of neutralization. One is that grammatical computation is based on discrete categories—the representations and the rules referring to these are spelled out in the language of symbols. The variable phonetic outputs in the neutralization facts cannot be fully described using categories. The final [t] of [ʁat] 'wheel-Nominative' is voiceless but it is quantitatively different from other voiceless segments, for example, [t] of [ʁat] 'advice-Nominative'. The other property is that phonological computation precedes the phonetic implementation of grammatical outputs. Recall that in the output of phonological computation, the voicing value of the final consonant of *Rad* 'wheel-Nominative' is identified with the featural symbol [–Voiced]. This eliminates the contrast between the final consonants of *Rad–Rat* at the output of phonological computation, as a neutralization rule should do. Consequently, the phonetic implementation transducer, whose role is to flesh out the phonological output as vocal tract action, is now unable to deliver the difference observed in the realization of the final consonants in *Rad* versus *Rat*.[4]

We thus revisit the neutralization problem with the goal of accounting for the fact that the speakers' phonetic outputs adapt flexibly and in a purposive way to demands dictated by the communicative context. This requires a more flexible computational system able to deliver differences of a continuous nature and an explicit link between context and grammar so that the two can interact in ways that derive the observed systematicities.

The starting point of our proposal is a fundamental insight about grammars, namely, the idea that linguistic patterns can be seen as the result of optimization under a set of soft constraints or competing principles. This is the idea embodied in independent proposals by Goldsmith (1993), Burzio (1994), and in its most developed form by Prince and Smolensky's (1993/2004, 1997) optimality theory. According to this theory, grammars are best formalized as constraint optimization systems rather than systems of rewrite rules. Two broad classes of constraints are recognized. Markedness constraints define dimensions of harmony in linguistic structures, such as "a syllable must have an onset" or "a vowel must be oral (not nasalized)." Faithfulness constraints require that the output of phonological computation preserves properties of its input in the mental lexicon. To exemplify with voicing neutralization, there are two relevant constraints. The markedness constraint requires coda consonants to be voiceless. The faithfulness constraint requires identity between the representation of a consonant in the mental lexicon and its surface form. When the input is /ʁad/, the constraints are in conflict. Output [ʁad] preserves the voicing as demanded by faithfulness but violates markedness. Output [ʁat] satisfies markedness but violates faithfulness. Because in German the surface form is [ʁat] the constraints are prioritized so that markedness is ranked higher than faithfulness. This is written as markedness » faithfulness. In principle, optimality theory maintains that all languages draw from the same set of constraints and cross-linguistic differences can only derive from different prioritization of these universal constraints. For example, when the constraints in our neutralization example are ranked differently (i.e., faithfulness » markedness), the result is preservation of voicing from the underlying level to the output, /ʁad/ → [ʁad]. This is the ranking that would be appropriate for those languages that do not neutralize voice (e.g. English). Other less trivial typological results have been achieved using this basic formalism. As a consequence of its gen-

erality, optimality theory has obviated a range of highly specific tools and postulates of previous theories of grammar. It is remarkable that a theory so parsimonious in the number of basic principles seems to provide sufficient predictive power to account for the range of observed linguistic specificity within and across languages.

Adopting the insight of constraint optimization, the plan now is to build a dynamical formulation of the constraint-based view, integrating the discreteness of the phonological system with the flexibility of phonetic performance. The essential units of phonological cognition, the phonological representations and the grammar constraints on these, will be formulated in dynamical terms.[5] Specifically, the faithfulness and markedness constraints will be situated in the continuous realm of forces on phonetic variables as opposed to symbols. This provides a basis for capturing the systematic scaling of the phonetic output as contextual parameters change.

We begin by a dynamical formulation of the faithfulness constraint. Consider a communicative act wherein the speaker's goal is to convey the word *Rad*. Because in the mental lexicon *Rad* is stored with a voiced final segment, /ʁad/, we can express the intention to convey this form by an intentional dynamics that contributes an attractor at the required value of voicing. The same applies to *Rat,* where the intention is voiceless. This idea is depicted in Fig. 6, showing the potential functions for voiced and voiceless intentions. The *x* axis represents the space of all possible voicing values, which we assume can be indexed by the degree of glottal aperture. The voiceless intention is represented with an attractor, a minimum of the potential function, at some positive value of glottal opening x_0. Similarly, the voiced intention is represented with an attractor at some negative value of glottal opening, $-x_0$ (the actual numeric values are not crucial in this context).

We can visualize the glottal aperture variable as the position of a particle placed in one of the preceding potentials. Intention acts as a force constraining the glottal aperture variable to fall
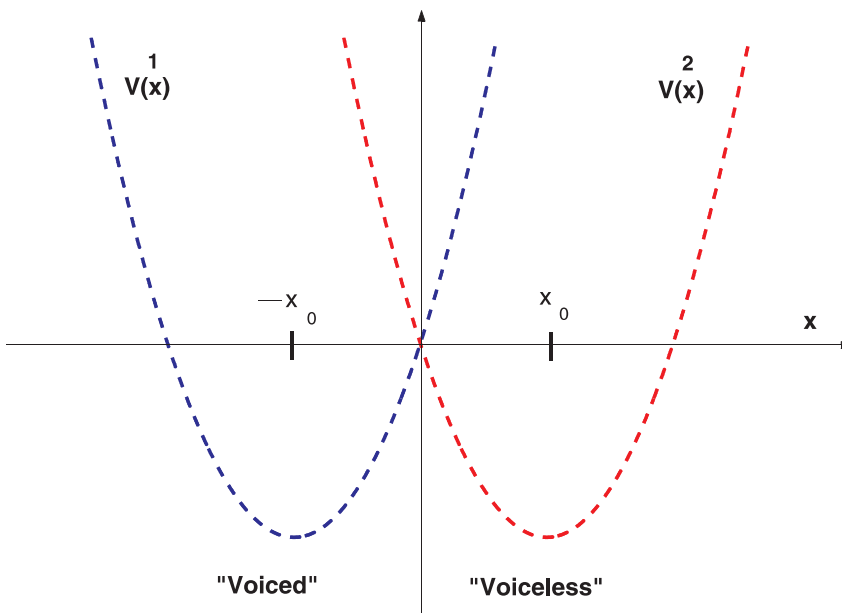
Fig. 6. Potential functions for voiced and voiceless intentions.

within a region of values around the mean characteristic for the intended voiced or voiceless consonant. Hence, the attractors at the mean values x_0, $-x_0$ describe qualitatively distinct states of the voicing system or, in other words, they describe a dimension of order in the continuous space of phonetic variables. In a symbolic statement of the voicing contrast, the situation is written in terms of the mental binary distinction [±Voiced]. This notation describes two discrete categories abstracted away from phonetic substance but ultimately translated to that substance via a phonetic implementation transducer. The difference between the symbolic and the dynamical formulation is that in the latter (but crucially not in the former) the distinct modes of voicing are inseparably linked with the phonetic substance. These modes are not derivationally antecedent to that substance and therefore they do not need to be translated to that substance.

Formally, as in any (autonomous) dynamical system, intention is modeled by a differential equation of the form $\dot{x} = F(x)$. $F(x)$ is the simplest function that admits a stable fixed point at the (intentionally) required value of voicing, or $F(x) = \theta(x^{REQ} - x)$. The term x^{req} takes values from $\{-x_0, x_0\}$, that is, the glottal aperture values corresponding to [+Voiced], [–Voiced] consonants, and the coefficient θ represents the relative strength of the intentional contribution. To derive the intention potential, we use $\dot{x} = F(x) = \theta(x^{REQ} - x) = -dV(x)/dx$, and by basic calculus we can derive $V_F(x) = \theta x^2/2 - \theta x^{REQ}x + C$, up to the constant C, which can be dropped because it is of no qualitative significance in the context of this discussion. The potentials for two values of x^{REQ}, $\{-x_0, x_0\}$, are shown in Fig. 6.

Next, we describe the continuous equivalent of the markedness constraint, requiring coda consonants to be voiceless. To state this constraint in dynamical terms, we specify a potential function that contributes an attractor at the appropriate value of voicing. The required potential is shown in Fig. 7. Because codas can only be voiceless, the attractor is at a value characteristic of voiceless consonants. A particle left in this potential ends up at the minimum, representing a value of glottal aperture appropriate for voicelessness.

Formally, the markedness constraint can be defined by a differential equation of the general form $\dot{x} = M(x)$. As a working hypothesis, we assume that $M(x) = \dot{x} = -k + x - x^3$. Given that $-dV(x)/dx = \dot{x}$, we can compute by integration the potential $V_M(x) = kx - x^2/2 + x^4/4$ (see Tuller, Case, Ding, & Kelso, 1994, for an application to perception). This is the potential shown in Fig. 7. The attractor is at some positive value of glottal opening on the x axis characteristic of voiceless obstruents.[6]

We can now look at the interaction between the markedness and faithfulness forces in the dynamical setting. Formally, the simplest model for this interaction corresponds to a linear combination of the M, F forces, that is, $\dot{x} = M(x) + F(x)$ (following Schöner & Kelso, 1988, on coordinated movement by humans). Consider what this model predicts for the intention to convey *Rad* 'wheel'. In German, as suggested by the optimality theory analysis, the M force dominates the F force. Thus, when the two forces combine, the result is pulled toward the voiceless attractor. However, the presence of the dominated F force is nonetheless felt. Because intention contributes an attractor at the intended [d] form, the attractor for the combined system is a little closer to that of [d] than in the case when the intention is *Rat* 'advice'. This competition between the M, F forces is schematized in Fig. 8 for various values of intentional degree, the control parameter θ in the dynamics for intention $F(x) = \theta(x^{REQ} - x)$ representing the relative strength of the intentional contribution. A value of intent close to 0 corresponds to a context where the speaker's intention to communicate the contrast between *Rat* and *Rad* is weak, as
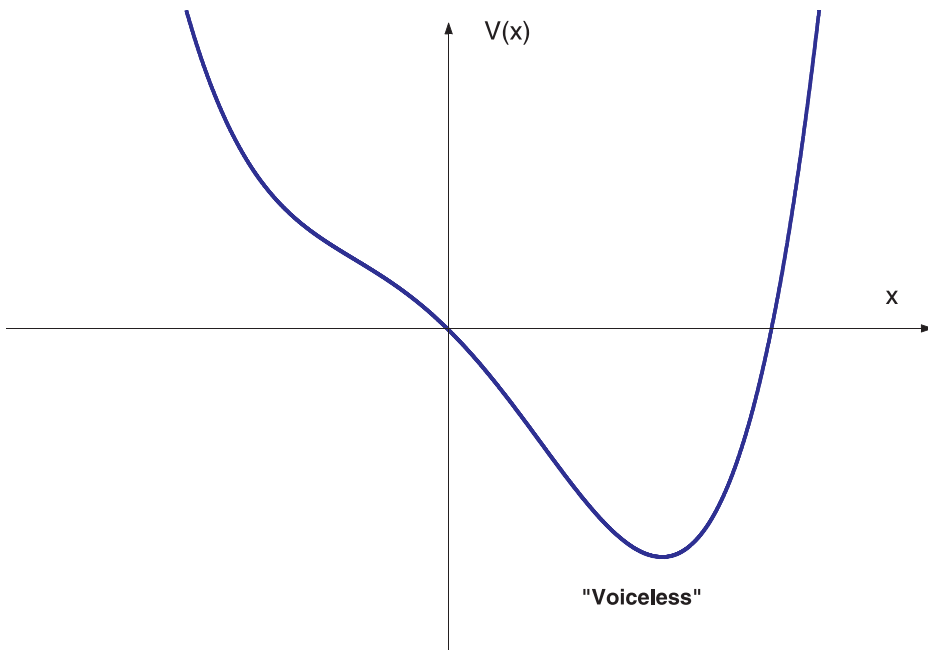
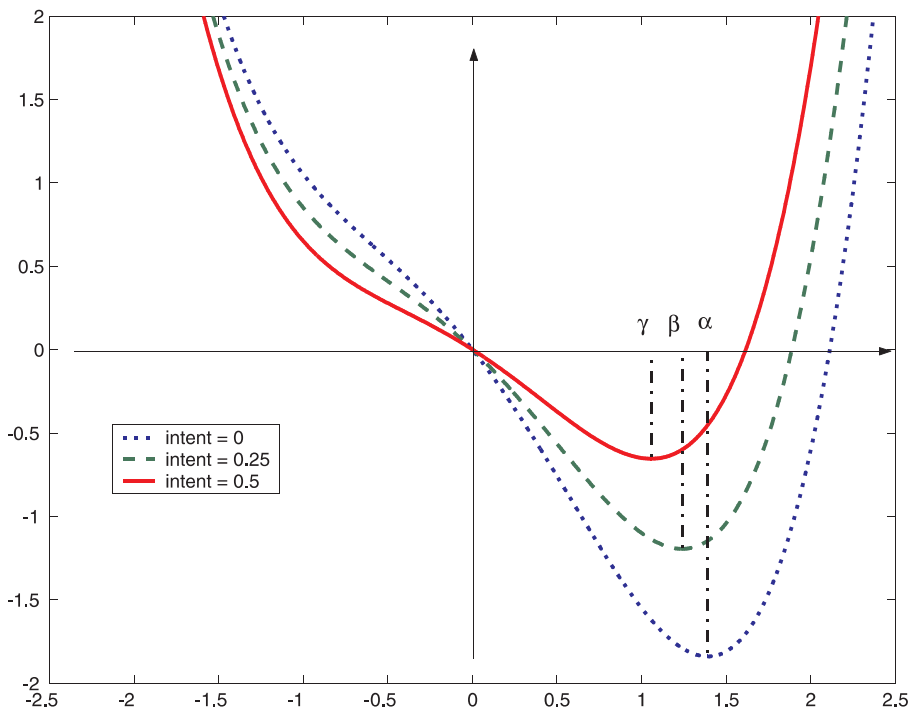Fig. 7. Dynamical model of a markedness constraint.

Fig. 8. Drifting of phonetic output for different intentional degrees.

would be the case in the word list reading, assistant-absent condition. Higher values correspond to communicative contexts with stronger requirements for expressing the contrast as would be the case in the assistant-present condition. It is observed that, as intentional strength increases, the potential gradually changes so that the attractor drifts toward the left or toward more voicing ($\alpha \rightarrow \beta \rightarrow \gamma$).

The drift in the attractor of the phonetic output is a reflex of constraint conflict between the M, F forces relative to different contextual conditions. This drift is beyond the scope of the standard optimality theory model where constraint conflict is at the symbolic level. When constraint interaction is trapped at the level of symbols, it can only yield the same categorical representation [ʁat] for *Rad*, *Rat*.

The dynamical formulation of constraint conflict maintains the discreteness of the symbolic formulation inspired by optimality theory while embedding it in a continuous realm. The discreteness inheres in the presence of distinct constraints (M, F) and in the attractors of the dynamical systems defining these. Crucially, these attractors are embedded in continuous phonetic dimensions. Consequently they can be flexibly and gradiently modulated by variation in context in a way that derives the subtle differences in *Rad* versus *Rat* and their dependency on the communicative context.

At a more broad level, the main proposal here is that it is both necessary and promising to do away with the metaphor of precedence between the qualitative phonology and the quantitative phonetics, without losing sight of the essential distinction between the two. The end result is a "symbiosis of the symbolic and subsymbolic paradigms: The symbolic paradigm offers concepts for better understanding subsymbolic models, and those concepts are in turn illuminated with a fresh light by the subsymbolic paradigm" (Smolensky, 1988, p. 19). Attaining this requires a coherent way to make discreteness and continuity coexist within the same formal language. The mathematics of nonlinear dynamics satisfies this requirement.

5. Dynamical links between categorical alternations and continuous phonetics

We now turn to the problem of how low-level spatial phonetic properties of vowels relate to the high-level phonological behavior of suffix choice in the phenomenon of vowel harmony. Although this problem derives from a rather different phonological phenomenon from that of consonant devoicing, discussed in the previous section, the two share a common formal challenge: relating continuous phonetic variables to discrete phonological patterns. As in the previous section, the proposed dynamical model formalizes the link between the discreteness of phonological form and the continuity of phonetic substance in which that form is embedded. The emerging picture is that the interfaces between phonology and phonetics may be many and variegated (Ladefoged, 1988) but the mathematical principles remain invariant across disparate domains.

5.1. Vowel harmony and transparency

Vowel harmony is a systematic regularity found in many languages requiring vowels in certain grammatical domains to agree in terms of specific phonetic properties. For example, the

vowel set of Hungarian can be divided into two subsets for the purposes of vowel harmony: the subset of front vowels /i í e é ö ő ü ű/ articulated with a frontward movement of the tongue body and transcribed in IPA as [i iː ɛ eː ø øː y yː] and the subset of back vowels /u ú o ó a á/ articulated with a backward movement of the tongue body and transcribed as [u uː o oː ɔ ɑː] (Siptár & Törkenczy, 2000, Vago, 1980). In terms of phonological features, the front vowels share the feature [−back], and the back vowels share the feature [+back]. The phonological consequences of vowel harmony are most readily observed in suffix vowel alternations where the [±back] quality of the suffix vowel is determined by the [±back] quality of the stem vowel. For example, the Dative suffix alternates between two forms, one with the front /e/ and another with the back /a/, as a function of the stem vowel: *ház-nak* 'house-Dative', *város-nak* 'city-Dative', but *kéz-nek* 'hand-Dative', *öröm-nek* 'joy-Dative'.[7] Because it determines the suffix form, the first stem vowel is called the trigger and the suffix vowel is called the target of the harmony pattern.

A central aim of phonological theory is to characterize the range and possible forms of such rule-like patterns in different languages (Chomsky & Halle, 1968/1991). A particularly fruitful research strategy in this regard has been to explore the *phonetic basis* of sound patterns; that is, the extent to which such patterns can be seen as adaptations to biological constraints on speech production and perception (Lindblom, 1983; cf. Anderson, 1981). For vowel harmony, specifically, it has been proposed that a natural basis for it can be traced to the low-level phonetic influences among vowels in consecutive syllables (Fowler, 1983; Ohala, 1994). The crucial fact is that vowels exert influences on neighboring vowels across intervening consonants, the so-called *V-to-V coarticulation* (Öhman, 1966). However, V-to-V coarticulation is a quantitative pattern whose degree varies depending on the specific language, quality of intervening consonants, stress distribution, and other factors (e.g., Recasens, 1999). It thus remains to be shown how such variable and quantitative coarticulation effects are to be linked to the binary [±back] character of suffix alternations. As we have emphasized all along, addressing this challenge requires the appropriate formal tools for relating discrete and continuous aspects of complex systems.

A second problem with the proposed phonetic basis of vowel harmony is that many languages with harmony include vowels that disagree with their adjacent vowels. These vowels are called *transparent* because their most well-known property is that they may intervene between the trigger and the target vowel even when they bear the opposite value for the harmonizing feature. For example, in Hungarian, *papír* selects [+back] suffixes, such as *nak* 'Dative', *ház* 'Allative', *tól* 'Ablative', *ban* 'Inessive', in agreement with the [+back] value of the initial stem vowel and despite the intervening [−back] value of /í/. The rest of the Hungarian front unround vowels, that is, /i, é, e/, behave similarly. Here are some representative examples of words with these vowels: *gumi-nak* 'rubber-Dative', *kávé-nak* 'coffee-Dative', and *hárem-nak* 'harem-Dative'. Transparent vowels, then, present a challenge to the proposal that vowel harmony has its basis in V-to-V coarticulation effects between consecutive vowels. This is because the [+back] quality of the suffixes in the preceding words cannot plausibly be derived via V-to-V coarticulation when their preceding vowels are specified as [−back].

At the heart of the problem that transparent vowels pose for the phonetic basis of vowel harmony is an implicit assumption about their representation. This assumption is that the phonological category of a transparent vowel is invariant across different contexts and irrelevant to

the quality of the suffix following the transparent vowel. In an impressionistic sense, the transparent vowels in words like bul**i**-nak 'party-Dative', *hĭd-nak* 'bridge-Dative' or *mamicsi-nak* 'mother-Diminutive-Dative' are not perceptually different from those in *bili-nek* 'pot-Dative' or *vĭz-nek* 'water-Dative'. Hence, they are assumed to be invariant across these different contexts. However, it is well known that for vowels a relatively stable acoustic output can be produced using multiple articulatory strategies and construction locations. Independent work by Stevens (1972, 1989) using simple tubes, and Wood (1979) using natural human vocal tract profiles, has shown that the acoustic outputs for nonlow front vowels—exactly the transparent vowels of languages like Hungarian and Finnish—are insensitive to a limited amount of variation in the horizontal position of the tongue body. Therefore, the impressionistic perceptual invariance of transparent vowels in different contexts does not necessarily imply their articulatory invariance.

Our hypothesis, pursued in experimental work (Benus, Gafos, & Goldstein, 2004; Gafos & Benus, 2003), is that transparency emerges from nonlinearities in the relation between articulation and sound. In a nutshell, we hypothesize that the /ĭ/ in *zafĭr-ban* 'sapphire-Inessive' is retracted articulatorily as compared to /ĭ/ in *zefĭr-ben* 'zephyr-Inessive', but that this retraction falls within that limited region of articulatory variation that does not result in any significant acoustic consequences. If this hypothesis is correct it would provide a basis for a principled understanding of the cooccurrence of two properties of the phenomenon, the nature of the harmonizing parameter (tongue body retraction) and the set of transparent vowels in Hungarian (/ĭ, i, é, e/).

The hypothesis that transparent vowels do participate in vowel harmony by subcategorical changes in their tongue body position may help understand other heretofore recalcitrant generalizations in the Hungarian vowel harmony system (Benus, 2005). First, stems with only transparent vowels (T stems) may trigger both front and back suffixes (Vago, 1980). The majority of T stems trigger front suffixes (*cĭm-nek* 'address-Dative', *szél-nek* 'wind-Dative'), but approximately 60 T stems trigger back suffixes (*sĭp-nak* 'whistle-Dative', *cél-nak* 'aim-Dative'). From the perspective of categorical representations, this situation is paradoxical. Because the vowels in *szél, cél* are represented identically, they are expected to select the same suffix, but they do not. However, if there are systematic phonetic differences in tongue body position between these two stem groups, they may provide a basis for their divergent suffix choices.

Second, the height of the transparent vowels in stems where they are preceded by a back vowel (BT stems) affects the choice of the suffix (Hayes, 2004; van der Hulst, 1988). The lower the vowel, the more likely it is that the suffix following BT stems is front. For example, Hungarian speakers accept only back suffixes after *papĭr* but allow free variation between a front and a back suffix after *hotel*. This latter pattern is known as *vacillation* in the Hungarian literature. What is the difference between /ĭ/ and /e/ that may be responsible for their distinct suffix selection patterns? An answer might come from differences in the articulatory-acoustic properties of high vowels such as /ĭ/ versus lower vowels such as /e/. Specifically, the lower and more retracted tongue body of /e/ reduces the region of acoustic insensitivity to articulatory variation, thus predicting that /e/ cannot be retracted to a degree comparable to that of /ĭ/.

Finally, increasing the number of transparent vowels following a back vowel decreases the likelihood of selecting back suffixes. For example, *mam-i, mam-csi,* both meaning 'mother-Diminutive', select back suffixes as in *mam-i-nak, mam-csi-nak*

'mother-Diminutive-Dative'. However, when *-i* and *-csi* are combined as in *mam-i-csi*, both front and back suffixes are acceptable, *mamicsi-nak* and *mamicsi-nek*. Compare also *kabin-nak* 'cabin-Dative' versus *aszpirin-nak/nek* 'aspirin-Dative' and *Acél-nak* 'Acél-Dative' versus *Acélék-nek/nak* 'Acél-Collective-Dative' (Farkas & Beddor, 1987; Hayes, 2004; Kaun, 1995; Ringen & Kontra, 1989). Under the assumption that all vowels, including transparent vowels, participate in vowel harmony, the tongue body position for the stem-final vowel in stems like *mami* is predicted to be different from that of the stem-final vowel in *mamicsi*. The vowel /i/ in *mami* is less advanced or more retracted than the stem-final /i/ in *mamicsi* because the additional front vowel in the second syllable of *mamicsi* eliminates partially the influence of the initial back vowel on the final /i/.

In short, our hypothesis predicts that transparent vowels show systematic articulatory differences in different contexts and that these differences are linked to the harmonic behavior of these vowels. In the next section, we present the results of our experiments on the articulatory properties of transparent vowels.

5.2. *Articulatory characteristics of transparent vowels*

To examine the articulation of transparent vowels in vowel harmony, we used two experimental techniques, electromagnetic articulometry and ultrasound imaging. Before a presentation of our results, we give a brief overview of these techniques.

In electromagnetic articulometry (EMMA; Perkell et al., 1992), an electromagnetic field is used to track movements of small receiver coils attached to the speech articulators. Three transmitter coils are fixed on a plastic apparatus surrounding the speaker's head. The transmitter coils produce alternating magnetic fields at different frequencies in the range of about 10 KHz. Small receivers, about 2 mm in diameter, are attached on the speech articulators using a special adhesive. The electromagnetic field from the transmitter coils passes through the receiver coils and generates an electric signal. The voltage of this signal is inversely related to the distance of the receiver relative to the transmitter coils. This relation is used to calculate the position of the receivers as a function of time. The voltages in the receiver coils are captured at a sampling rate of 500 Hz. Due to their high sampling rate, EMMA systems are currently the most accurate tools for the collection of real-time lingual articulatory movement data (Tiede, Vatikiotis-Bateson, Hoole, & Yehia, 1999).

A representative recording from the EMMA system is shown in Fig. 9. This figure shows one instance of a sentence containing the target word *zafírban*. The signals from top to bottom represent the acoustic waveform of the entire sentence, the acoustic waveform of the target word with an approximate segmentation into phonemes, and the vertical (solid curve) and horizontal (dashed curve) position of the receivers attached on the tongue tip (TT), body (TB), dorsum (TD), and upper and lower lips (UL, LL). The axis at the bottom depicts time in milliseconds. As the tongue body smoothly moves from vowel to vowel in the sequence *a-í-a* of *zafírban*, the receivers on the TB and TD can be seen to trace a bell-shaped trajectory from a retracted position for /a/, with low horizontal values of TB and TD, to an advanced position for /í/, with higher horizontal positions for TB and TD, and back to a retracted position of the final /a/. To quantify the spatial properties of transparent vowels, here the /í/ in *zafírban,* we identified the maximal horizontal positions of the TB and TD receivers during that vowel. These are
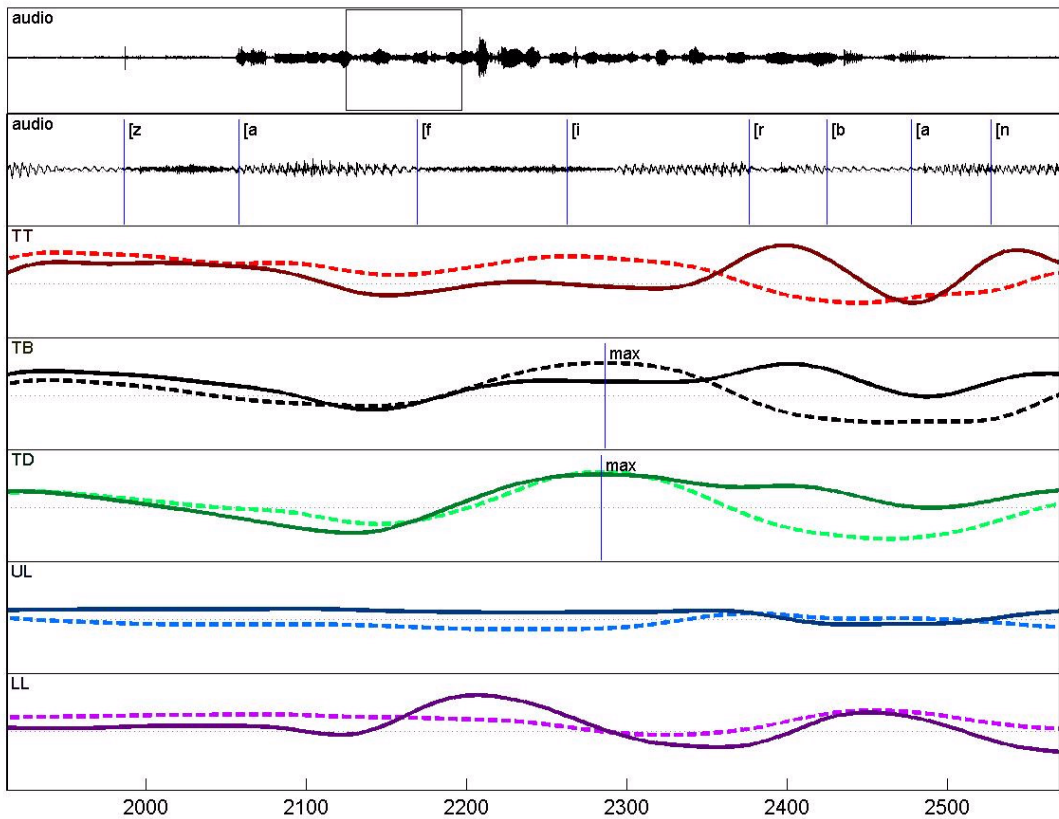
Fig. 9. Articulator kinematics recorded with EMMA.

shown by the max labels in Fig. 9, corresponding to the peaks in the horizontal trajectories of TB and TD. At these time points the TB and TD receivers are at their most advanced position.

The other experimental method is ultrasound (Stone, 1997). In ultrasound, a probe with a pi-ezoelectric crystal is placed below the participant's chin, emitting ultrahigh frequency waves. These waves travel through soft tissue and reflect when they reach an interface with a matter of different density such as air. The reflected echo is used to construct a bright white line that shows the boundary between the tongue surface and the air above it. In our experiments, ultra-sound images of the tongue were collected at a 30 Hz rate, video recorded, and then digitized. The spatiotemporal resolution of ultrasound is low relative to that of EMMA. The advantage of ultrasound is that it allows visualization of the back region of the tongue's surface from the root to the dorsum. This region is crucially involved in vowel production, but it is usually inaccessi-ble with EMMA due to difficulties with participants tolerating a receiver attached to that part of the tongue.

To estimate tongue shapes, we first identified the frame with the most advanced tongue posi-tion during the production of the transparent vowel (left, Fig. 10). The tongue edge in this frame was then traced using methods developed by Iskarous (2005) that determine the points of maximal contrast within the selected region (middle, Fig. 10) and fit multiple snakes into a
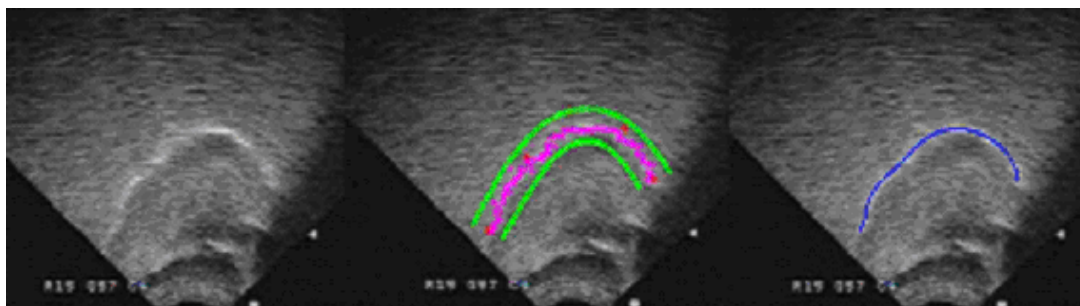
Fig. 10. Tongue surface captured with ultrasound and estimation of tongue contours.

curve that balances the distance of the points from the curve with the curve's smoothness (right, Fig. 10).

We constructed two sets of stimuli consisting of word pairs where transparent vowels occur in stems triggering front or back suffixes. In the first set, all words were trisyllabic. An example pair from this set is *zefír-ben* 'zephyr-Inessive' / *zafír-ban* 'sapphire-Inessive', where /í/ occurs in *zefír*, triggering a front suffix, and in *zafír*, triggering a back suffix. Such pairs allow us to compare the tongue posture for /í/ in the two vowel harmony contexts, front and back. The second set of stimuli consisted of monosyllabic words. For example, /é/ in *szél* 'wind' was compared to /é/ in *cél* 'aim'. The forms *szél*, *cél* correspond to the nominative case of the respective nouns, where there is no overt suffix. In other cases with overt suffixes, *szél* triggers a front and *cél* triggers a back suffix: *szél-nek* 'wind-Dative' versus *cél-nak* 'aim-Dative'. Once again, such pairs allow us to compare the tongue posture for /é/ in the two vowel harmony contexts, front and back. However, they potentially allow us to establish another point. In pairs like *zefír* versus *zafír*, the difference in suffix choice, front for *zefír* versus back for *zafír*, is typically ascribed to the presence of a front versus back stem-initial vowel. In our *szél* versus *cél* pairs of stimuli, if systematic subcategorical differences are found in the transparent vowel, then the distinct suffix choices in *szél-nek* versus *cél-nak* should be related to those subcategorical differences.

EMMA data from 3 participants and ultrasound data from 1 participant were analyzed. Table 1 shows the data from the EMMA experiments. For the two stimuli sets, trisyllabic (3-syll) and monosyllabic (1-syll), the rows show the mean receiver positions on the tongue tip (TT), body (TB), and dorsum (TD) in the front (F) and back (B) harmony context as well as their mean difference (MD = F – B). The absolute value of MD corresponds to the size of the effect, and its sign shows the direction of the effect. If the MD value is positive, the relevant receiver in the back environment is retracted relative to its position in the front environment.[8]

All MD cells contain positive values except for one cell in CK's TT monosyllabic data. TT is the receiver attached at the most anterior part of the tongue. Thus, its MD values are least relevant to the posture of the tongue body and dorsum, the main determinants of vowel quality (Harshman, Ladefoged, & Goldstein, 1977, p. 702). Moreover, for that same participant, the TD value in the monosyllabic data is positive and also shows a significant mean difference between F and B contexts (0.69). Hence, the main result is that the tongue is more advanced for transparent vowel in stems triggering front suffixes than for transparent vowels in stems triggering back suffixes. This effect is significant for 12 out of 16 MD cells.

Table 1
Direction and size of the effect of harmonic environment in EMMA data

	Subject ZZ			Subject BU			Subject CK		
Rec.	F	B	MD	F	B	MD	F	B	MD
Trisyllabic (3-syll)									
TD	–48.02	–48.97	0.95**	–43.12	–43.51	0.39**	–24.59	–25.58	0.99*
TB	–38.65	–40.05	1.40**	–30.89	–31.48	0.59**	—	—	—
TT	–23.41	–24.73	1.32**	–21.68	–22.07	0.39**	–21.83	–22.08	0.23
Monosyllabic (1-syll)									
TD	–46.67	–46.93	0.26	–42.08	–42.61	0.53**	–22.25	–22.94	0.69*
TB	–36.17	–36.81	0.64*	–29.54	–30.38	0.84**	—	—	—
TT	–20.35	–20.62	0.27	–20.09	–20.6	0.51**	–20.00	–19.78	–0.22

Note. F = front; B = back; MD = mean difference; TT = tongue tip; TB = body; TD = dorsum. The values under F, B are negative because the origin of the coordinate system is approximately at the subject's upper incisors with receiver positions farther inside the mouth represented with progressively decreasing values.

The same effect is observed with the ultrasound data. In Fig. 11, we have superimposed tongue postures of /í/, /é/ in a back (dotted lines) versus front harmony (solid lines) context from participant ZZ. The direction of the effect is the same as that observed with the EMMA data. Transparent vowels in the front context are more advanced than those in the back context.

We observed differences in effect size between the two methodologies. For example, in ultrasound, averaged differences in tongue position between the two harmonic contexts reach up to 2.5 mm. In EMMA, the maximal average difference was 1.4 mm. Such differences in magnitude are at least in part due to the fact that ultrasound allows access to almost the entire surface of the tongue as opposed to the position of a few flesh points with EMMA. Details of the
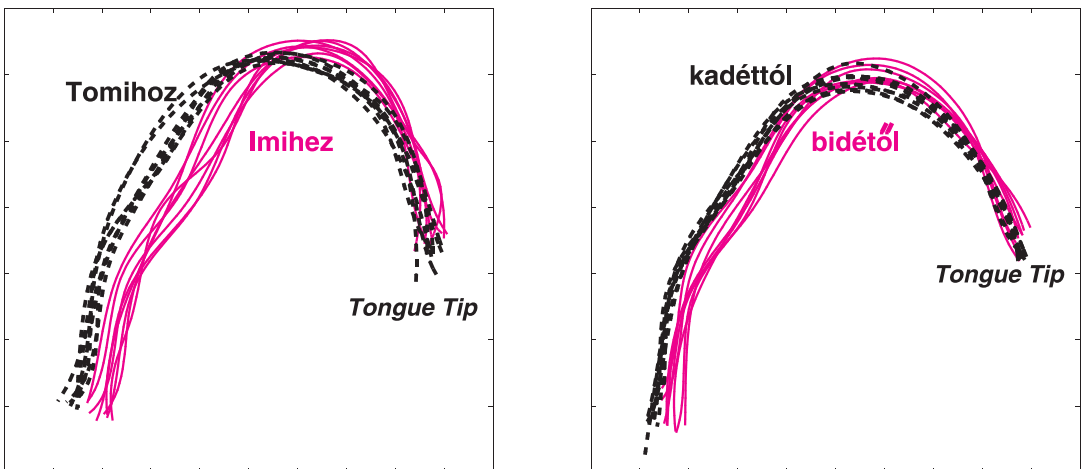

Fig. 11. Effect of environment on /í/ (left) and /é/ (right).

ultrasound data and their relation to the EMMA data are not discussed due to space limitations (see Benus, 2005).

5.3. Synthesis: Dynamical links between continuity and discreteness

The main generalization from our study is that continuous differences in tongue body posture of stem-final vowels are linked to alternations in discrete suffix form. A more advanced transparent vowel selects a front suffix and a less advanced one selects a back suffix. We now propose a model that derives this generalization. Our presentation proceeds in two steps. To capture differences in degree of tongue position for stem-final vowels, we first model harmony between stem vowels as perturbations of consecutive vowel gestures due to coarticulation. These perturbations effect differences in the spatial targets of stem-final vowels in different environments. We then model the relation between these continuous differences and the binary ([±back]) nature of the suffix alternations.

In the proposed model of stem-internal harmony, articulatory gestures are formalized using point-attractor dynamics. As discussed earlier, a gesture has both spatial and temporal dimensions. Because in vowel harmony the combinatorial patterns on vowels are described in terms of the spatial dimension of construction location (CL), we focus on CL here. For any given gesture, we model the dynamics of CL using a monostable potential of the form $V(x) = \alpha(x - x_0)^2$, where α expresses the strength with which a given gesture imposes its control over the tongue body articulator and x_0 represents the target CL value. For exposition, we choose $CL = 2$ as a representative value for front or [–back] vowels and $CL = -2$, a constriction farther inside the vocal tract, as a representative value for [+back] vowels. We can visualize the movement of the tongue body articulator toward its target location by a particle left in the potentials of Fig. 12.
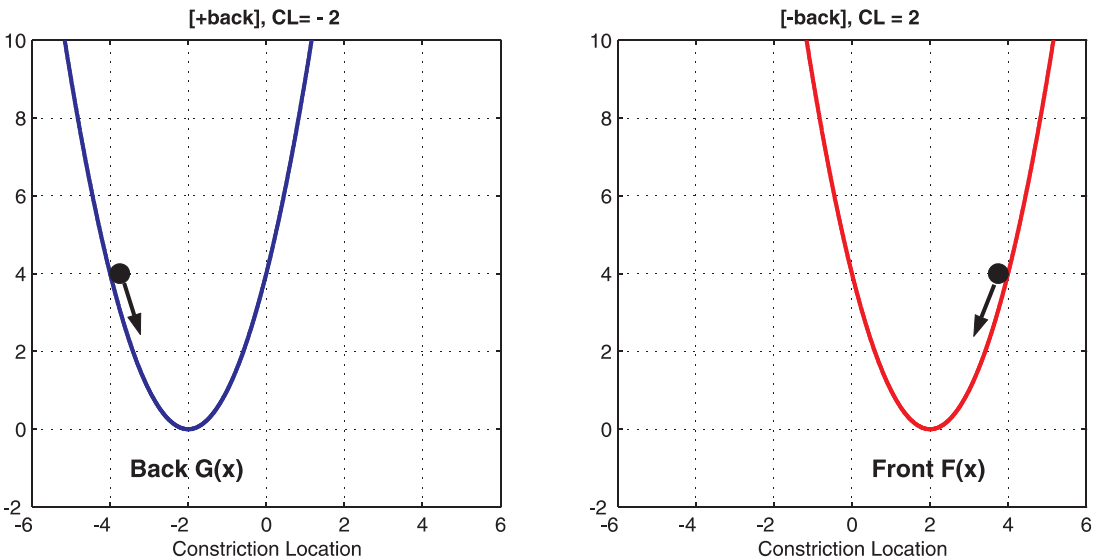

Fig. 12. Dynamics for constriction location.

As first shown in Öhman (1966), due to coarticulation, vowels within a word exert spatial influences on each other even when they are separated by consonants (see also Fowler, 1983). For example, when a back vowel is followed by a front vowel, as in the sequence /a–i/, the tongue body posture of the front vowel is relatively retracted due to the demands on the tongue body from the overlapping back vowel. Such spatial influences among gestures are known as blending (see also section 3). Blending of vowel gestures in consecutive syllables is a core component of our vowel harmony model. Consider two vowel gestures in adjacent syllables as in the stem *papír*. The first is a back vowel with the potential $V_1(x) = \alpha(x + 2)^2$, and the second is a front vowel with the potential $V_2(x) = \beta(x - 2)^2$. We focus here on the effects of blending on the second or stem-final vowel.[9] The simplest hypothesis for formalizing the perturbation of the second gesture due to the presence of the first vowel is to take the linear combination of the unperturbed potentials where both gestures contribute equally to the blended output for the second gesture ($\alpha = \beta$). The potential in the left panel of Fig. 13, marked with the dashed line, represents the perturbed front vowel gesture under this blending hypothesis.

It can be seen that if $\alpha = \beta$, the attractor of the resulting potential would be at the midpoint between the FRONT and the BACK attractors. However, for stems like *papír*, our experiments show that the second vowel is a slightly retracted version of a front vowel rather than a vowel with a tongue body posture midway between a front and back vowel. Therefore, a minimal extension of the blending function is required so that the two potentials influencing the blended output of the second gesture carry different weights. This is a reasonable extension because our aim is to model the perturbation of the second gesture due to the first gesture. As the second gesture is produced, the strength of the control imposed on the tongue by the first gesture fades away and that of the second gesture increases. A simplified but adequate way of capturing this in our blending model is to increase the weight of the second gesture relative to that of the first.
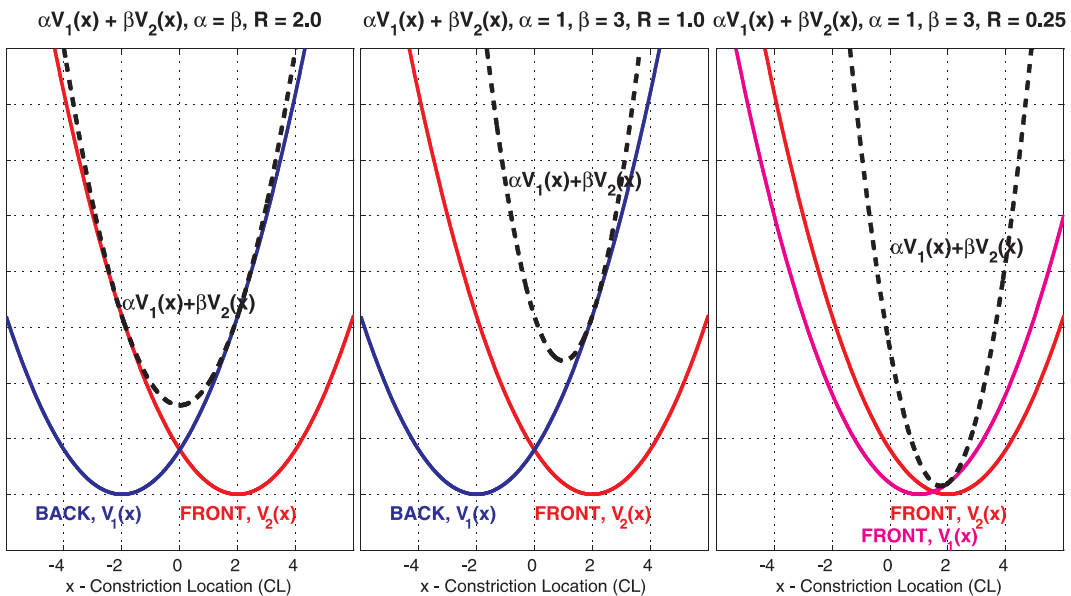

Fig. 13. Blending of two gestures represented with potentials $V_1(x)$ and $V_2(x)$.

This is shown in the middle panel of Fig. 13, where now the gesture of the second front vowel is weighted more than that of the first back vowel ($\alpha = 1$, $\beta = 3$). Consequently, the result of blending, the potential $\alpha V_1(x) + \beta V_2(x)$ shown with the dashed line, has its minimum tilted more toward the attractor of the second front gesture. The right panel of Fig. 13 shows the result of blending on the second or stem-final vowel for a stem with two front vowels such as *emir* 'emir'. In this example, the first vowel is slightly less advanced ($CL = 1$) than the second one ($CL = 2$). Given that both V_1 and V_2 are front vowels, the difference between their respective attractors is small and the displacement of V_2 from its canonical horizontal position is minimal.

Gestural blending effects perturbations in vocalic tongue body position of stem-final vowels; for example, the /í/ in *papír* is retracted to some degree because it follows a back vowel. The degree to which a vowel is retracted is captured by the parameter R, the difference between the attractor position of the vowel gesture after blending and its attractor position before blending. Our experiments indicate that fine differences in articulatory retraction of transparent vowels are linked to the discrete alternation in suffix form. A more advanced transparent vowel selects a front suffix and a less advanced one selects a back suffix. How can small differences in articulation be related to a categorical ([±back]) alternation in suffixes? Informally, the relation obtained between degree of advancement of the stem-final vowel and suffix selection is nonlinear. Small changes in the former can cause large (nonlinear) changes in the latter. We now formalize this idea using nonlinear dynamics.

The core fact in our case of vowel harmony is that suffix vowels alternate between a front and a back version. In the proposed model for suffix selection, the two discrete forms of an alternating suffix (e.g., Dative -*nak* vs. -*nek*) should correspond to the attractors of a dynamical system. To model the dependence between the continuous parameter of retraction degree R of the stem-final vowel and the discrete form of the suffix, we require that the choice of the suffix attractor be modulated by variation in the control parameter R. Following the discussion in section 2, these ideas can be stated in the form of the equation $\dot{x} = f(x,R) + Noise$.

Our goal now is to determine a good candidate for the function $f(x,R)$. A proposed dynamical model of some phenomenon is a good model to the extent that aspects of the phenomenon in question correspond well with qualitative properties of its mathematical formulation. An appropriate dynamical system for the suffix alternation is required to have a bistable potential to capture the presence of two stable forms of a suffix, front and back. A polynomial of order less than three allows for at most one attractor (Arnold, 2000). Hence, the simplest model for suffix choice can be specified by a cubic polynomial. A good candidate for $f(x,R)$ is the tilted anharmonic oscillator whose dynamics are described by $f(x,R) = \lambda R + x - x^3$ where λ is a factor linearly proportional to R (see also Tuller et al., 1994). Because $\dot{x} = f(x,R) = -dV(x)/dx$, we can compute the potential $V(x) = -\lambda Rx - x^2/2 + x^4/4$ by integrating $f(x,R)$. Using concepts from section 2, the value of the CL for a suffix vowel is interpreted by the position of a particle running downhill in this potential and the asymptotic behavior of x in this equation can be visualized in Fig. 14. In these plots, the control parameter R varies between 0.2 and 1.2, corresponding to minimal and maximal retraction, respectively.

The graph on the left shows the behavior of the system for stems like *Tomi*, which take back suffixes. We observed experimentally that in such stems the transparent vowel is retracted. In our model, retraction enters the dynamics via R. The function $f(x,R)$ for a range of R values, $R \approx$
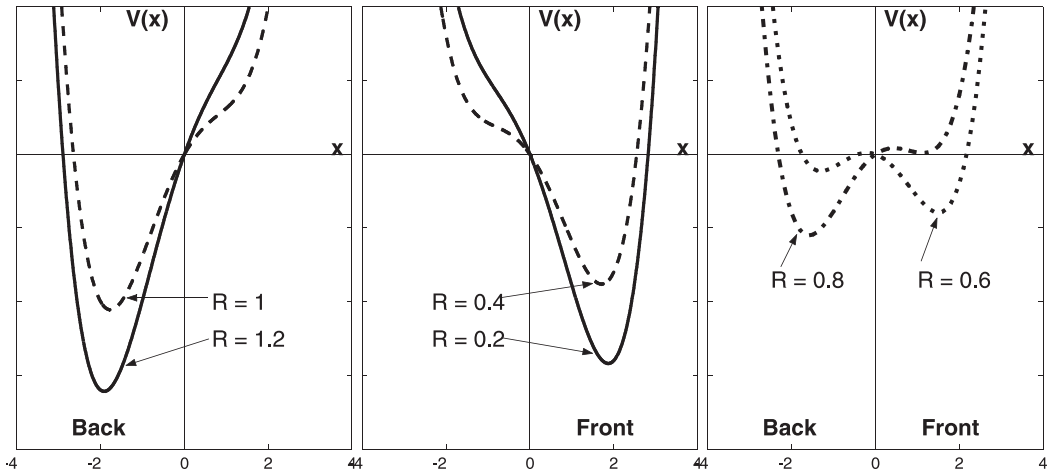
Fig. 14. Suffix form as a function of retraction degree R.

1 (significant retraction), provides a potential $V(x)$ with an attractor close to the value of $CL =$ –2 (BACK), corresponding to the back variant of the suffix. The probability that a particle left in this potential ends up in the vicinity of the BACK attractor is very high. Because the position of the particle represents the [±back] form of the suffix, it is predicted that the suffix is back, (e.g., *Tomi-hoz*).

The graph in the middle panel shows how the potential $V(x)$ changes for stems whose final vowels show minimal or no retraction like *emir*. For minimal retraction, modeled as $R \approx 0.2$, a qualitative change is evident in the shape of $V(x)$. The BACK attractor has been replaced by a FRONT attractor. A stem with minimal retraction of its final vowel is thus predicted to select front suffixes (e.g., *emir-hez*).

The rightmost graph in Fig. 14 shows the behavior of the system for intermediate values of the control parameter R ($R \approx 0.7$). In nonlinear dynamics, a change from one macroscopic state of the system to another implies an intermediate stage of fluctuation. We see that there are now two minima representing the presence of two stable states, FRONT and BACK. To see the consequences of this for suffix choice, we must take into account the effects of noise and initial position of the particle. For example, consider a particle at a position around (0, 0) in any of the potentials of the right panel. Due to the random kicks introduced by the fluctuations, the particle ends up in either the FRONT or the BACK attractor. We illustrated this with analytical and simulation tools described in section 2. Thus, for intermediate R values our model predicts that the suffix can vary between a front and a back version.

Before turning to the specifics of this prediction, consider what Fig. 14 tells us about the relation between the control parameter R and the order parameter of suffix quality. Equal changes in the control parameter do not always effect comparable changes in the order parameter. For example, both $R = 1.2$ and $R = 1$ result in qualitatively the same potential with the single back attractor, albeit with different stability. But as R changes from $R = 1$ to $R = 0.8$ or from $R =$ 0.4 to $R = 0.6$ the potential changes qualitatively from a monostable regime to a bistable regime. Hence, a change of R by 0.2 leaves the qualitative form of the system unaltered within a

certain region of the control parameter values. However, within a different region of control parameter values, a change of the same magnitude causes a qualitative change in the behavior of the system. This is another illustration of the fundamental property of nonlinearity.

We now turn to instantiate the prediction on vacillation in suffix selection with Hungarian data. As discussed earlier, we find two sources of vacillation in Hungarian: Be stems where a back vowel is followed by the low /e/ (e.g., *hotel*) and BTT stems, where a back vowel is followed by two transparent vowels (e.g., *aszpirin*, *mamicsi*). For Be stems, Benus (2005) argued that the low and relatively less front tongue body posture for /e/ allows for only limited retraction when /e/ blends with a preceding back vowel. Effectively, /e/ can be retracted less than the other transparent vowels /i í é/. This is arguably related to differences in the articulation-to-acoustics mapping between these vowels (see Benus, 2005). In effect, there is a plausible phonetic basis for the differences in suffix selection between Be stems and stems where the back vowel is followed by some other higher transparent vowel.

We concentrate here on the explanation that our model provides for the robust generalization that BTT stems such as *mamicsi* are more likely to select front suffixes than BT stems such as *mami*. In our model, greater tongue body advancement (i.e., smaller R values) of the stem-final vowel corresponds to greater probability of selecting front suffixes. If the final vowel in stems like *mamicsi* can be shown to be less retracted compared to the final vowel in stems like *Tomi,* this prediction of the model would be confirmed.

Fig. 15 shows how our model derives the retraction degree of the final vowel in BTT stems such as *mamicsi.* Because in /a-i-i/ there are two pairs of adjacent vowels, our figure shows two panels corresponding to the two blending sites. On the left panel, we show the blending between /a/ and the following /i/, with /i/'s potential after blending shown with the dashed line. We see that the blended gesture is retracted compared to the canonical /i/ gesture. Its attractor is at $CL = 1$, with a retraction degree $R_i = 1$.

If the perturbed /i/ vowel from the first blending were the stem-final vowel, then for its R value our model for suffix selection would yield a suffix with a back vowel. As a result, a stem like *mami* is predicted to select back suffixes, consistent with the data. However, the /i/ in the second syllable of *mamicsi* is followed by another vowel. The right panel of Fig. 15 shows the second blending between the first retracted /i/, the output of the first blending, and the final /i/ vowel. This blending yields the potential shown with the dotted line whose attractor is at $CL = 1.4$. The degree of retraction of the stem-final vowel in *mamicsi* is thus $R = 0.6 (2.0 - 1.4)$. The main point is that, effectively, the second transparent vowel is less retracted than the first ($R_{T1} = 1.0$, $R_{T2} = 0.6$). As shown in Fig. 14 the potential $V(x)$ for the suffix vowel following a stem-final vowel with $R = 0.6$ is bistable, but with a bias toward the FRONT attractor. The probability density function corresponding to this potential predicts that the probability of x being in a region around the FRONT attractor is higher than the probability of x being around the BACK attractor. This was illustrated with simulations in Fig. 4. Hence, the difference in the potentials for $R = 0.6$ and $R = 1$ translates into a difference in suffix selection between BTT and BT stems. The former are more likely to select front suffixes than the latter.

We thus see that by linking suffix selection with subcategorical features of transparent vowels, the proposed model derives the difference between BT and BTT stems. All stem vowels participate in harmony because all vowel gestures undergo blending. Because there are more front vowels in BTT than in BT stems, the stem-final vowel in BTT stems is less retracted than in BT stems.
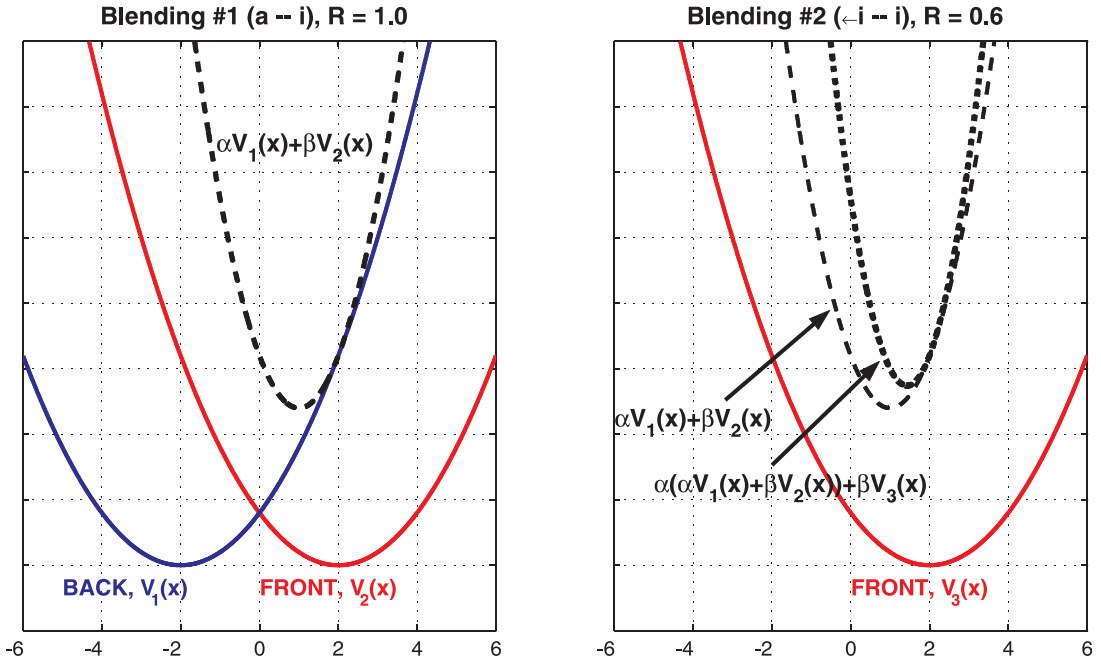
Fig. 15. Stem-internal blending in BTT stems.

In the proposed model, this quantitative difference in retraction corresponds to qualitatively different suffix choices. We plan to test the empirically predicted values of retraction in BTT versus BT stems in a future study. These differences will in turn allow us to fine-tune quantitatively the values of *R* generating a monostable versus a bistable potential (see Fig. 14).

To sum up, the starting point of this section was that small changes in tongue body CL of Hungarian transparent vowels are related to qualitative suffix alternations. This is the property of nonlinearity, a hallmark of complexity in natural systems in general and spoken language in particular. We presented a model that allows one to relate continuous phonetic distinctions to discrete phonological form using the mathematics of nonlinear dynamics. Our model accounts for the patterns of suffix selection in stems like *Tomi* (back suffix) versus *Imi* (front suffix), and makes plausible predictions for the patterns of vacillation (*hotel-nak/nek, mamicsi-nak/nek*). Overall, the proposed model provides an explicit link between the quantitative and qualitative aspects of the relevant patterns and makes specific predictions leading to new experimental studies.

6. Conclusion

Whereas the phonetic signal is a complex of continuous information, the phonological structure of languages is best described as a qualitative system of abstract, symbol-like prop-

erties. What kinds of formal tools are best in dealing with this duality? At a broad level, the main proposal of this article is that a single formal language, nonlinear dynamics, makes it possible to model the relation between the discreteness of phonological form and the continuity of phonetic substance in which that form is embedded. The essential constructs of phonological cognition, the phonological representations and the grammar constraints on these, are formulated in dynamical terms. A corollary of our proposal is that it opens the way for a parallel or nonderivational way of relating phonology and phonetics. Crucially, this view does away with the problematic metaphor of implementation or precedence between phonology and phonetics without losing sight of the essential distinction between the two (qualitative, discrete vs. quantitative, continuous).

Notes

1. Smolensky (1988) made a similar point on the notion of symbol. In a connectionist-based architecture, dynamical entities such as mental states or schemas that play a role at the higher level of description are defined in terms of context-specific activity of continuous vectors. See Smolensky (1988, p. 17) and antecedent notions in Hofstadter (1979).
2. There is an extensive literature on the phonetics of incomplete neutralization. See, among others, Fourakis and Iverson (1984), Dinnsen and Charles-Luce (1984), Dinnsen (1985), and Blumstein (1991). See Kim and Jongman (1996) for a case of complete neutralization in Korean.
3. There is considerable debate on whether the feature change in the German neutralization facts should be expressed in terms of the features [Voiced], [Tense] or [Spread Glottis] (Iverson, 1997; Jessen, 1998; Kloeke, 1982; Kohler, 1984; Wiese, 1996). This issue is orthogonal to the argument made here. Assume that there is a right way of formalizing final devoicing as a feature-change rule. The crucial point for the forthcoming discussion is that this rule would be discrete in nature, changing one categorical representation to another.
4. The incompleteness of final devoicing has led to arguments for relaxing some of the key assumptions of the standard view on the phonology–phonetics relation (see Gafos, 2006). For example, an elaboration of the analysis discussed in the text would revise phonetic implementation so that it has access not just to the output of the phonological derivation but also to the underlying form. At the phonetic level, the voicing of a syllable-final consonant would be computed by a continuous interpolation between its voicing feature in the underlying form and its voicing feature in the output of the phonological computation. This could provide for the phonetic differences between underlying voiced and voiceless consonants. Although this revision would be a step in the right direction, it would also make the rule-based plus revised phonetic implementation analysis hard to distinguish from the proposal to be developed. Furthermore, such an elaboration would not work for our next case study in section 5. Crucially, however, the formal tools used therein are the same as in the proposal developed in this section (see Browman & Goldstein, 1995a, for relevant discussion of methodological issues with

elaborating the notion of phonetic implementation). Ultimately, however, our goal is to build on the insights of the symbolic view, rather than to argue that they are irrelevant, as discussed in this section.

5. See Elman (1995) for a connectionist-based dynamical approach to representations and rules in syntax.

6. The potential for the faithfulness constraint is a polynomial of the second order, but that of the markedness constraint is a polynomial of the fourth order. Whence this difference? By definition, faithfulness constraints prescribe a unique value in the state space of an order parameter. In the dynamical equivalent of this uniqueness, the force function has a single zero and the corresponding potential is monostable—there is a single (stable) fixed point and thus the polynomial must be of the second order. The dynamical formulation of markedness, on the other hand, encodes more complexity than the faithfulness constraint. In German, syllable onset consonants can be voiced or voiceless, so the potential in that environment must be bistable, with attractors at voicing values appropriate for voiced and voiceless consonants. This implies a force function of the third order or equivalently a potential of the fourth order. The potential for the markedness constraint used here is a special case of the onset potential, and it is derived from it as described in Gafos (2006).

7. In the description of harmony patterns we use Hungarian orthography where the acute accent denotes length, and the umlaut denotes front round vowels.

8. The values under F, B are negative because the origin of the coordinate system is approximately at the participant's upper incisors with receiver positions farther inside the mouth represented with progressively decreasing values.

9. See Benus (2005) on how this model can also capture bidirectional blending in vowel sequences.

Acknowledgments

References

Abraham, R. H., & Shaw, C. D. (1982). *Dynamics: The geometry of behavior.* Santa Cruz, CA: Aerial.
Anderson, S. R. (1974). *The organization of phonology.* New York: Academic.
Anderson, S. R. (1976). Nasal consonants and the internal structure of segments. *Language, 52,* 326–344.
Anderson, S. R. (1981). Why phonology isn't "natural". *Linguistic Inquiry, 12,* 493–539.

Arnold, V. I. (2000). Nombres d'Euler, de Bernoulli et de Springer pour les groupes de Coxeter et les espaces de morsification: Le calcul de serpents [Euler, Bernoulli and Springer numbers for Coxeter groups and morsification spaces: The calculus of snakes]. In É. Charpentier & N. Nikolski (Eds.), *Leçons de Mathématiques d'Aujourd'Hui* (pp. 61–98). Paris: Cassini.

Beckman, M. E., Edwards, J., & Fletcher, J. (1992). Prosodic structure and tempo in a sonority model of articulatory dynamics. In G. J. Docherty & D. R. Ladd (Eds.), *Papers in laboratory phonology II: Gesture, segment, prosody* (pp. 68–86). Cambridge, UK: Cambridge University Press.

Beckman, M. E., & Kingston, J. (1990). Introduction. In J. Kingston & M. E. Beckman (Eds.), *Papers in laboratory phonology I: Between the grammar and the physics of speech* (pp. 1–16). Cambridge, UK: Cambridge University Press.

Benus, S. (2005). *Dynamics and transparency in vowel harmony.* Unpublished doctoral dissertation, New York University, New York.

Benus, S., Gafos, A., & Goldstein, L. (2004). *Phonetics and phonology of transparent vowels in Hungarian.* In P. M. Nowak, C. Yoquelet, & D. Mortensen (Eds.), *Proceedings of the 29th annual meeting of the Berkeley Linguistic Society* (pp. 485–497). Berkeley, CA: Berkeley Linguistic Society.

Benus, S., Smorodinsky, I., & Gafos, A. (2004). Gestural coordination and the distribution of English "geminates." In S. Arunachalam & T. Scheffler (Eds.), *Proceedings of the Twenty-Seventh Penn Linguistics Colloquium* (pp. 33–46). Philadelphia: University of Pennsylvania.

Bloomfield, L. (1933). *Language.* New York: Holt, Rinehart & Winston.

Blumstein, S. E. (1991). The relation between phonetics and phonology. *Phonetica, 48,* 108–119.

Browman, C. P., & Goldstein, L. (1986). Towards an articulatory phonology. *Phonology Yearbook, 3,* 219–252.

Browman, C. P., & Goldstein, L. (1990). Tiers in articulatory phonology, with some implications for casual speech. In J. Kingston & M. Beckman (Eds.), *Papers in laboratory phonology I: Between the grammar and the physics of speech* (pp. 341–397). Cambridge, UK: Cambridge University Press.

Browman, C. P., & Goldstein, L. (1991). Gestural structures: Distinctiveness, phonological processes, and historical change. In I. G. Mattingly & M. Studdert-Kennedy (Eds.), *Modularity and the motor theory of speech perception* (pp. 313–338). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Browman, C. P., & Goldstein, L. (1995a). Dynamics and articulatory phonology. In R. F. Port & T. van Gelder (Eds.), *Mind as motion* (pp. 175–193). Cambridge, MA: MIT Press.

Browman C. P., & Goldstein, L. (1995b). Gestural syllable position effects in American English. In F. Bell-Berti & L. J. Raphael (Eds.), *Producing speech: Contemporary issues* (for Kathering Safford Harris) (pp. 19–33). Woodbury, NY: AIP.

Browman, C. P., & Goldstein, L. (2000). Competing constraints on intergestural coordination and self-organization of phonological structures. *Les Cahiers de l'ICP, Bulletin de la Communication Parlée, 5,* 25–34.

Burzio, L. (1994). *Principles of English stress.* Cambridge, UK: Cambridge University Press.

Byrd, D. (1996). Influences on articulatory timing in consonant sequences. *Journal of Phonetics, 24,* 209–244.

Byrd, D., & Saltzman, E. (1998). Intergestural dynamics of multiple phrase boundaries. *Journal of Phonetics, 26,* 173–199.

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

Chomsky, N., & Halle, M. (1991). *The sound pattern of English.* Cambridge, MA: MIT Press. (Original work published 1968)

Clements, N. G. (1985). The geometry of phonological features. *Phonology Yearbook, 2,* 225–252.

Cohn, A. C. (1990). *Phonetic and phonological rules of nasalization.* Unpublished doctoral dissertation, University of California, Los Angeles.

Coleman, J. (1992). The phonetic interpretation of headed phonological structures containing overlapping constituents. *Phonology, 9,* 1–44.

Davidson, L. (2003). *The atoms of phonological representation: Gestures, coordination, and perceptual features in consonant cluster phonotactics.* Unpublished doctoral dissertation, Johns Hopkins University, Baltimore.

Dinnsen, D. A. (1985). A re-examination of phonological neutralization. *Journal of Linguistics, 21,* 265–279.

Dinnsen, D. A., & Charles-Luce, J. (1984). Phonological neutralization, phonetic implementation and individual differences. *Journal of Phonetics, 12,* 49–60.

Draper, M., Ladefoged, P., & Whitteridge, D. (1960). Expiratory pressures and airflow during speech. *British Medical Journal, 1,* 1837–1843.

Elman, J. L. (1995). Language as a dynamical system. In R. F. Port & T. van Gelder (Eds.), *Mind as motion* (pp. 195–223). Cambridge, MA: MIT Press.

Farkas, D., & Beddor, P. (1987). Privative and equipollent backness in Hungarian. In A. Bosch, B. Need, & E. Schiller (Eds.), *23rd annual regional meeting of the Chicago Linguistics Society: Part 2. Parasession on autosegmental and metrical phonology* (pp. 90–105). Chicago: Chicago Linguistics Society.

Fodor, J. A. (1975). *The language of thought.* Cambridge, MA: Harvard University Press.

Fodor, J. A., & Pylyshyn, Z. W. (1981). How direct is visual perception? Some reflections on Gibson's "ecological approach." *Cognition, 9,* 139–196.

Fourakis, M., & Iverson, G. K. (1984). On the "incomplete neutralization" of German final obstruents. *Phonetica, 41,* 140–149.

Fowler, C. (1983). Converging sources of evidence for spoken and perceived rhythms of speech: Cyclic production of vowels in sequences of monosyllabic feet. *Journal of Experimental Psychology, 112,* 384–412.

Fowler, C. (1986). An event approach to a theory of speech perception from a direct-realist perspective. *Journal of Phonetics, 14,* 3–28.

Fowler, C., 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.

Freidlin, M., & Wentzell, A. (1984). *Random perturbations of dynamical systems.* New York: Springer-Verlag.

Gafos, A. (2006). Dynamics in grammar: Comments on Ladd and Ernestus & Baayen. In L. Goldstein, D. Whalen, & C. Best (Eds.), *Varieties of phonological competence (Laboratory phonology No. 8)* (pp. 51–79). Berlin: Mouton de Gruyter.

Gafos, A. I. (2002). A grammar of gestural coordination. *Natural Language and Linguistic Theory, 20,* 269–337.

Gafos, A., & Benus, S. (2003). On neutral vowels in Hungarian. In M.-J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 77–80). Barcelona, Spain: Universitat Autònoma de Barcelona.

Goldsmith, J. (1976). An overview of autosegmental phonology. *Linguistic Analysis, 2,* 23–68.

Goldsmith, J. (1993). Harmonic phonology. In J. Goldsmith (Ed.), *The last phonological rule* (pp. 21–60). Chicago: University of Chicago Press.

Goldstein, L. (2003). Emergence of discrete gestures. In M.-J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 85–88). Barcelona, Spain: Universitat Autònoma de Barcelona.

Goodwin, B. C. (1970). Biological stability. In C. H. Waddington (Ed.), *Towards a theoretical biology: Vol. 3. Drafts* (pp. 1–17). Chicago: Aldine.

Haken, H. (1977). *Synergetics: An introduction.* Heidelberg, Germany: Springer-Verlag.

Hall, N. (2003). *Gestures and segments: Vowel intrusion as overlap.* Unpublished doctoral dissertation, University of Massachusetts, Amherst, MA.

Harnad, S. (1990). The symbol grounding problem. *Physica D, 42,* 335–346.

Harshman, R. A., Ladefoged, P., & Goldstein, L. (1977). Factor analysis of tongue shapes. *Journal of the Acoustical Society of America, 62,* 693–707.

Haugeland, J. (1985). *Artificial intelligence: The very idea.* Cambridge, MA: MIT Press.

Hayes, B. (2004, April 23). *Stochastic phonological knowledge: The case of Hungarian vowel harmony.* Lecture presented at New York University, New York.

Higham, J. D. (2001). An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM Review, 43,* 525–546.

Hofstadter, D. R. (1979). *Gödel, Escher, Bach: An eternal golden braid.* New York: Basic Books.

Honorof, D. N., & Browman, C. P. (1995). The center or the edge: How are consonant clusters organized with respect to the vowel? In K. Elenius & P. Branderud (Eds.), *Proceedings of the XIIIth International Congress of Phonetic Sciences* (Vol. 3, pp. 552–555). Stockholm, Sweden: Stockholms Universitet.

Iskarous, K. (2005). Detecting the edge of the tongue: A tutorial. *Clinical Linguistics and Phonetics, 19,* 555–565.

Iverson, G. K. (1997). Review of W. Brockhaus, Final devoicing in the phonology of German. *American Journal of Germanic Linguistics and Literatures, 9,* 255–264.

Jessen, M. (1998). *Phonetics and phonology of tense and lax obstruents in German.* Amsterdam: Benjamins.

Kaun, A. (1995). *The typology of rounding harmony: An optimality theoretic approach.* Unpublished doctoral dissertation, University of California, Los Angeles.

Keating, P. A. (1990). Phonetic representations in generative grammar. *Journal of Phonetics, 18,* 321–334.

Kelso, J. A. S. (1984). Phase transitions and critical behavior in human bimanual coordination. *American Journal of Physiology, 15,* R1000–R1004.

Kelso, J. A. S. (1995). *Dynamic patterns: The self-organization of brain and behavior.* Cambridge, MA: MIT Press.

Kim, H., & Jongman, A. (1996). Acoustic and perceptual evidence for complete neutralization of manner of articulation in Korean. *Journal of Phonetics, 24,* 295–312.

Kloeke, W. (1982). *Deutsche phonologie und morphologie* [German phonology and morphology]. Tübingen, Germany: Max Niemeyer.

Kohler, K. (1984). Contrastive phonology and the acquisition of phonetic skills. In S. Eliasson (Ed.), *Theoretical issues in contrastive phonology* (pp. 73–84). Heidelberg, Germany: Julius Groos.

Krakow, R. A. (1989). *The articulatory organization of syllables: A kinematic analysis of labial and velar gestures.* Unpublished doctoral dissertation, Yale University, New Haven, CT.

Krakow, R. A. (1993). Nonsegmental influences on velum movement patterns: Syllables, sentences, stress, and speaking rate. In M. A. Huffman & R. A. Krakow (Eds.), *Nasals, nasalization, and the velum* (pp. 87–116). New York: Academic.

Kugler, P. N., & Turvey, M. T. (1987). *Information, natural law, and the self-assembly of rhythmic movement.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Ladefoged, P. (1988). The many interfaces between phonetics and phonology. *UCLA Working Papers in Phonetics, 70,* 13–23.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21,* 1–36.

Lindblom, B. (1983). Economy of speech gestures. In P. F. MacNeilage (Ed.), *The production of speech* (pp. 217–245). New York: Springer-Verlag.

McClelland, J. L., Rumelhart, D. E., & and the PDP Research Group. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 2. Psychological and biological models.* Cambridge, MA: MIT Press.

Nam, H., & Saltzman, E. (2003). A competitive, coupled oscillator of syllable structure. In M.-J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 2253–2256). Barcelona, Spain: Universitat Autònoma de Barcelona.

Newell, A., & Simon, H. A. (1976). Computer science as empirical inquiry: Symbols and search. *Communications of the ACM, 19,* 113–126.

Ohala, J. (1990). There is no interface between phonetics and phonology: A personal view. *Journal of Phonetics, 18,* 153–171.

Ohala, J. (1994). Towards a universal, phonetically-based, theory of vowel harmony. In *Proceedings of International Conference on Spoken Language Processing* (pp. 491–494).

Öhman, S. (1966). Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America, 39,* 151–168.

Percival, I., & Richards, D. (1982). *Introduction to dynamics.* Cambridge, UK: Cambridge University Press.

Perkell, J., Cohen, M., Svirsky, M., Matthies, M., Garabieta, I., & Jackson, I. (1992). Electromagnetic midsaggital articulometer (EMMA) systems for transducing speech articulatory movements. *Journal of the Acoustical Society of America, 92,* 3078–3096.

Petitot-Cocorda, J. (1985). *Les catastrophes de la parole: De Roman Jakobson à René Thom* [The catastrphes of speech. From Roman Jakobson to René Thom]. Paris: Maloine.

Pierrehumbert, J. B. (1990). Phonological and phonetic representations. *Journal of Phonetics, 18,* 375–394.

Port, R. F., & Crawford, P. (1989). Incomplete neutralization and pragmatics in German. *Journal of Phonetics, 17,* 257–282.

Port, R. F., & van Gelder, T. (Eds.). (1995). *Mind as motion: Explorations in the dynamics of cognition.* Cambridge, MA: MIT Press.

Pouplier, M., & Goldstein, L. (2005). Asymmetries in the perception of speech production errors. *Journal of Phonetics, 33,* 47–75.

Prince, A., & Smolensky, P. (1997). Optimality: From neural networks to universal grammar. *Science, 275,* 1604–1610.

Prince, A., & Smolensky, P. (2004). *Optimality theory: Constraint interaction in generative grammar.* Oxford, UK: Blackwell. (Original work published 1993)

Recasens, D. (1999). Lingual coarticulation. In W. J. Hardcastle & N. Hewlett (Eds.), *Coarticulation: Theory, data and techniques in speech production* (pp. 78–104). Cambridge, UK: Cambridge University Press.

Ringen, C. O., & Kontra, M. (1989). Hungarian neutral vowels. *Lingua, 78,* 181–191.

Rumelhart, D. E., McClelland, J. L., & the PDP Research Group. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 1. Foundations.* Cambridge, MA: MIT Press.

Rumelhart, D. E., Smolensky, P., McClelland, J. L., & Hinton, G. E. (1986). Schemata and sequential thought processes in parallel distributed processing. In J. L. McClelland, D. E. Rumelhart, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 2, pp. 7–57). Cambridge, MA: MIT Press.

Sagey, E. (1991). *The representation of features and relations in nonlinear phonology.* New York: Garland. (Original work published 1986)

Saltzman, E. (1986). Task dynamic coordination of the speech articulators: A preliminary model. In H. Heuer & C. Fromm (Eds.), *Experimental brain research series* (Vol. 15, pp. 129–144). New York: Springer-Verlag.

Saltzman, E., & Byrd, D. (2000). Task-dynamics of gestural timing: Phase windows and multifrequency rhythms. *Human Movement Science, 19,* 499–526.

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.

Schöner, G., & Kelso, J. A. S. (1988). A dynamic pattern theory of behavioral change. *Journal of Theoretical Biology, 135,* 501–524.

Siptár, P., & Törkenczy, M. (2000). *The phonology of Hungarian.* Oxford, UK: Oxford University Press.

Smith, C. L. (1991). The timing of vowel and consonant gestures in Italian and Japanese. In *Actes du XIIème Congrès International des Sciences Phonétiques, Aix-en-Provence, France* (Vol. 4, pp. 234–237).

Smolensky, P. (1986). Information processing in dynamical systems: Foundations of harmony theory. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 1. Foundations* (pp. 194–281). Cambridge, MA: MIT Press.

Smolensky, P. (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences, 11,* 1–74.

Sproat, R., & Fujimura O. (1993). Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics, 21,* 291–311.

Stetson, R. H. (1951). *Motor phonetics.* Amsterdam: North-Holland.

Stevens, K. N. (1972). The quantal nature of speech: Evidence from articulatory-acoustic data. In E. David & P. Denes (Eds.), *Human communication: A unified view* (pp. 51–66). New York: McGraw-Hill.

Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics, 17,* 3–45.

Stone, M. (1997). Laboratory techniques for investigating speech articulation. In J. Hardcastle & J. Laver (Eds.), *The handbook of phonetic sciences* (pp. 11–32). Oxford, UK: Blackwell.

Tiede, M. K., Vatikiotis-Bateson, E., Hoole, P., & Yehia, H. (1999). Magnetometer data acquisition and analysis software for speech production research (ATR Tech. Rep. No. TR-H 1999). Kyoto, Japan: ATR Human Information Processing Labs.

Tuller, B., Case, P., Ding, M., & Kelso, J. A. S. (1994). The nonlinear dynamics of speech categorization. *Journal of Experimental Psychology, 20,* 3–16.

Vago, R. M. (1980). *The sound pattern of Hungarian.* Washington, DC: Georgetown University Press.

van der Hulst, H. G. (1988). The geometry of vocalic features. In H. van der Hulst & N. Smith (Eds.), *Features, segmental structure and harmony processes* (Vol. 2, pp. 77–125). Dordrecht, The Netherlands: Foris.

Wiese, R. (1996). *The phonology of German.* Oxford, UK: Clarendon.

Wood, S. (1979). A radiographic analysis of constriction location for vowels. *Journal of Phonetics, 7,* 25–43.

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