Speech dynamics: Converging evidence from syllabification and categorization

Betty Tuller *a, Leonardo Lancia b

a National Science Foundation, 4201 Wilson Boulevard, Arlington, VA 22230, USA
b Laboratoire de Phonétique et Phonologie (CNRS, Sorbonne Nouvelle), 19 rue des Bernardins, 75005 Paris, France

ARTICLE INFO

Article history:
Received 11 March 2016
Received in revised form 7 February 2017
Accepted 9 February 2017
Available online xxxx

Keywords:
Dynamical systems
Models of speech perception
Models of speech production
Syllabification
Categorization

ABSTRACT

The present paper explores the dynamics of speech production and perception in the context of syllabification and categorization. The selective review includes empirical work and dynamical models that account for changes in the perception and production of syllable structure as transitions between attractors in a dynamical system and that highlight the role of instabilities as a mechanism for regulating flexibility and change. Different conceptual approaches to changes in perceptual categorization are reviewed, including a nonlinear dynamic model, a related Bayesian approach, and a hybrid approach. Of particular importance are recent models that incorporate cognitive factors (such as attention, expectation, and memory) and that change slowly or quickly relative to the changing acoustic input. These dynamical models allow phenomena such as self-organization, emergence, and other hallmarks of complex adaptive systems and may also suggest a mechanism linking speech production and perception, providing an alternative description to the internal models often invoked.

1. Introduction

Speech communication, from the movements of the articulators to taking turns in conversation, involves an intricate dance in time. Small differences in the timing of “velar” movement in articulatory synthesis, for example, can shift perception of the resulting acoustic signal among banana, bandana, bad#nana and bad#data (Rubin & Vatikiotis-Bateson, 1998) and alteration in the short intervals between turns of conversational participants can severely disrupt social interaction (Levinson & Torreira, 2015). In this paper, we consider as fundamental the notion that speech is a dynamic process that evolves in time and that the temporal flow is inseparable from the processes that regulate when production and perception are stable and that allow both stability and plasticity/flexibility to coexist. We will describe work that assesses and models how both stability and flexibility arise, coexist, and influence speech communication.

Over the last decades, we have explored the mechanisms involved in the stability and flexibility of speech communication, guided by concepts from the theory of nonlinear dynamical systems (e.g., Case, Tuller, Ding, & Kelso, 1995; Lancia, Nguyen, & Tuller, 2008; Nguyen, Lancia, Bergounioux, Wauquier-Gravelines, & Tuller, 2005; Nguyen, Wauquier-Gravelines, & Tuller, 2009; Tuller, 2003, 2004; Tuller, Case, Ding, & Kelso, 1994; Tuller & Kelso, 1990, 1991). These studies emphasize the role of interactions among the multiple heterogeneous processes that affect speech communication at different levels of organization and that unfold over different time scales. These time scales range from the fast time scale of milliseconds (characterizing neuronal firing) to the much slower time scales of years (characterizing learning, developmental processes, and language change). Although the interdependencies among processes that underlie the observed behaviors are a source of complexity when considered from the point of view of symbolic computations, mutual interactions can be a source of order when considered from a dynamical view. The sensorimotor system can reorganize in a task-specific manner through changes in the parameters governing the interactions, thus producing stable, yet flexible, behavior. This idea is grounded in theories of self-organization and pattern-formation in open systems far from equilibrium (particularly Haken’s Synergetics, 1977/1983). In what follows, we will summarize and illustrate the main principles governing this approach.

* Corresponding author. Fax: +1 703 292 9068.
E-mail address: btuller@nsf.gov (B. Tuller).
using examples from the production and perception of syllabification and perceived categorization (including illusory changes in categorization). In addition, we will describe different approaches to nonlinear dynamic modeling of these phenomena.

2. Non-equilibrium phase transitions in sensorimotor processes

Processes that unfold over time can be modeled as dynamical systems. In a dynamical system, the present state of the system depends in some rule-governed way on previous states. Differential equations or maps of essential variables offer a mathematical description of how a behavior’s essential parameters change as time passes and contextual parameters change. Interactions among dynamical systems are modeled as coupling relations: systems are coupled when the current state of one system depends on its own past values as well as on the past values of another system. Coupling relations can be linear or nonlinear, but the presence of nonlinear coupling relations is a necessary condition for the observation of non-equilibrium phase transitions, as described below.

Dynamical models of particular relevance for speech are complex, open systems (e.g., living systems) that require interaction with their environment, exchanging matter or energy to maintain an organized structure. This organized structure is described by collective variables, which capture the macroscopic behavior of the many individual degrees of freedom; collective variables obey a lower-dimensional dynamics than that describing the behavior of the individual components. The behavior of complex systems is also influenced by control parameters, which may be fixed from the outside (i.e., they quantify the influences of the environment) or may be generated within the system under consideration. Across some range of values, changes in a control parameter might have little or no observable change in system behavior despite ever-present fluctuations, i.e., the system remains stable. But when control parameters reach specific critical values, a system may become unstable and undergo qualitative changes in organization or behavior. These qualitative changes, or transitions between macroscopic patterns, reveal a new state of collective variables, which capture the new macroscopic behavior of the individual system components. This qualitative change in the behavior of a system driven by a change in a control parameter is termed a phase transition. An important herald of an approaching phase transition is the growth in instability of the current state (organization/behavior/attractor) until, when the threshold is crossed, the initially stable behavior vanishes and the stable system states shift until behavior adopts a new stable organization.

The evaluation of these principles in speech communication leads to a conception of speech production and perception as characterized by a limited number of stable states, or attractors, which allow the system to perform a discretization of articulatory gestures and perceptual space and which are associated with abstract speech categories. Changes in articulatory and/or perceptual state may occur as a pattern formation process resulting from a non-equilibrium phase transition. The control parameters governing phase transitions in speech can represent, for example, features of the input signals or top-down patterns of activation. In this view, coupling relations and speech-relevant control parameters act as constraints, harnessing the underlying dynamics so that their joint action produces stable patterns that are associated to a symbol or category. As a consequence, the same spatiotemporal pattern can be described as due to the interaction of many partially independent macroscopic dynamical systems or as due to the action of a low-dimensional macroscopic system in the achievement of a symbolic goal (such as a speech utterance or perception of a speech sound). The two levels of description are not redundant because the function served by the macroscopic system can be understood at the symbolic level but not at the macroscopic level (Pattee, 1972). In other words, we will usually find no traces of the goal defined at the macroscopic level in the equations that describe the microscopic systems. The integration between the microscopic and the macroscopic levels of description permits grounding abstract representations on physical reality (Rączaszek-Leonardi, 2012). It reconciles the lack of detail implicit in abstract symbolic representations with the sensitivity of the sensorimotor processes to the details of their physical implementation (Nguyen et al., 2009).

3. Production dynamics and syllabification

The principles of dynamical systems, introduced above, allow an explanation of the shifts in perception of the utterances banana, bandana, badanana and baddata with acoustic changes consequent to changes in the timing of model parameters akin to velar movement: the temporal structure of movements, in particular their relative phase, might act as a collective variable that indexes distinct production states with perceptual consequences. In flesh-and-blood speakers, small differences in the phasing of velar movement relative to the timing of alveolar closure may act as a control parameter, moving the speaker/listener through several macroscopic system states. In order to explore whether these different states emerge through a non-equilibrium phase transition, it is necessary to lead the speaker or listener through the states and assess how the switches in production and perception occur. However, there is usually little reduplication of syllables in English conversational speech, so it is unclear what might constitute a base cycle within which phase would be assessed. A base cycle can be induced in experiment by repeating an event, such as a word, syllable, or phrase (e.g., Cummins, 2009; Cummins & Simko, 2009). Repeating the event at a variety of rates can induce a varying base cycle. One well-known example comes from Stetson (1951) who observed that when subjects repeat groups of syllables at a range of rates, syllables such as /at/ that are clearly distinct from /ta/ at slow rates become perceptually identical to /ta/ at fast rates. As a subject produces a VC syllable (such as /at/) repetitively, with gradually increasing speaking rate, the syllable affiliation of the consonant appears to change, producing a CV series (/ta, ta.../). Stetson proposed that the need to simplify coordination caused the syllable-final consonant to become syllable-initial because the final consonant was “off phase, out of step with the syllable movement” (p. 96). Another interpretation of this observation is that the phasing among component gestures is a collective variable capable of exhibiting multiple patterns and complex behaviors. In this interpretation, the role of speech rate is that of an external control parameter whose
دریافت فوری متن کامل مقاله

امکان دانلود نسخه تمام متن مقالات انگلیسی
امکان دانلود نسخه ترجمه شده مقالات
پذیرش سفارش ترجمه تخصصی
امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
امکان دانلود رایگان ۲ صفحه اول هر مقاله
امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
دانلود فوری مقاله پس از پرداخت آنلاین
پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات