The gesture as an autonomous nonlinear dynamical system

Tanner Sorensen
Signal Analysis and Interpretation Laboratory, University of Southern California, 3740 McClintock Ave., EEB 400, Los Angeles, CA 90089-2564, USA

Adamantios Gafos
Universität Potsdam, Department Linguistik, Haus 14, Karl-Liebknecht-Straße 24-25, 14476 Potsdam, Germany
Haskins Laboratories, 300 George Street, Suite 900, New Haven, CT 06511, USA

Send reprint requests to Tanner Sorensen, Signal Analysis and Interpretation Laboratory, University of Southern California, 3740 McClintock Ave., EEB 400, Los Angeles, CA 90089-2564, USA
At various stages in its development, this paper has benefited from meetings with Carol Fowler, Mark Tiede, Elliot Saltzman, Doug Whalen, Khalil Iskarous, and Louis Goldstein, as well as through presentations to the working group of Doug Whalen at Haskins Laboratories, to the Speech Production and Perception Laboratory of Doug Whalen at the City University of New York, to the Workshop on Dynamic Modeling at the Institute for Phonetics in Cologne, and to the phonology-phonetics group at the University of Potsdam. This research has been supported in part by a grant from the European Research Council (http://erc.europa.eu/), Advanced Grant STIMOS 249440.
Abstract

We propose a theory of how the speech gesture determines change in a functionally relevant variable of vocal tract state (e.g., constriction degree). A core postulate of the theory is that the gesture determines how the variable evolves in time independently of any executive time-keeper. That is, the theory involves intrinsic timing of speech gestures. We compare the theory against others in which an executive time-keeper determines change in vocal tract state. Theories which employ an executive time-keeper have been proposed to correct for disparities between theoretically predicted and experimentally observed velocity profiles. Such theories of extrinsic timing make the gesture a nonautonomous dynamical system. For a nonautonomous dynamical system, the change in state depends not just on the state, but also on time. We show that this nonautonomous extension makes surprisingly weak kinematic predictions both qualitatively and quantitatively. We propose instead that the gesture is a theoretically simpler nonlinear autonomous dynamical system. For the proposed nonlinear autonomous dynamical system, the change in state depends nonlinearly on the state and does not depend on time. This new theory provides formal expression to the notion of intrinsic timing. Furthermore, it predicts experimentally observed relations among kinematic variables.
The gesture as an autonomous nonlinear dynamical system

Introduction

Extrinsic timing of speech gestures has been proposed to correct for disparities between theoretically predicted and experimentally observed speech kinematics (Kröger, Schröder, & Opgen-Rhein, 1995; Byrd & Saltzman, 1998, 2003). Extrinsic timing makes the speech gesture a nonautonomous dynamical system. This means that the state of the system does not fully determine how it evolves and that the system is further influenced by a process which depends explicitly on time. This process is external to the gesture, but possibly internal to the phonological system. This time-dependent process influences how the gesture evolves, but the gesture does not in turn influence this process. Thus, the time-dependent process plays the role of an external time-keeper, governing the evolution of the gesture as a sort of clock.

In principle, nonautonomy may be necessary in order to correct for disparities between theoretically predicted and experimentally observed speech kinematics. However, nonautonomy is undesirable for theoretical reasons. For one thing, nonautonomy severely constrains analytical options. A deeper conceptual problem with nonautonomy is that it implies a regress of control structures. If an external time-keeper influences the evolution of the gesture, then we ought to identify the time-keeper of the time-keeper, and so on ad infinitum (Beek, Turvey, & Schmidt, 1992, p. 69). Conversely, autonomy has no external time-keeper, thus halting the regress of control structures.

We argue that, despite making the gestural dynamics more complex, the kinematic predictions of the nonautonomous extension are surprisingly weak both qualitatively and quantitatively. We propose a revised nonlinear dynamical system which maintains that the gesture is an autonomous dynamical system. The system is a model of intra-gestural dynamics, that is, of the single speech gesture. The system does not determine inter-gestural dynamics, that is, how gestures are coordinated.

The following two sections distinguish autonomous from nonautonomous and
nonlinear from linear theories of the speech gesture. We propose that the gesture is a nonlinear autonomous dynamical system. We lay out the predictions of the proposed dynamical system and qualitatively evaluate these predictions. Using the University of Wisconsin X-ray microbeam database (Westbury, Milenkovic, Weismer, & Kent, 1990), we then quantitatively evaluate the fit of the proposed dynamical system to the movement kinematics of tongue dorsum raising and lowering. Finally, we explain how, in an isochronous speech task (e.g., ‘bapabapa...’), where the proposed revised dynamical system is driven by a periodic external force, trajectories become aperiodic. Phase portraits and Hooke diagrams of the proposed driven nonlinear system are consistent with empirical observations. In sum, we illustrate the revised system’s fit to the kinematics in both non-cyclic speech and cyclic tasks (i.e., regularly-timed speech with a metronome).

The proposed nonlinear dynamical system gives both formal expression and empirical justification to Carol Fowler’s vision of intrinsic timing at the level of a single gestural event. It restores the autonomy postulate and offers a basis for moving forward with that vision.

**Autonomous versus nonautonomous**

Theoretical work suggests that the gesture has intrinsic timing (Fowler, 1980). This means that the spatial and temporal extent of the gesture is determined by the gesture itself, not by any system external to the gesture. In contrast, an extrinsic timing theory of the gesture would mean that the spatial and temporal extent of the gesture is influenced by systems which are external to the gesture. This section defines terms and introduces the autonomous versus nonautonomous distinction as it has been applied to gestures.

A *dynamical system* is a formal model which expresses a rule for how the *state* of a system changes in time. The state is the minimal set of variables required to predict how the system evolves. The set of all possible states is the *phase space*. For one-dimensional systems, where there is only one state variable $x$, the phase space is the line of real
numbers. Change in the state of the system is due to presumed causes, represented by the force field $f(x)$ of the dynamical system. As the state of the system changes, it traces out a path in phase space called a trajectory.

Autonomous dynamical systems have the form of Equation 1. Specifically, this equation states that at any time instant $t$ the rate of change $\dot{x} = dx/dt$ of $x$ is a function $f(x)$ which depends only on the state $x$ and not on time $t$.

$$\dot{x} = f(x)$$ (1)

This means that in a short time interval $dt$, the change $dx$ in the state of the autonomous dynamical system is $f(x)dt$. In an autonomous system, we can fully describe the behavior of the state variable $x(t)$ by considering just three cases. If $f(x)$ is positive, then $x$ will increase by $f(x)dt$, where $dt$ is a short time step. If $f(x)$ is negative, then $x$ will decrease, again by $f(x)dt$ where $dt$ is a short time step. Finally, if $f(x)$ is zero, $x$ stays the same (this $x$ is called a fixed point). Thus, $f(x)$ points in the direction of change for $x$, positive or negative, and specifies the magnitude of the change. The force field $f(x)$ of an autonomous dynamical system is stationary (i.e., does not vary over time). As a consequence of stationarity, no two trajectories cross (Coddington & Levinson, 1955). This is because at any given state $x$, the change in state dictated by $f(x)$ is fixed. For a crossing point to exist would require that $f(x)$ is not unique at that crossing point. This is not possible in autonomous dynamics. The coordinates of the state in the phase space fully determine the evolution of the system regardless of time. The no crossing consequence of autonomy plays a crucial role in the qualitative understanding of dynamical systems.

Nonautonomous dynamical systems have the form of Equation 2. Specifically, this equation states that at any time instant $t$ the rate of change $\dot{x} = dx/dt$ of $x$ is a function $f(x,t)$ which depends both on the state $x$ and on the time $t$.

$$\dot{x} = f(x,t)$$ (2)

The state of a nonautonomous dynamical system does not uniquely determine the
evolution of the system because the force field is nonstationary (i.e., varies over time). As a consequence of nonstationarity, trajectories of the system can cross in the phase space and multiple trajectories can pass through each point. This is because at any given state $x$, the change in state dictated by $f(x,t)$ depends on the time instant $t$ at which the system is in state $x$. This means that $f(x,\cdot)$ is not unique, so multiple trajectories can evolve in different ways starting from state $x$.

We illustrate autonomy and nonautonomy with some examples. Consider the harmonic oscillator as a dynamical model of the spring. The system has the form

$$m\ddot{x} + kx = 0,$$

where $x$ is a state variable representing displacement from the resting position of the spring, $m$ is mass and $k$ is the stiffness parameter of the oscillating spring. This equation derives from Newton’s law,

$$m\ddot{x} = F,$$

where the force $F$ is a restoring force $F = -kx$. The farther the displacement or distance $x$ of the mass $m$ from the fixed point, the stronger the restoring force. Equation 4 is a second-order equation (i.e., the highest derivative is a second derivative) which we can rewrite as a system of two first-order equations (i.e., as two equations whose highest derivatives are first derivatives) by introducing a new state variable $y$ and setting $\dot{x} = y$, along with some rearrangement of terms in the original second-order equation to express

$$\dot{x} = y$$

$$\dot{y} = -kx/m$$

Each equation in the resulting system of equations has the form of Equation 1 and the system is thus autonomous (henceforth and without loss of generality we are assuming unit mass, $m = 1$). The number of such equations corresponds to the dimension of the phase space. The harmonic oscillator has two equations and is thus two-dimensional. Because all examples we consider below are two-dimensional, we refer to the phase space as the phase
plane with coordinates $x$ and $y = \dot{x}$, displacement and its derivative (i.e., velocity). An example solution to the harmonic oscillator is given in Figure 1, with displacement and velocity as a function of time. In the phase plane, the trajectories which correspond to these time-solutions are loops, as shown in Figure 2. Assume that during some time interval $[t_a, t_b]$ we observe partly two loops with one tracing a trajectory inside the other. Then, as a consequence of autonomy, no trajectories can cross, and thus the inner loop must stay inside the outer loop for all time. The equation of motion for the harmonic oscillator can be seen to impose this particular geometric structure on trajectories in the phase plane. The equation is an invariant statement characterizing an infinity of trajectories. That is, regardless of initial conditions, the way $x$ and $y = \dot{x}$ change must follow the pattern shown in Figure 2.

As a second example of an autonomous dynamical system, consider the damped version of the harmonic oscillator is the system

$$\ddot{x} + b\dot{x} + kx = 0$$

with $b$ a damping constant. For a damped system, in addition to the restoring force $F_1 = -kx$, there is another force $F_2 = -b\dot{x}$, which is a function of velocity $\dot{x}$. Solutions to this system, shown in Figure 3, are oscillations which decay in amplitude due to damping and approach a fixed point at the origin of the state space. As an autonomous system, the trajectories never cross in the phase plane, as shown in Figure 4, keeping in mind that approaching the fixed point is not the same as crossing. Any trajectory that is traced below or above some other trajectory stays in that region of the phase space and this holds true for all time.

An example of a nonautonomous system is the Duffing oscillator (Duffing, 1918).

$$\ddot{x} + b\dot{x} + x^3 = \alpha \cos t$$

The right-hand side is a force which is periodic in $t$. Due to the explicit dependence of the right-hand side term on $t$, the system is not self-governed. The right-hand side term plays
the role of external time-keeper and contributes to the oscillation. A solution to this system is shown in Figure 5. The solution becomes aperiodic as the result of forcing. In the phase plane, trajectories can cross. This is shown by the single trajectory plotted in Figure 6, which concretely illustrates a consequence of nonautonomy: any given crossing point \((x_i, \dot{x}_i)\) on the phase plane in Figure 5 is a state at which the nonautonomous dynamical system dictates different evolutions emanating from that state. This is the result of the time-dependent force function. The trajectory crossings graphed in Figure 6 are possible only when the force depends on time \(t\).

With these examples of autonomous and nonautonomous systems as a reference point, we now consider the gesture as a dynamical system. The gesture is a dynamical system whose state is a variable which has functional relevance for speech production. Motivated by Articulatory Phonology (Browman & Goldstein, 1990) and Task Dynamics (Saltzman & Munhall, 1989), as well as by studies in non-speech motor control (e.g., limb control in humans and monkeys, Georgopoulos, Kalaska, & Massey, 1981; Morasso, 1981; Abend, Bizzi, & Morasso, 1982), this functionally relevant variable may be a spatial variable (e.g., constriction degree, vocal tract length, lower tooth height). Spatial variables of vocal tract shape interact with airflow to determine the acoustics and aerodynamics of the vocal tract as described in the acoustic theory of speech production (Fant, 1971; Stevens, 2000), and thus aerodynamic and acoustic variables may likewise take on the role of functionally relevant variables (Guenther, 1995). This paper considers the case where the gesture controls a spatial variable as its state. By control, we mean that the gesture moves the state along trajectories of the force field, starting at some initial position and stabilizing the state at a target value. At the target, movement of the vocal tract ceases under the influence of balanced and opposing forces, corresponding to a posture of the vocal tract at which the state of the gesture has been stabilized (see Feldman, 1966, for a related idea). For example, a bilabial closure gesture stabilizes lip aperture at 0 mm. The target value is a fixed point of the force field. Being able to move
the state along a trajectory of a force field involves solving an inverse dynamics problem, namely, how to coordinate muscle activity in the vocal tract so as to guide the state along the trajectory in the force field. For example, when producing a bilabial closure, the jaw and lips must move together as a coordinative structure or synergy (Fowler, Rubin, Remez, & Turvey, 1980; Fowler, 1980; Latash, 2008; Diedrichsen, Shadmehr, & Ivry, 2010) in order to close the lips. Inverse dynamics problems commonly arise in motor systems, and methods exist to characterize aspects of the inverse dynamics problem in speech (Ouni & Laprie, 2005; Lammert, Goldstein, Narayanan, & Iskarous, 2013; Sorensen, Toutios, Goldstein, & Narayanan, 2016). A motor system which solves the problem is able to relate the motor commands to the muscles with the corresponding changes in the state of the system and with the corresponding somatosensory consequences. This involves an interplay between state prediction on the one hand and auditory and somatosensory feedback on the other, which we do not take up here. See Houde and Nagarajan (2011) and Tourville and Guenther (2011) for speech and Todorov (2005) and Shadmehr and Mussa-Ivaldi (2012) for general treatments of this interplay.

One basic dynamical model of the gesture is a special case of the damped oscillator

$$\ddot{x} + b\dot{x} + kx = 0,$$  \hspace{1cm} (8)

where $b = 2\sqrt{mk}$ (Browman & Goldstein, 1990; Saltzman & Munhall, 1989; Fowler et al., 1980). When these parameters satisfy this relation, the system is critically damped and $x$ approaches zero but does not oscillate about zero (Fowler et al., 1980, p. 396). This is an autonomous dynamical system. A solution to this equation is shown in Figure 7 (contrast with the damped oscillations in Figure 3). We set the initial conditions to $(x = 1, \dot{x} = 0)$. That is, if $x$ represents for example the tract variable of lip aperture, then $x = 1$ means that lip aperture starts at the value of 1 and converges in the way shown in this plot at the target of 0 which corresponds to lip closure. A snapshot of the phase plane, with trajectories traced from six different initial conditions, is shown in Figure 8.

We highlight some aspects of this model that anticipate developments in the
forthcoming sections. It is useful to consider the gesture to set up a potential $V(x)$ in which the motion of the state $x$ takes place. The potential exerts a restoring force which pulls $x$ towards a fixed point $x^*$, the minimum of the potential (see Gafos, 2006, for basic notions). The restoring force is the negative gradient of the potential $V(x)$, that is,

$$f(x) = \dot{x} = dx/dt = -\nabla V(x).$$

Thus, the potential of the gestural model in Equation 8 is

$$V(x) = kx^2/2, \quad (9)$$

where $k$ is the stiffness of the gesture and $x$ is displacement from the target of the gesture (cf. Figure 9). This is the so-called harmonic potential because the harmonic oscillator and its damped variants discussed above make $x$ move in this potential. Recall that the force exerted by this potential is one of the forces contributing to the motion of the system:

$$F_1 = -kx,$$

where $x$ is the distance from the fixed point. The farther the displacement away from the fixed point, the greater the restoring force (Hooke’s law, see Figure 10). This force represents elasticity. The other force due to damping, $F_2 = -b\dot{x}$, represents viscosity. In viscous behavior, the force depends (only) on velocity, not on displacement (cf. McNeill, 1992, pp. 82-83). The essential point is that both $F_1$ and $F_2$ are independent of time and are linear. The former property corresponds to the autonomous versus nonautonomous distinction which we emphasize in this section. The latter property corresponds to the linear versus nonlinear distinction we take up in the next section.

Making use of the notion of potential, we can rewrite Equation 8, the equation of motion of the critically damped oscillator, as the autonomous dynamical system in Equation 10.

$$\ddot{x} + b\dot{x} + \nabla(kx^2/2) = 0 \quad (10)$$

If gestural units are defined via this or some other autonomous dynamical system, then the spatial and temporal properties of the gesture are fully determined by the parameters $b$, $k$ internal to this model. The spatial and temporal extent is not influenced by some external variable referring to time $t$. Thus, for example, for the model of gestures introduced above,
the gestural duration or settling time of the gesture, which in Figure 7 corresponds to the
time needed for displacement to go from the initial position at \( x_0 = 1 \) to the fixed point at
\( x^* = 0 \), is a quantity that can be expressed analytically by the parameters of the model and
is independent of time \( t \). The same holds for other temporal aspects of the gesture such as
the time to peak velocity or the magnitude of the peak velocity. All of these properties are
independent of time \( t \).

Speech involves ensembles of gestures organized in time and thus a mechanism which
activates and deactivates gestures is needed. Any such mechanism is implicated in the
notion of inter-gestural timing. One such mechanism switches the force field of a gesture on
and off as

\[
\ddot{x} + a(t) \left( \frac{b\dot{x}}{-F_2} + \nabla \left( kx^2 / 2 \right) \right) = 0
\]  

(11)

where \( a(t) \) is the step function

\[
a(t) = \begin{cases} 
1 & t \in [t_a, t_b] \\
0 & \text{otherwise}
\end{cases}
\]  

(12)

which activates the gesture over some time interval (Saltzman & Munhall, 1989). Figure 11
graphs activation over time, and Figure 12 graphs a particular solution to Equation 11,
given this step pattern of activation.

Turning gestures on and off as shown above appears to make Equation 10
nonautonomous as Equation 11, which can be made to have the form of Equation 2 by
introducing a new state variable \( y \) and setting \( \dot{x} = y \), along with some rearrangement of
terms in order to express Equation 10 as two first-order equations.

\[
\begin{align*}
\dot{x} &= y \\
\dot{y} &= -a(t) \left( by + \nabla (kx^2 / 2) \right)
\end{align*}
\]  

(13)

The system in 13 is nonautonomous because the change in state expressed in its second
equation is not only a function of the state variables \( x, y \), but also a function of time \( t \).
Nevertheless, Equation 11 is equivalent to the following piecewise autonomous system.

\[
\dot{x} = \begin{cases} 
- b \dot{x} - k x & \text{for } t \in [t_a, t_b] \\
0 & \text{otherwise}
\end{cases}
\]  

(14)

Over the interval \([t_a, t_b]\) we observe the intrinsic dynamics of the gesture which arises from the damped motion of \(x\) in the potential \(V(x)\). The intrinsic gestural dynamics is an autonomous dynamical system. During the interval of nonzero activation, no coefficient in the dynamical law above depends on time (i.e., stiffness \(k\) and damping \(b\) are constant).

Equation 11 with step activation fails to predict a property of speech movements, namely, that near symmetry is typical of speech velocity profiles (i.e., velocity graphed as a function of time). Ostry, Cooke, and Munhall (1987) report proportional times to peak velocity (i.e., time to peak velocity divided by duration; 0.50 is symmetric) in the range of 0.36 to 0.43 for tongue dorsum lowering at fast and slow speech rates in discrete and repetitive tongue dorsum movement tasks. Nearly symmetrical velocity profiles have also been reported for the speech movements of jaw lowering (Ostry et al., 1987), tongue dorsum movement in vowels (Munhall, Ostry, & Parush, 1985), glottal abduction (Munhall et al., 1985), and labial constriction (Byrd & Saltzman, 1998). In contrast to these reports, Equation 11 with step activation predicts proportional time to peak velocity of 0.20 (see Figure 13). The reported deviations from 0.20 indicate that we need a correction to Equation 11 with step activation.

One correction for short proportional time to peak velocity is to make gestural activation continuous. In the context of a model of the speech gesture, this was anticipated by Perrier, Abry, and Keller (1988) and formal implementations were later proposed by Kröger et al. (1995) and Byrd and Saltzman (1998, 2003). Whereas step activation makes the gesture autonomous during its interval of activation, these proposals make the gesture nonautonomous during this interval. For instance, Equation 15 defines activation
as a continuous function of time (Kröger et al., 1995).

\[
a(t) = \begin{cases} 
0, & \text{if } t < t_a \\
\sin \left( \frac{2\pi(t-t_a)}{4(t_b-t_a)} \right), & \text{if } t_a \leq t < t_b \\
1, & \text{if } t_b \leq t < t_c \\
\sin \left( \frac{2\pi(t-t_c)}{4(t_c-t_d)} \right), & \text{if } t_c \leq t < t_d \\
0, & \text{if } t \geq t_d 
\end{cases}
\]

Figure 14 graphs this particular continuous activation function over time. The graph displays a quarter sine rise over the interval \([t_a, t_b]\) and a quarter sine fall over the interval \([t_c, t_d]\). Figure 15 graphs a particular solution to Equation 11 given the continuous pattern of activation.

Continuous activation makes the solution to Equation 11 take on a range of proportional times to peak velocity for varying \(t_b\) and \(t_c\) in Equation 15 (i.e., for varying activation frequency). Examples are the solid velocity profiles of Figure 16. These are corrections to Equation 11 with step activation, which has proportional time to peak velocity of 0.20 (see the dashed velocity profile of Figure 16).

Continuous activation corrects for short proportional time to peak velocity, but it makes the gesture a nonautonomous dynamical system which involves an executive time-keeper. Whereas the nonlinear system with step activation admits a piecewise autonomous definition as Equation 14, the linear system with continuous activation does not admit a piecewise autonomous definition. The next section proposes a different model of the gesture which retains autonomy and uses step activation. The model accounts for kinematic properties of the speech gesture such as the typically large proportional time to peak velocity and characteristic relationships between the kinematic variables of movement amplitude, peak velocity, and movement time. These properties characterize the autonomous, damped motion of the state in the potential of the proposed dynamical system.
Nonlinear versus linear

Equation 11 with step activation is the simplest dynamical system for discrete goal-directed tasks. The two forces of this system are the restoring force $F_1 = -kx$ and the damping force $F_2 = -b \dot{x}$. Both of these are linear. Hence, two types of nonlinearity extend the linear system: nonlinear damping and nonlinear restoring force. The former may introduce oscillations even without external periodic forcing. This property is undesirable for discrete goal-directed tasks. In the latter extension, a departure from the space of linear models is expressed in the most general way by considering different forms of the potential function $V(x)$ as shown below.

\[ \ddot{x} + b \dot{x} + \nabla V(x) = 0 \]

(16)

Any potential other than the potential of the harmonic oscillator $V(x) = kx^2/2$ is called anharmonic (Haken, 1973, p. 108). This section argues that the gesture has an anharmonic potential. Specifically, we propose that a negative quartic term introduces a correction to the harmonic potential of Equation 9. This corrects for the short proportional time to peak velocity of Equation 11 with step activation.

\[ \ddot{x} + b \dot{x} + \nabla (kx^2/2 - dx^4/4) = 0 \]

(17)

Figure 17 graphs the force of the revised model

\[ F(x) = -\nabla V(x) = -kx + dx^3 \]

(18)

as a function of displacement. In the neighborhood of $x^* = 0$, the linear term attracts $x$ to $x^*$. For small displacement, the cubic deviation $dx^3$ is negligible and thus the restoring force $F$ is approximately linear. As displacement increases, the nonlinear term opposes the linear restoring force, and $F$ bows back toward the $x$-axis. This revision to the linear system has a number of consequences to be illustrated in the next section. It is also the simplest such revision for two reasons. On the one hand, increasing the degree of the force polynomial to higher than cubic (higher than quartic in the potential) contributes
quantitative, not qualitative, distinctions within the space of expanded models. The key departure from the linear model is that the restoring force is weakened for relatively large displacements by the addition of these higher degree terms. Higher degree corrections are inessential in qualitative terms. On the other hand, a quadratic term in the force would give us $F = -kx + cx^2$ which is qualitatively incorrect. While the point $x = k/c$ is a maximum of the potential $V(x) = kx^2/2 - cx^3/3$, the point $x = -k/c$ is not a maximum, and thus $V(x)$ is asymmetric about $x^*$.

The qualitative distinction at hand can also be described in terms of potential differences. The slope of the anharmonic potential

$$V(x) = kx^2/2 - dx^4/4$$

has absolute value less than the slope of the harmonic potential for all $x \neq 0$ in the basin of attraction (i.e., for $0 < |x| < \sqrt{k/d}$). See Figure 18. This means that acceleration arising from displacement is less in the anharmonic potential than in the harmonic potential.

The following sections evaluate the predictions of the nonlinear dynamical system of Equation 17.

**Qualitative evaluation**

A qualitative indication that the nonlinear system is on the right track comes from its predictions for characteristic relations among kinematic variables. Specifically, the nonlinear system predicts

1. that velocity profiles are nearly symmetric, with proportional times to peak velocity close to 0.50,
2. that amplitude and peak velocity covary nonlinearly, and
3. that the ratio of peak velocity to amplitude varies inversely with movement duration (i.e., the shorter the duration, the greater the ratio of peak velocity to amplitude).
We now turn to illustrate Prediction 1, Prediction 2, and Prediction 3 of the nonlinear system. Evidence that these predictions are borne out in speech kinematics derives from observations of oral and laryngeal speech gestures, as reviewed earlier. The linear system makes the first prediction but not the second and third predictions, if corrected with a continuous activation function, and makes none of the predictions otherwise.

We first illustrate Prediction 1, namely, that the nonlinear system predicts velocity profiles which are nearly symmetric, with proportional times to peak velocity close to 0.50. Figure 19 compares representative proportional times to peak velocity of the nonlinear Equation 17 with that of the linear Equation 11 with step activation. When $d = 0$, corresponding to the harmonic potential of the linear system, proportional time to peak velocity is 0.20. When $d > 0$, corresponding to the anharmonic potential of the nonlinear system, proportional time to peak velocity increases. The proportional times to peak velocity of 0.36 and 0.50 are plotted as representative examples. This is consistent with the findings surveyed earlier that velocity profiles are nearly symmetric.

Next, we illustrate Prediction 2, namely, that the nonlinear system predicts that amplitude and peak velocity covary nonlinearly. Figure 20 plots peak velocity as a function of displacement. When $d = 0$, corresponding to the harmonic potential of the linear system, the plot is a line. When $d > 0$, corresponding to the anharmonic potential of the nonlinear system, peak velocity undergoes soft saturation at large displacement. The predictions of both the linear system (with or without continuous activation) and the nonlinear system are consistent with the finding that peak velocity and movement amplitude correlate positively for tongue dorsum raising and lowering (Ostry and Munhall, p. 644; Ostry, Keller, and Parush, pp. 629-630; Munhall et al., p. 467) and vocal fold adduction and abduction (Munhall et al., 1985, p. 462, p. 467). Crucially, however, only the prediction of the nonlinear system is consistent with a quadratic trend in the nonlinear regression of peak velocity against movement amplitude displayed by two of three speakers of Ostry and Munhall (1985, p. 644). We do not believe that testing the generality of this relation is a
trivial task. In particular, evaluation of a nonlinear relationship between peak velocity and movement amplitude is likely not as straightforward as performing a polynomial regression, as done, for instance, in Ostry and Munhall (1985). The main challenge is to estimate the parameter values for the movements and determine where in the parameter space the system was when that movement was generated. Parameter instantiation is important because not all parameter values will lead to a nonlinear relationship between peak velocity and movement amplitude. Indeed, for initial constriction degrees sufficiently close to the constriction degree target, the relationship is approximately linear. Hence, cases in past literature where a nonlinear relation is not reported do not imply that the saturation of peak velocity at large amplitudes is false. What is crucial for current purposes is that this relation is outside the scope of the linear model (with or without continuous activation) whereas it is within the scope of the revised nonlinear model.

Finally, we illustrate Prediction 3, namely, that the ratio of peak velocity to amplitude varies inversely with movement duration (i.e., the shorter the duration, the greater the ratio of peak velocity to amplitude). Figures 21 and 22 plot amplitude-normalized peak velocity against gesture settling time (i.e., the time from movement onset to target achievement) for trajectories of the linear system with continuous activation and for trajectories of the nonlinear system, respectively. As settling time increases, amplitude-normalized peak velocity decays for the nonlinear system but not for the linear system. This nonlinear relation between amplitude-normalized peak velocity and settling time is outside the scope of the linear system, with or without continuous activation. Thus, the prediction of the nonlinear system is consistent with the finding that although peak velocity, movement amplitude, and gesture settling time vary depending on initial conditions, the relation among them is invariant (Ostry & Munhall, 1985; Munhall et al., 1985). This relation is described by an equation of constraint in the sense of Fowler et al. (1980, p. 401). This equation is \( \frac{\text{peak velocity}}{\text{amplitude}} = \frac{c}{\text{settling time}} \), where \( c \) is a constant of proportionality. A consequence of intrinsic timing is that this
equation of constraint is not the result of external forcing. Rather, the equation of
constraint is a consequence of the anharmonic potential of the intrinsic gestural dynamics.
This equation of constraint characterizes the kinematics of the tongue dorsum (Ostry &
Munhall, 1985; Munhall et al., 1985) and the glottis (Munhall et al., 1985). In turn, this
invariance across effectors hints to the abstractness of the law(s) that underwrite speech
units. In contrast, the linear system does not entail this equation of constraint on
movements of different amplitude, with or without continuous activation.¹

In sum, nonautonomy in the intrinsic gestural dynamics via continuous activation
corrects for short proportional time to peak velocity but fails to predict characteristic
relations among kinematic variables. In contrast, autonomy and nonlinearity correct for
short proportional time to peak velocity and predict characteristic relations among the
kinematic variables of movement amplitude, duration, and peak velocity.

Quantitative evaluation

We compare the nonlinear system with anharmonic potential against the linear
system with harmonic potential using an X-ray microbeam dataset with recordings from 43
speakers. This dataset exhibits inter-speaker variability in velocity profiles. We fit the
trajectories of the linear and nonlinear systems to the velocity profiles of tongue dorsum
raising and lowering.

The dataset comes from the X-ray Microbeam Speech Production
Database (Westbury et al., 1990). The X-ray Microbeam Speech Production Database
consists of 57 speakers carrying out 118 speech or speech-related tasks. We analyze the
tongue dorsum raising and lowering movements of Task 16, a citation task with target
items [a'ka] and [a'ga]. Forty-three of the 57 total speakers performed Task 16. The tongue

¹Ostry and Munhall (1985) and Munhall et al. (1985) note that the equation of constraint is a consequence
of varying the parameters of the undamped harmonic oscillator, and Fuchs, Perrier, and Hartinger (2011)
ote note the same for the damped harmonic oscillator. However, the equation of constraint is not a consequence
of varying the movement amplitude.
dorsum (T4) pellet was mistracked by the X-ray microbeam system for three speakers during [ag], for four speakers during [ga], for four speakers during [ak], and for two speakers during [ka] (see Westbury et al., 1990, Section 6.7.1 for probable causes of mistracking). Pellet positions are expressed in a two-dimensional cranial coordinate system (see Westbury et al., 1990, Section 6.3). The origin of the coordinate system is the caudal-most edge of the central maxillary incisors. The anteroposterior axis is the intersection of the midsagittal and maxillary occlusal planes. The superior-inferior axis is normal to the maxillary occlusal plane and passes through the origin. Table 1 lists the numerical ID codes of the speakers whose data was not mistracked. We use the T4 pellet in the 159 resulting X-ray microbeam recordings from 43 different speakers to compare the observed tongue dorsum lowering and tongue dorsum raising kinematics with predictions of the linear and nonlinear dynamical systems.

The gestural dynamics has one degree of freedom. This means that the dimension of the movement traces must be reduced from two to one. A non-arbitrary way of reducing the two-dimensional movement trace $\gamma$ to one dimension is to project it onto the principal component $u$ of movement during the interval $I = [\tau_a, \tau_b]$, where movement onset time $\tau_a$ and movement offset time $\tau_b$ are the sample time points at which $d\gamma(t)/dt$ passes some velocity criterion at the onset and offset of movement. After identifying $\tau_a$ and $\tau_b$, we approximate the velocity profile $d\text{proj}_u \gamma(t)/dt$ with central differences over the samples $d\text{proj}_u \gamma(t)/dt$, $\tau_a \leq t \leq \tau_b$. For tongue dorsum lowering movements, the minimum is the peak velocity because the superior-inferior coordinate of the X-ray microbeam pellet is negative-going. For tongue dorsum raising movements, the maximum is the peak velocity because the superior-inferior coordinate of the X-ray microbeam pellet is positive-going.

The velocity criterion mentioned in the above paragraph is a relative criterion, determined for each velocity profile separately as 20% of peak velocity. This contrasts with an absolute criterion, which would have that a movement begins when the X-ray microbeam pellet rises above and falls below a specified tangential or component velocity
(e.g., 5 cm/s), independently of peak velocity. We do not use such an absolute criterion. The reason for choosing the relative velocity criterion rather than the absolute velocity criterion is that the velocity of the tongue is known to differ from speaker to speaker in the database which we used (Simpson, 2001). In particular, male speakers had significantly greater tongue dorsum velocity than female speakers. Such differences likely reflect the influence of vocal tract morphology on speech articulation, as the adult male pharynx is larger than the adult female pharynx (Fitch & Giedd, 1999). Applying an absolute velocity criterion to identify movement onset and offset does not take into account that the velocity at movement onset and offset is smaller for a small vocal tract than for a large vocal tract.

We choose a different criterion for each velocity profile for identifying movement onset and offset because tongue dorsum velocity at movement onset and offset varies from speaker to speaker. As we measured two velocity profiles per speaker (i.e., one for vocal tract constriction and another for release), we note that in principle we could use a single criterion for both velocity profiles. However, the movements are of qualitatively different type, with constriction involving tongue dorsum raising, whereas the release involves tongue dorsum lowering. Therefore, a single velocity criterion may not be appropriate for both velocity profiles. Rather than use a single criterion for two different types of movement, we use different criteria for constriction and release movements.

The histograms of Figure 23 show that the principal components \( \mathbf{u} \) are nearly parallel to the superior-inferior axis \( (90^\circ) \) for most tongue dorsum raising and lowering movements. Figure 24 graphs the 80 observations of \( \text{proj}_\mathbf{u} \gamma(t) \) for tongue dorsum raising and lowering, centered on the time of onset of movement toward the oral vowel target.

We use the simplex search method (Lagarias, Reeds, Wright, & Wright, 1998) as implemented in MATLAB to optimize the parameters of the linear and nonlinear systems. The objective function which we optimized is

\[
f(\theta) = \sum_{t \in \mathcal{I}} \left( \frac{d\gamma(t)}{dt} - d\text{proj}_\mathbf{u} \gamma(t)/dt \right)/|\mathcal{I}|^2,
\]

where \( \theta = k \) for the linear system and \( \theta = (k, d)^T \) for the nonlinear system; \( d\text{proj}_\mathbf{u} \gamma(t)/dt \) is
the observed velocity profile; \( \frac{d\hat{\gamma}(t)}{dt} \) is the velocity profile of the dynamical system solved for parameters \( \theta \), with target \( x^* = \text{proj}_u \gamma(t_b) - \text{proj}_u \gamma(t_a) \), and with initial conditions \((x, \dot{x}) = (\text{proj}_u \gamma(t_a), d\text{proj}_u \gamma(t_a)/dt)\); \( I \) is the discrete grid of time points in \( I \) which is associated with the pellet positions in the given observed movement trace \( \gamma(t) \); and \( |I| \) is the cardinality of \( I \). By definition, if the ODE solver cannot solve the differential equation for parameters \( \theta \), then \( f(\theta) = \infty \).

Figure 25 shows the sample distribution of log error for the linear and nonlinear systems whose trajectories we fit to the tongue dorsum raising and lowering velocity profiles observed in the X-ray microbeam dataset. The logarithmic scale is used for visualization. Comparison shows that the nonlinear system had smaller objective function evaluations and thus fit the observed velocity profiles more closely than the linear system.

We illustrate the closeness of fit which the nonlinear system achieves by examining the worst, average, and best fit of the nonlinear system to the tongue dorsum raising and lowering velocity profiles observed in the X-ray microbeam dataset. Figure 26 graphs the velocity profiles of optimized linear and nonlinear system trajectories against the observed tongue dorsum raising and lowering velocity profiles. The left, middle, and right columns are associated with the minimum, median, and maximum error for the nonlinear system over the 80 observations of the X-ray microbeam dataset.

Averaging over all velocity profiles of the X-ray microbeam dataset, we find a mean±SD proportional time to peak velocity of 0.49 ± 0.07. These measurements agree with earlier findings of nearly symmetric velocity profiles of jaw lowering (Ostry et al., 1987), tongue dorsum movement in vowels (Munhall et al., 1985), glottal abduction (Munhall et al., 1985), and labial constriction (Byrd & Saltzman, 1998).

**Isochronous speech task**

This section sets up a model of an isochronous speech task using the proposed intrinsic gestural dynamics of Equation 17. We base the model on electromagnetic
articulography (EMA) recordings from three adult native English speakers AP, JB, LK of the Harvard-Haskins database of Regularly-Timed Speech (Patel, Löfqvist, & Naito, 1999). Each speaker has four recordings of the consonant-vowel [ba] alternating with each of [ba], [pa], and [ma]. The disyllable repeats eight times in each recording (e.g., ‘ba-ma-ba-ma-ba-ma-ba-ma-ba-ma-ba-ma-ba-ma-ba-ma-ba-ma’). See Figure 27 for lip aperture graphed as a function of time for an example trial. The instructions were to produce syllables as evenly spaced in time as possible. Thus, the hypothetical gestural score for lip aperture in Figure 28 has alternating lip aperture constriction and release gestures and is periodic in time with period 2p.

We characterize the isochrony constraint as a process which is periodic in time. This process couples uni-directionally to the intrinsic gestural dynamics, as modeled by Equation 17. The model neglects the influence of the vocal tract on the parts of the central nervous system involved in the isochrony task and retains only the uni-directional coupling of those parts of the central nervous system to the vocal tract. This makes the task constraint a periodic external force \( \sin(\omega t) \) on the lip aperture task variable \( x \). This force drives the damped task variable in the potential \( V(x) \) of Equation 19. The resulting equation of motion is

\[
\ddot{x} + b\dot{x} + \nabla\left(kx^2/2 - dx^4/4\right) = \Gamma \sin \omega t,
\]

where \( \omega < \sqrt{k/m} \) because the periodic external force is slower than the natural frequency of the intrinsic gestural dynamics.

This is a nonautonomous system. This system is chaotic for \( \Gamma \) large enough that \( x \) visits the anharmonic part of the potential \( V(x) \) of Equation 19 (i.e., the edge of the basin of attraction) (Huberman & Crutchfield, 1979). Figure 29 plots the chaotic attractor in the projection of the three-dimensional space onto the \((x, \dot{x})\) phase plane. The trajectories can and do cross as shown in Figure 29. Amplitude and peak velocity vary from cycle to cycle and make the trajectories aperiodic. The Hooke diagram (acceleration graphed against displacement) of Figure 30 shows N-shaped curves. The Hooke diagram of the harmonic
oscillator is a straight line; see Figure 10. The N-shapes in Figure 30 indicate anharmonicity (Mottet & Bootsma, 1999).

These predictions of Equation 21 are consistent respectively with the phase portraits and Hooke diagrams of Figures 31 and 32 derived from the recordings of the Harvard-Haskins database of Regularly-Timed Speech (Patel et al., 1999). In particular, the quasi-ellipsoidal phase portraits resemble the phase portrait projection of Figure 29 and the N-shaped Hooke diagrams resemble the Hooke diagrams of Figure 30.

This section has emphasized that intrinsic gestural timing does not exclude the gesture coupling to other coordinative structures. In fact, some theories of inter-gestural coordination involve the uni-directional coupling of planning oscillators to gestures (Saltzman & Byrd, 2000; Nam & Saltzman, 2003; Nam, 2007). Planning oscillators determine when speech gestures start and stop controlling task variables, but the task variable state does not influence planning oscillators. Of course, gestures can exert mutual influence on each other, as indicated by the compensation of one gesture to the perturbation of another (Kelso, Tuller, Vatikiotis-Bateson, & Fowler, 1984). Applications of the nonlinear autonomous dynamical system proposed in this paper may treat such interacting gestures as bidirectionally coupled autonomous dynamical systems. Furthermore, systems external to the gesture (but not necessarily external to the phonological system) may determine variability in the duration and coordination of gestures (Turk & Shattuck-Hufnagel, 2014; Tilsen, 2015). Evidence that at least some of these interactions involve bidirectional coupling between the gesture and larger units such as syllables and feet comes from correlations between variability in the relative timing of gestural movements and variability in the timing of larger units (Tilsen, 2009).

**Conclusion**

We follow Fowler (1980) and Fowler et al. (1980) in connecting autonomous dynamical systems with the theory of intrinsic timing and self-organization in coordinative
structures (Turvey, 1977; Greene, 1972; Easton, 1972). This connection is the theoretical basis for the proposal of this paper. The proposal is that the gesture is a nonlinear autonomous dynamical system with an anharmonic monostable potential. The damped motion of a task variable in this potential agrees qualitatively and quantitatively with the observed kinematics of speech gestures.
References


Fitch, W. T., & Giedd, J. (1999). Morphology and development of the human vocal tract:


and Performance, 10(6), 812.


Tilsen, S. (2009). Multitimescale dynamical interactions between speech rhythm and


Table 1

<table>
<thead>
<tr>
<th>[a'ka]</th>
<th>11, 12, 13, 14, 15, 16, 18, 19, 21, 24, 26, 27, 28, 29, 30, 31, 33, 34, 35, 36, 37, 39, 40, 43, 45, 46, 48, 49, 51, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62</th>
</tr>
</thead>
<tbody>
<tr>
<td>[a'ka]</td>
<td>11, 12, 13, 14, 15, 16, 18, 19, 21, 24, 25, 26, 27, 28, 29, 30, 31, 33, 34, 35, 36, 37, 39, 40, 43, 44, 45, 46, 48, 49, 51, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62</td>
</tr>
<tr>
<td>[a'ga]</td>
<td>11, 12, 14, 15, 16, 18, 19, 20, 21, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 39, 40, 44, 45, 46, 48, 49, 51, 53, 55, 56, 57, 58, 59, 60, 61, 62</td>
</tr>
<tr>
<td>[a'ga]</td>
<td>12, 14, 15, 16, 18, 19, 20, 21, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 39, 40, 44, 45, 46, 48, 49, 51, 53, 55, 56, 57, 58, 59, 60, 61, 62</td>
</tr>
</tbody>
</table>

*Speakers which produced each target item*
Figure 1. Harmonic oscillator displacement (solid) and velocity (dashed)

Figure 2. Harmonic oscillator trajectories in the phase plane
Figure 3. Damped harmonic oscillator displacement (solid) and velocity (dashed)

Figure 4. Damped harmonic oscillator trajectories in the phase plane
Figure 5. Driven Duffing oscillator displacement (solid) and velocity (dashed)

Figure 6. Driven Duffing oscillator trajectory in the phase plane
Figure 7. Critically damped harmonic oscillator displacement (solid) and velocity (dashed)

Figure 8. Critically damped harmonic oscillator phase plane
Figure 9. Harmonic potential

Figure 10. Hooke’s law: the restoring force is a linear function of distance $x$ from equilibrium (here assumed to be $x^* = 0$)
Figure 11. Step activation

Figure 12. Solution to Equation 11 with the step pattern of activation

Figure 13. Proportional time to peak velocity for step activation
Figure 14. Continuous activation

Figure 15. Solution to Equation 11 with the continuous pattern of activation

Figure 16. Change in proportional time to peak velocity for step activation (dashed) and different continuous patterns of activation (solid)
Figure 17. Stiffness function

Figure 18. Harmonic potential (dashed) and anharmonic potential (solid)
Figure 19. Change in proportional time to peak velocity for step activation with $d = 0$ (dashed), $d = 0.7 \, k/m$ (dotted), and $d = 0.95 \, k/m$ (solid).

Figure 20. Peak velocity against amplitude in harmonic potential ($d = 0$) and increasingly anharmonic potentials ($d$ increasing).
Figure 21. Amplitude-normalized peak velocity against movement time for the linear system. Dots correspond to different movement amplitudes. Lines correspond to different activation functions (see labels).

Figure 22. Amplitude-normalized peak velocity against movement time for the nonlinear system. Dots correspond to different movement amplitudes.
Figure 23. Angle of principal component of movement relative to the line formed by the intersection of the midsagittal and maxillary occlusal planes for tongue dorsum raising (left) and lowering (right)
Figure 24. Displacement along the principal component of movement for tongue dorsum raising (left) and lowering (right)
Figure 25. Histogram of log objective function value for the linear system (top row) and nonlinear system (bottom row) for tongue dorsum raising (left column) and for tongue dorsum lowering (right column)
Figure 26. Minimum (left column), median (middle column), and maximum (right column) error for tongue dorsum raising (top row) and tongue dorsum lowering (bottom row) in the X-ray microbeam dataset.
Figure 27. Lip aperture graphed as a function of time for an example trial of the isochronous speech task.
Figure 28. 2\(p\)-periodic gestural score for the isochronous speech task
Figure 29. Phase portrait of the driven nonlinear dynamics

Figure 30. Hooke diagram of the driven nonlinear dynamics
**Figure 31.** Lip aperture (horizontal axis) and its velocity (vertical axis) for speakers AP (rows 1-3), JB (rows 4-6), and LK (7-9) producing four trials each of ‘papa’, ‘bapa’, and ‘mapa’
Figure 32. Lip aperture (horizontal axis) and its acceleration (vertical axis) for speakers AP (rows 1-3), JB (rows 4-6), and LK (7-9) producing four trials each of ‘papa’, ‘bapa’, and ‘mapa’