

The gesture as an autonomous nonlinear dynamical system

February 10, 2016

1 Introduction

Extrinsic timing of speech gestures has been proposed to correct for disparities between theoretically predicted and experimentally observed velocity profiles. Extrinsic timing makes the dynamics of the gesture nonautonomous. This is theoretically undesirable. We argue that, despite making the 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 restores the autonomy of the gesture. This 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.

Sections 2 and 3 distinguish autonomous from nonautonomous and nonlinear from linear theories of the speech gesture. We then propose that the gesture is a nonlinear autonomous dynamical system. Sections 4 and 5 qualitatively and quantitatively evaluate the proposed dynamical system. Section 6 shows that, in an isochronous speech task (e.g., ‘bapabapa...’) where the dynamical system is driven by a periodic external force, solutions become aperiodic.

2 Autonomous versus nonautonomous

Theoretical work suggests that the gesture has intrinsic timing [7]. This means that the gesture intrinsically has both spatial and temporal extent and that the gesture does not merely inherit temporal extent from an external system. In contrast, an extrinsic timing theory of the gesture would mean that the gesture intrinsically has only spatial extent and that the gesture inherits temporal extent from an executive time-keeper. This section introduces the autonomous versus nonautonomous distinction as it has been applied to gestures and defines terms to be used in forthcoming sections.

Autonomous dynamical systems have the form of Equation 1.

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

The evolution of x as a function of time, $x(t)$, also known as a trajectory of x , is completely determined by the equation above. Specifically, this equation states that at any time instant t the rate of change of x , $\dot{x} = dx/dt$, is a function of x , $f(x)$, but not also of t . We can fully describe the behavior of the system by considering just three cases. If $f(x)$ is positive, then x will increase. If $f(x)$ is negative, then x will decrease. Finally, if $f(x)$ is zero, x stays the same (this x is called a fixed point of the dynamics). Thus, $f(x)$ points in the direction of change for x and specifies the magnitude of the change. Therefore, $f(x)$ is known as a force field of the dynamical system. The force field of an autonomous dynamical system is stationary (i.e., does not vary over time). As a consequence of stationarity, no two trajectories cross [5]. The coordinates of the state in the phase space fully determine the evolution of the system. 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.

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

This means that the evolution of the system from a given point in the state space depends not only on the coordinates of that point, but also on the time at which that point was reached. The force field of a nonautonomous dynamical system 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.

A gesture is a dynamical system which controls a functionally relevant variable x (e.g., a tract variable [27, 2], or a task variable more generally [26]). This dynamical system makes x move in the potential

$$V(x) = kx^2/2, \tag{3}$$

where k is the stiffness of the gesture and x is displacement from the target of the gesture (Figure 1). The dynamical law for x moving in the potential of Equation 3 is given in Equation 4. This is the well-known equation of the damped linear spring. It is an autonomous dynamical system which can be made¹ to have the form of Equation 1. When $b = 2\sqrt{mk}$, the spring is critically damped and x approaches zero and does not oscillate about zero ([8], p. 396).

$$m\ddot{x} + b\dot{x} + \nabla(kx^2/2) = 0. \tag{4}$$

Equation 4 derives from Newton's law, $m\ddot{x} = F$, where the force F is the sum of two forces: a restoring force $F_1 = -kx$ (Hooke's law, see Figure 2), which is a function of the distance x from equilibrium, and a damping force $F_2 = -b\dot{x}$, which is a function of velocity \dot{x} . The former force represents elasticity. The

¹The second-order Equation 4 can be rewritten as a system of two first-order equations by introducing a new state variable y and setting $\dot{x} = y$, along with some rearrangement of terms in the original second-order equation. The dynamical system has become two-dimensional with state space coordinates x and y

farther the displacement away from equilibrium, the greater the restoring force. The latter represents viscosity. In viscous behavior, the force depends (only) on velocity, not on displacement (cf. [15], pp. 82-83). Both F_1 and F_2 are independent of time and are linear.

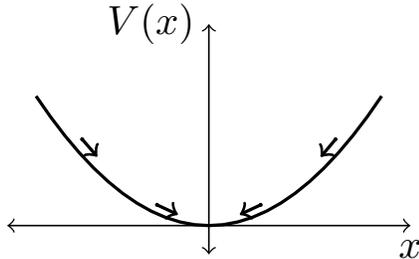


Figure 1: Harmonic potential

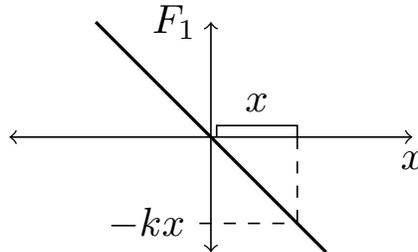


Figure 2: Hooke's law

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

$$m\ddot{x} + a(t)\left(\underbrace{b\dot{x}}_{-F_2} + \underbrace{\nabla(kx^2/2)}_{-F_1}\right) = 0, \quad (5)$$

where $a(t)$ is the step function

$$a(t) = \begin{cases} 1 & t \in [t_a, t_b] \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

which activates the gesture over some time interval [27]. Figure 3 graphs activation over time, and Figure 4 graphs a particular solution to Equation 5 given this step pattern of activation.

Turning gestures on and off as shown above appears to make Equation 4 nonautonomous as Equation 5, which can be made to have the form of Equation 2. Nevertheless, Equation 5 is equivalent to the following piecewise autonomous system.

$$m\ddot{x} = \begin{cases} -b\dot{x} - kx & \text{for } t \in [t_a, t_b] \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

Over the interval $[t_a, t_b]$ we observe the intrinsic dynamics of the gesture which arises from the damped movement 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., k, b are constant).

Equation 5 with step activation fails to predict a property of speech movements, namely that near symmetry is typical of speech velocity profiles (i.e.,

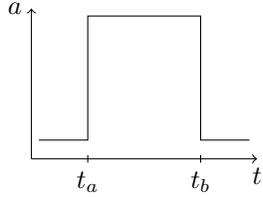


Figure 3: Step activation

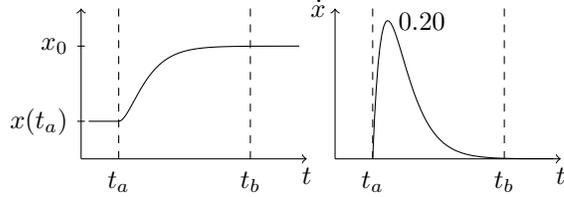


Figure 4: Solution to Equation 5 with the step pattern of activation

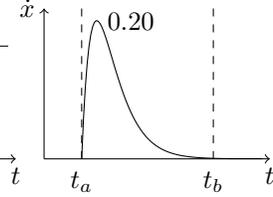


Figure 5: Proportional time to peak velocity for step activation

velocity graphed as a function of time). [20] 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 [20], tongue dorsum movement in vowels [17], glottal abduction [17], and labial constriction [3]. In contrast to these reports, Equation 5 with step activation predicts proportional time to peak velocity of 0.20 (see Figure 5). The reported deviations from 0.20 indicate that we need a correction to Equation 5 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 in [24] and a formal implementation was proposed in [13, 3, 4]. Whereas step activation makes the gesture autonomous during its interval of activation, these proposals make the gesture nonautonomous during this interval. For instance, Equation 8 defines activation as a continuous function of time [13].

$$a(t) = \begin{cases} 0, & \text{if } t < t_a \\ \sin\left(\frac{2\pi(t-t_a)}{4(t_b-t_a)}\right) & \text{if } t_a \leq t < t_b \\ 1, & \text{if } t_b \leq t < t_c \\ \sin\left(\frac{2\pi(t-t_d)}{4(t_c-t_d)}\right) & \text{if } t_c \leq t < t_d \\ 0, & \text{if } t \geq t_d \end{cases} \quad (8)$$

Figure 6 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 7 graphs a particular solution to Equation 5 given the continuous pattern of activation.

Continuous activation makes the solution to Equation 5 take on a range of proportional times to peak velocity for varying t_b and t_c in Equation 8 (i.e., for varying activation frequency). Examples are the solid velocity profiles of Figure 8. These are corrections to Equation 5 with step activation, which has proportional time to peak velocity of 0.20 (see the dashed velocity profile of

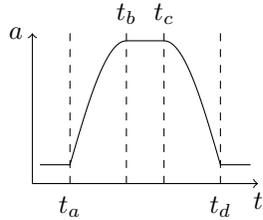


Figure 6: Continuous activation

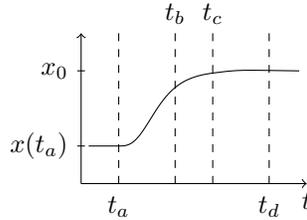


Figure 7: Solution to Equation 5 with the continuous pattern of activation

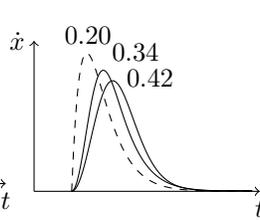


Figure 8: Change in proportional time to peak velocity for step activation (dashed) and different continuous patterns of activation (solid)

Figure 8).

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 7, the linear system with continuous activation does not admit a piecewise autonomous definition. Section 3 proposes a different correction for short proportional time to peak velocity which keeps the gesture autonomous and uses step activation. The intrinsic gestural dynamics reflects only the autonomous, damped movement of x in the potential $V(x)$.

3 Nonlinear versus linear

Equation 5 with step activation is the simplest dynamical system for the discrete goal-directed movement task. 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, despite keeping with damping, introduces oscillations even without external periodic forcing. This property is undesirable for the discrete goal-directed movement task. 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.

$$m\ddot{x} + b\dot{x} + \nabla V(x) = 0 \quad (9)$$

Any potential other than the potential of the harmonic oscillator $V(x) = kx^2/2$ is known as anharmonic ([10], p. 108). This section argues that the gesture has an anharmonic potential. Specifically, we propose that a negative

quadratic term introduces a correction to the harmonic potential of Equation 3. This corrects for the short proportional time to peak velocity of Equation 5 with step activation.

$$m\ddot{x} + b\dot{x} + \nabla(kx^2/2 - dx^4/4) = 0 \quad (10)$$

Figure 9 graphs the force

$$F(x) = -\nabla V(x) = -kx + dx^3 \quad (11)$$

as a function of displacement. In the neighborhood of the stable fixed point $x = 0$, the linear term attracts x to $x = 0$. For small displacement, the cubic deviation 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. Increasing the degree of the force polynomial to higher than cubic (higher than quadratic 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 the stable fixed point $x = 0$.

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

$$V(x) = kx^2/2 - dx^4/4 \quad (12)$$

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 10. 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 10 against those of the linear dynamical system of Equation 4.

4 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 (i) that proportional time to peak velocity is nearly symmetric, (ii) that amplitude and peak velocity covary nonlinearly, and (iii) 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). In contrast, the linear system makes the first prediction but not

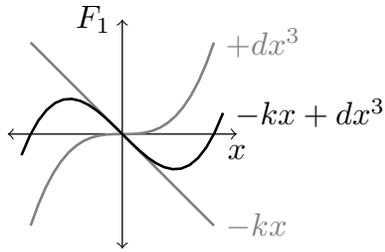


Figure 9: Stiffness function

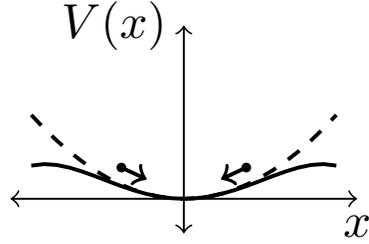


Figure 10: Harmonic potential (dashed) and anharmonic potential (solid)

the second and third predictions, if corrected with a continuous activation function, and makes none of the predictions otherwise. Evidence for these kinematic relationships comes from observation of oral and laryngeal speech gestures, suggesting shared principles of self-organization.

Figure 11 compares representative proportional times to peak velocity of the nonlinear Equation 10 with that of the linear Equation 5 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 in Section 2.

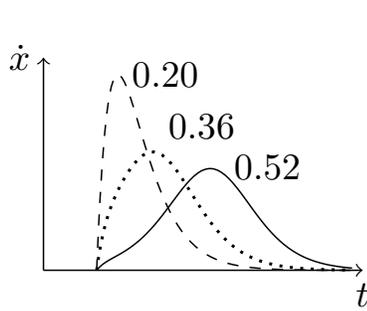


Figure 11: Change in proportional time to peak velocity for step activation with $d = 0$ (dashed), $d = 0.7 k/m$ (dotted), and $d = 0.95 k/m$ (solid)

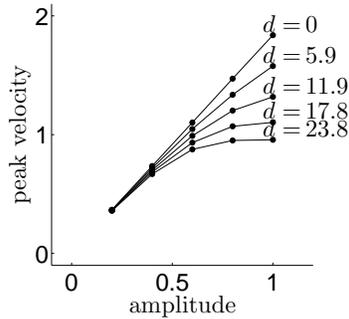


Figure 12: Peak velocity against amplitude in harmonic potential ($d = 0$) and increasingly anharmonic potentials (d increasing)

Figure 12 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 and nonlinear systems are consistent with the finding that peak velocity and movement amplitude correlate positively for tongue dorsum raising and lowering ([22], p. 644; [21], pp. 629-630; [17], p. 467) and vocal fold adduction and abduction ([17], 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 ([22], p. 644).

Figures 13 and 14 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 of the nonlinear system, respectively. As settling time increases, amplitude-normalized peak velocity decays quadratically 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 [22, 17]. This relation is described by an equation of constraint in the sense of [8] (p. 401). This equation is $(\text{peak velocity})/(\text{amplitude}) = c/(\text{settling time})$, 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 [22, 17] and the glottis [17]. In contrast, the linear system does not entail this equation of constraint, with or without continuous activation.²

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. Autonomy and nonlinearity achieve both.

5 Quantitative evaluation

We compare the nonlinear Equation 10 with anharmonic potential against the linear Equation 5 with harmonic potential using a dataset taken from the X-ray Microbeam Speech Production Database [30]. The X-ray Microbeam Speech Production Database consists of 57 subjects carrying out 118 speech or speech-related tasks. We analyze Task 16, a citation task with target items [a'ka] and [a'ga]. Forty-three of the 57 total subjects performed Task 16. The tongue dorsum (T4) pellet of two subjects was mistracked by the X-ray microbeam system over the interval from the release of [k] to the achievement of the oral

²[22, 17] note that the equation of constraint is a consequence of an undamped harmonic oscillator. However, the speech gesture needs damping terms to have discrete goal-directed movements as its solutions.

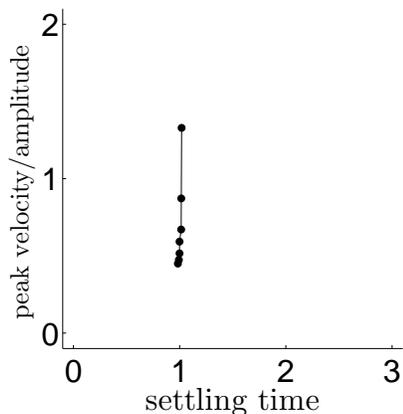


Figure 13: Amplitude-normalized peak velocity against movement time. Dots correspond to different movement amplitudes in the harmonic potential ($d = 0$)

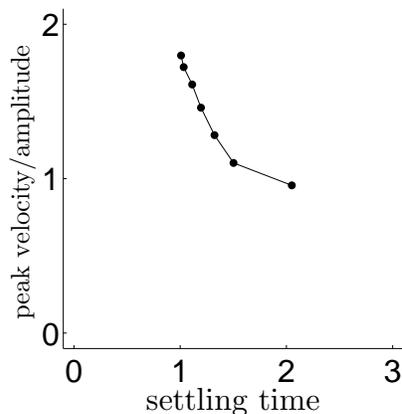


Figure 14: Amplitude-normalized peak velocity against movement time. Dots correspond to different movement amplitudes in the anharmonic potential ($d > 0$)

[a'ka]	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
[a'ga]	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

Table 1: Subjects which produced each target item

target of [a] and the same happened for four subjects in [a'ga] target items (see [30] Section 6.7.1 for probable causes of mistracking). Table 1 lists the numbers of the subjects whose data is available for each of the two target items. We use the T4 pellet in the 80 resulting X-ray microbeam recordings from 43 different speakers to compare the observed tongue dorsum lowering kinematics with predictions of the linear and nonlinear dynamical systems.

Pellet positions are expressed in a two-dimensional cranial coordinate system (see [30] 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.

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 $\gamma(t)$, $t \in I$ onto one dimension is to project it onto the principal component \mathbf{u} of movement during the interval $I = [t_a, t_b]$, where t_a and t_b are the times at which $\frac{d}{dt}\gamma(t)$

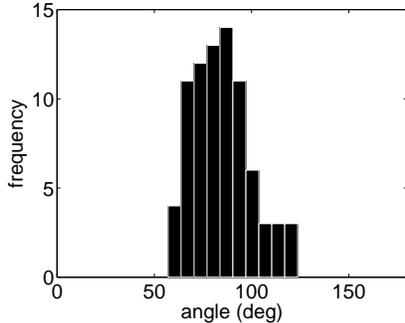


Figure 15: Angle of principal component of movement relative to the line formed by the intersection of the midsagittal and maxillary occlusal planes

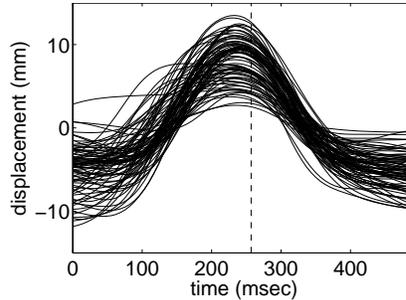


Figure 16: Displacement along the principal component of movement

falls below and rises above, respectively, $0.2 \min(\frac{d}{dt} \gamma(t))$, $t \in I$. The minimum is used because the superior-inferior coordinate of the T4 pellet is negative-going. The histogram of Figure 15 shows that the principal components \mathbf{u} are nearly parallel to the superior-inferior axis (90°) for most target items. Figure 16 graphs the 80 observations of $\text{proj}_{\mathbf{u}} \gamma(t)$, centered on the time of onset of movement toward the vowel target.

We estimate the velocity profile $\frac{d}{dt} \text{proj}_{\mathbf{u}} \gamma(t)$ by fitting sixth order smoothing splines to $\text{proj}_{\mathbf{u}} \gamma(t)$, differentiating analytically, and evaluating the derivative at the time points associated with the pellet positions in $\gamma(t)$ (see [30] Section 6.2 for the interpolation and resampling procedure).

We use the simplex search method of [14] as implemented in MATLAB to optimize the parameters of the linear and nonlinear systems. The objective function which we optimized is

$$f(\boldsymbol{\theta}) = \sum_{t \in \mathcal{I}} \left(\left(\frac{d}{dt} \text{proj}_{\mathbf{u}} \hat{\gamma}(t) - \frac{d}{dt} \text{proj}_{\mathbf{u}} \gamma(t) \right) / |\mathcal{I}| \right)^2, \quad (13)$$

where $\boldsymbol{\theta} = k$ for the linear system and $\boldsymbol{\theta} = (k, d)^\top$ for the nonlinear system; $\frac{d}{dt} \text{proj}_{\mathbf{u}} \gamma(t)$ is the observed velocity profile; $\frac{d}{dt} \text{proj}_{\mathbf{u}} \hat{\gamma}(t)$ is the velocity profile of the differential equation solved for parameters $\boldsymbol{\theta}$, with target $x_0 = \text{proj}_{\mathbf{u}} \gamma(t_b) - \text{proj}_{\mathbf{u}} \gamma(t_a)$, and with initial conditions $(x, \dot{x}) = (\text{proj}_{\mathbf{u}} \gamma(t_a), \frac{d}{dt} \text{proj}_{\mathbf{u}} \gamma(t_a))$; \mathcal{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 $|\mathcal{I}|$ is the cardinality of \mathcal{I} . By definition, if the ODE solver cannot solve the differential equation for parameters $\boldsymbol{\theta}$, then $f(\boldsymbol{\theta}) = \infty$.

Figure 17 has the sample distribution of $\log f$ for the linear system and Figure 18 has the sample distribution of $\log f$ for the nonlinear system. The logarithmic scale is used for visualization. Comparison shows that the nonlinear

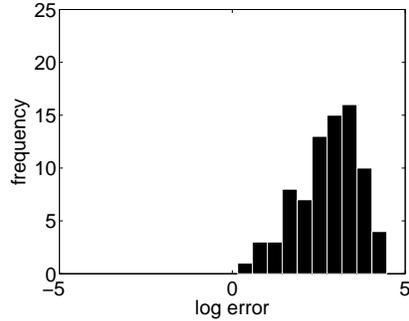


Figure 17: Histogram of log objective function value for the linear system

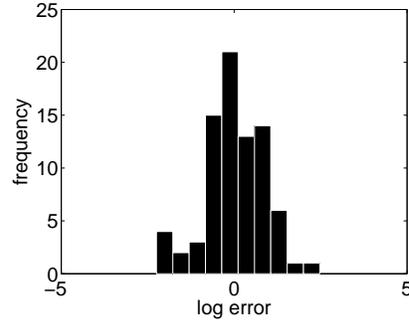


Figure 18: Histogram of log objective function value for the nonlinear system

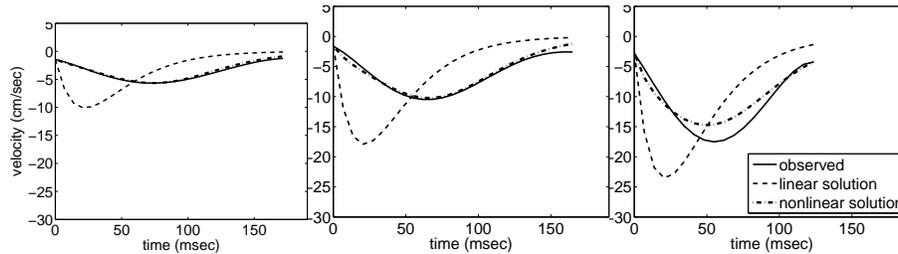


Figure 19: Minimum error Figure 20: Median error Figure 21: Maximum error

system fits the observed velocity profiles better in the sense that the nonlinear system had smaller objective function evaluations.

Figures 19, 20, and 21 graph the solutions to optimized linear and nonlinear systems against the observed velocity profiles. Figures 19, 20, and 21 are associated with the minimum, median, and maximum values of f for the nonlinear system over the 80 observations. The observed velocity profiles have mean proportional time to peak velocity of 0.44 (SD = 0.05), a measure which the solutions to the nonlinear — but not the linear — system match.

6 Model of an isochronous speech task

In general, characterizing a dynamical system as autonomous or nonautonomous depends entirely on the frame of reference [27]. This is particularly so in the case of uni-directional coupling. A system A of m equations which couples uni-directionally to another system B of n equations is equally well described as an autonomous system of $m + n$ equations or as a nonautonomous system of n equations. The former approach involves solving the uni-directionally coupled system of $m + n$ equations, while the latter approach involves solving the m

equations of system A, writing the uni-directional coupling of system A to system B as a function of time, and then solving the n equations of system B with the coupling term given as a function of time [1]. We follow the latter approach in the below case of uni-directional coupling.

This section sets up a model of an isochronous speech task using the proposed intrinsic gestural dynamics of Equation 10. 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 [23]. Each subject 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’). The instructions were to produce syllables as evenly spaced in time as possible. Thus, a hypothetical gestural score for lip aperture has alternating lip aperture constriction and release gestures and is periodic in time with period $2p$ (Figure 22).

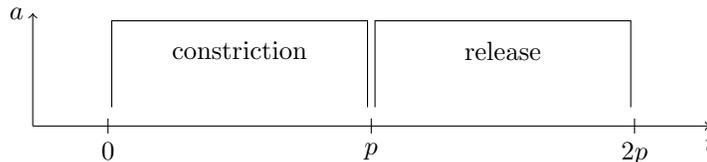


Figure 22: $2p$ -periodic gestural score for the isochronous speech task

The isochrony constraint couples uni-directionally to this dynamical system. 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 12. The resulting equation of motion is

$$m\ddot{x} + b\dot{x} + \nabla(kx^2/2 - dx^4/4) = \Gamma \sin \omega t, \quad (14)$$

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.³ Hence, its phase space is now the three-dimensional space (x, \dot{x}, t) . This system is chaotic for Γ large enough that x visits the anharmonic part of the potential $V(x)$ of Equation 12 (i.e., the edge of the basin of attraction) [11]. Figure 23 plots the chaotic attractor in the projection of the three-dimensional phase space onto the (x, \dot{x}) plane. In the three-dimensional space, the trajectories do not cross. The addition of the t axis in the phase space endows the system with the no crossing property. However, in the two-dimensional projection, the trajectories can and do cross

³This nonautonomous second-order equation can be rewritten as a system of three first-order autonomous equations by setting $\dot{x} = y$ and declaring t as one of the variables by the equation $\dot{t} = 1$.

as shown in Figure 23. Amplitude and peak velocity vary from cycle to cycle and make the solutions aperiodic. The Hooke diagram (acceleration graphed against displacement) of Figure 24 shows N-shaped curves. The Hooke diagram of the harmonic oscillator is a straight line; see Figure 2, remembering all along that force is mass multiplied by acceleration. The N-shapes in Figure 24 indicate anharmonicity [16].

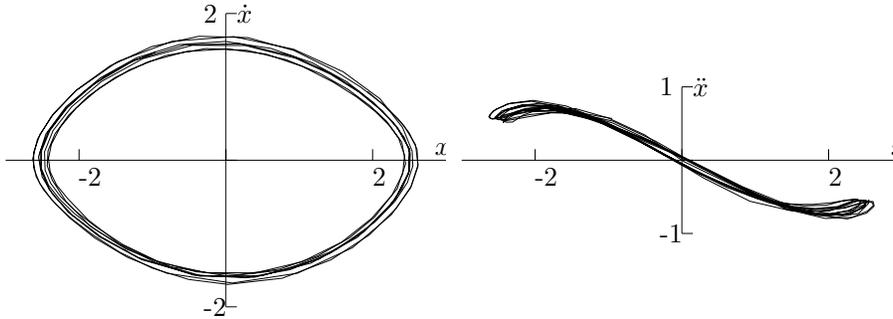


Figure 23: Phase portrait

Figure 24: Hooke diagram

These predictions of Equation 14 are consistent respectively with the phase portraits and Hooke diagrams of Figures 25 and 26 derived from the recordings of the Harvard-Haskins database of Regularly-Timed Speech [23]. In particular, the quasi-ellipsoidal phase portraits resemble the phase portrait projection of Figure 23 and the N-shaped Hooke diagrams resemble the Hooke diagrams of Figure 24.

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 [25, 19, 18]. Planning oscillators determine when speech gestures start and stop controlling task variables, but task variables do 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 [12]. Furthermore, systems external to the gesture determine variability in the duration and coordination of gestures [28]. Models of such systems are bi-directionally coupled autonomous dynamical systems.

7 Conclusion

We follow [7, 8] in connecting autonomous dynamical systems with the theory of intrinsic timing and self-organization in coordinative structures [29, 9, 6]. 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

- [1] Peter J Beek, Michael T Turvey, and RC Schmidt. Autonomous and nonautonomous dynamics of coordinated rhythmic movements. *Ecological Psychology*, 4(2):65–95, 1992.
- [2] Catherine P Browman and Louis M Goldstein. Gestural specification using dynamically-defined articulatory structures. *Journal of Phonetics*, 18:299–320, 1990.
- [3] Dani Byrd and Elliot Saltzman. Intragestural dynamics of multiple prosodic boundaries. *Journal of Phonetics*, 26(2):173–199, 1998.
- [4] Dani Byrd and Elliot Saltzman. The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics*, 31(2):149–180, 2003.
- [5] Earl A Coddington and Norman Levinson. *Theory of ordinary differential equations*. Tata McGraw-Hill Education, 1955.
- [6] Thomas A Easton. On the normal use of reflexes: The hypothesis that reflexes form the basic language of the motor program permits simple, flexible specifications of voluntary movements and allows fruitful speculation. *American Scientist*, pages 591–599, 1972.
- [7] Carol A Fowler. Coarticulation and theories of extrinsic timing. *Journal of Phonetics*, 8(1):113–133, 1980.
- [8] Carol A Fowler, Philip Rubin, Robert E Remez, and ME Turvey. Implications for speech production of a general theory of action. *Language Production: Speech and Talk*, 1, 1980.
- [9] Peter H Greene. Problems of organization of motor systems. *Progress in theoretical biology*, 2:123–145, 1972.
- [10] Herman Haken. Introduction to synergetics. In *Synergetics*, pages 9–19. Springer, 1973.
- [11] Bernardo A Huberman and James P Crutchfield. Chaotic states of anharmonic systems in periodic fields. *Physical Review Letters*, 43(23):1743, 1979.
- [12] J Scott Kelso, Betty Tuller, E Vatikiotis-Bateson, and Carol A Fowler. Functionally specific articulatory cooperation following jaw perturbations during speech: evidence for coordinative structures. *Journal of Experimental Psychology: Human Perception and Performance*, 10(6):812, 1984.

- [13] Bernd J Kröger, Georg Schröder, and Claudia Oppen-Rhein. A gesture-based dynamic model describing articulatory movement data. *The Journal of the Acoustical Society of America*, 98(4):1878–1889, 1995.
- [14] Jeffrey C Lagarias, James A Reeds, Margaret H Wright, and Paul E Wright. Convergence properties of the nelder–mead simplex method in low dimensions. *SIAM Journal on optimization*, 9(1):112–147, 1998.
- [15] Alexander R McNeill. *Exploring biomechanics*. Scientific American Library; Distributed by WH Freeman, 1992.
- [16] Denis Mottet and Reinoud J Bootsma. The dynamics of goal-directed rhythmical aiming. *Biological cybernetics*, 80(4):235–245, 1999.
- [17] Kevin G Munhall, David J Ostry, and Avraham Parush. Characteristics of velocity profiles of speech movements. *Journal of Experimental Psychology: Human Perception and Performance*, 11(4):457, 1985.
- [18] Hosung Nam. *A gestural coupling model of syllable structure*. PhD thesis, Yale University, 2007.
- [19] Hosung Nam and Elliot Saltzman. A competitive, coupled oscillator model of syllable structure. In *Proceedings of the 15th international congress of phonetic sciences*, volume 1, pages 2253–2256, 2003.
- [20] David J Ostry, James D Cooke, and Kevin G Munhall. Velocity curves of human arm and speech movements. *Experimental Brain Research*, 68(1):37–46, 1987.
- [21] David J Ostry, Eric Keller, and Avraham Parush. Similarities in the control of the speech articulators and the limbs: Kinematics of tongue dorsum movement in speech. *Journal of Experimental Psychology: Human Perception and Performance*, 9(4):622, 1983.
- [22] David J Ostry and Kevin G Munhall. Control of rate and duration of speech movements. *The Journal of the Acoustical Society of America*, 77(2):640–648, 1985.
- [23] Aniruddh D Patel, Anders Löfqvist, and Walter Naito. The acoustics and kinematics of regularly timed speech: A database and method for the study of the p-center problem. In *Proceedings of the 14th international congress of phonetic sciences*, volume 1, pages 405–408, 1999.
- [24] Pascal Perrier, Christian Abry, and Eric Keller. Vers une modélisation des mouvements de la langue. *Bulletin de la Communication Parlée*, 2:45–63, 1988.
- [25] Elliot Saltzman and Dani Byrd. Task-dynamics of gestural timing: Phase windows and multifrequency rhythms. *Human Movement Science*, 19(4):499–526, 2000.

- [26] Elliot Saltzman and JAS Kelso. Skilled actions: a task-dynamic approach. *Psychological review*, 94(1):84, 1987.
- [27] Elliot L Saltzman and Kevin G Munhall. A dynamical approach to gestural patterning in speech production. *Ecological psychology*, 1(4):333–382, 1989.
- [28] Alice Turk and Stefanie Shattuck-Hufnagel. Timing in talking: what is it used for, and how is it controlled? *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 369(1658):20130395, 2014.
- [29] Michael T Turvey. Preliminaries to a theory of action with reference to vision. *Perceiving, acting and knowing*, pages 211–265, 1977.
- [30] John Westbury, Paul Milenkovic, Gary Weismer, and Raymond Kent. X-ray microbeam speech production database. *The Journal of the Acoustical Society of America*, 88(S1):S56–S56, 1990.

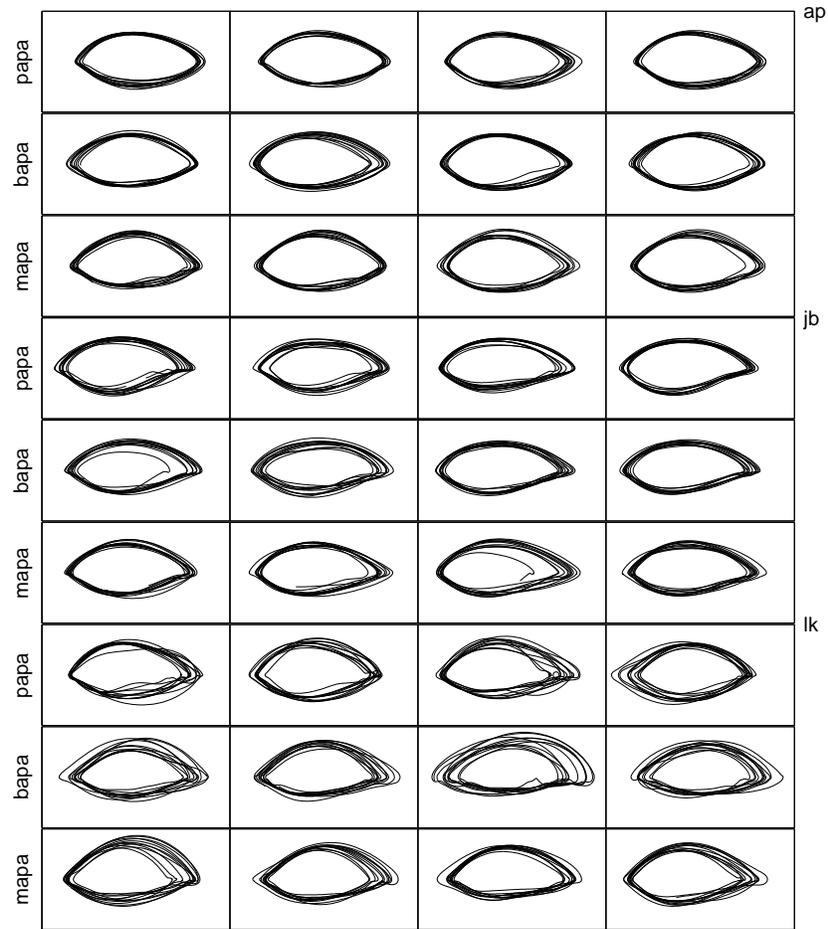


Figure 25: Lip aperture (horizontal axis) and its velocity (vertical axis) for subjects AP (rows 1-3), JB (rows 4-6), and LK (7-9) producing four trials each of 'papa', 'bapa', and 'mapa'

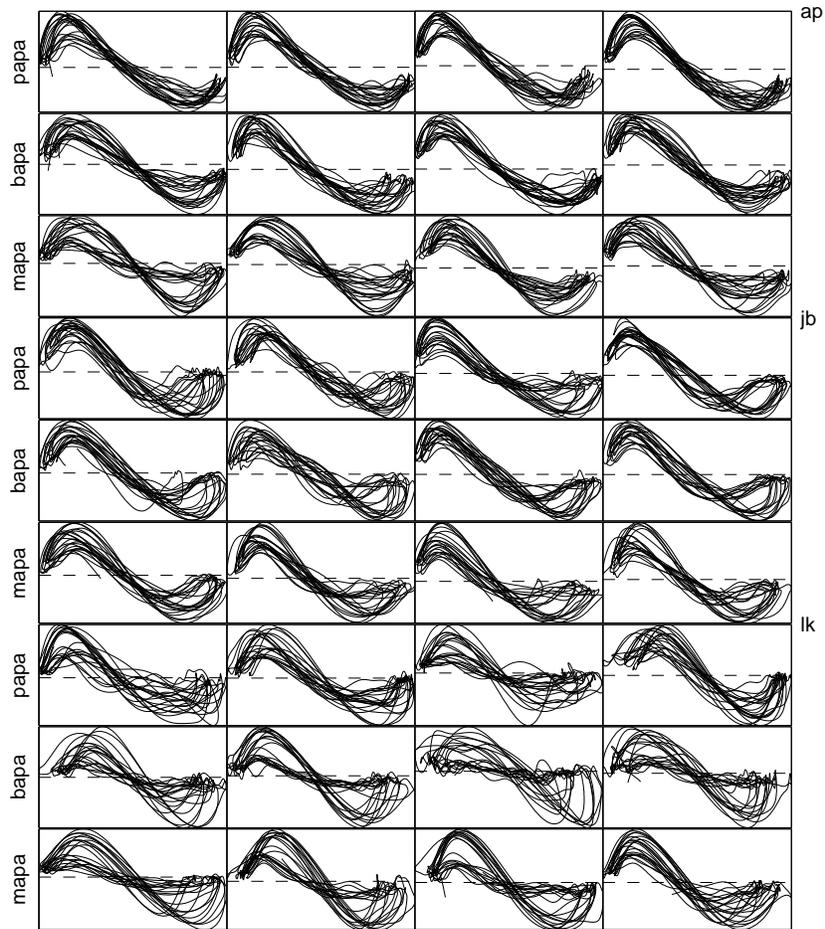


Figure 26: Lip aperture (horizontal axis) and its acceleration (vertical axis) for subjects AP (rows 1-3), JB (rows 4-6), and LK (7-9) producing four trials each of 'papa', 'bapa', and 'mapa'