Evaluation of a dynamical model of speech perception

PAMELA CASE, BETTY TULLER, MINGZHOU DING, and J. A. SCOTT KELSO
Florida Atlantic University, Boca Raton, Florida

Previous work (Tuller, Case, Ding, & Kelso, 1994) has revealed signature properties of nonlinear dynamical systems in how people categorize speech sounds. The data were modeled by using a two-well potential function that deformed with stimulus properties and was sensitive to context. Here we evaluate one prediction of the model—namely, that the rate of change of the potential's slope should increase when the category is repeatedly perceived. Judged goodness of category membership was used as an index of the slope of the potential. Stimuli from a "say"–"stay" continuum were presented with gap duration changing sequentially throughout the range from 0 to 76 to 0 msec, or from 76 to 0 to 76 msec. Subjects identified each token as either "say" or "stay" and rated how good an exemplar it was of the identified category. As predicted, the same physical stimulus presented at the end of a sequence was judged a better exemplar of the category than was the identical stimulus presented at the beginning of the sequence. In contrast, stimuli presented twice near the middle of a sequence with few (or no) stimuli between them, as well as stimuli presented with an intervening random set, showed no such differences. These results confirm the hypothesis of a context-sensitive dynamical representation underlying speech.

A basic issue in speech perception concerns how people sort a continuously varying acoustic signal into appropriate phonemic categories in a manner that is at once stable under acoustic variation and flexible in changing contexts. In a seminal study of the perceptual stability of speech, Liberman, Harris, Hoffman, and Griffith (1957) demonstrated the difficulty listeners have in discriminating between acoustic stimuli that are categorized as the same speech segment—the phenomenon known as categorical perception. A large body of work has followed and shown, among other things, that the assignment of an acoustic signal to a given category is flexible, adjusting with such factors as phonetic context, speaking rate, and linguistic experience (see Repp & Liberman, 1987, for a review). Typical studies involve a stimulus continuum that is created by systematically manipulating an acoustic variable. The stimuli from the continuum are randomized and presented to listeners for identification. When the continuum involves changes in consonant perception, subjects usually judge stimuli as belonging to the same consonantal category across relatively large variations of the acoustic parameter. Such stably perceived ranges are separated by a relatively narrow range of the acoustic parameter for which stimuli are sometimes heard as belonging to one category and sometimes

belonging to the other. The appearance of the identification functions obtained in such experiments indicates a rather abrupt and discontinuous change from one category to another. (See Figure 1 for an example.) The random presentation of stimuli that is designed to eliminate global contextual effects makes it difficult to investigate the dynamical processes underlying perceptual change.

In a recent study (Tuller, Case, Ding, & Kelso, 1994), we chose to examine directly dynamical processes in speech categorization. The stimulus continuum chosen, a "say"–"stay" continuum with variation of the silent gap duration after the initial fricative, is one for which categorical perception has previously been demonstrated (see, e.g., Best, Morongiello, & Robson, 1981). However, we varied the silent gap duration sequentially, not randomly, in order to explore the temporal evolution of phonemic categorization and change. The subject's task was to identify each stimulus as either "stay" or "say." Comparing the location of the category switch when gap duration increased from the shortest gap stimulus with the switch location when gap duration decreased from the longest gap stimulus can produce three patterns of results: (1) The boundary between responding "say" and responding "stay" remains fixed, regardless of whether the gap duration gets progressively longer or shorter, so that a given stimulus is always identified the same way (critical boundary). (2) In a pattern that exhibits an assimilative effect, a longer gap is necessary in order to induce a category change from "say" to "stay" when gap duration progressively lengthens than that which is necessary for the change from "stay" to "say" when gap duration shortens (hysteresis); the initial perceptual state persists even when the current gap duration favors the alternative re-

This research was supported by NIDCD Grant DC00411, NIMH Grant MH42900, and BRS Grant RR07258. We are grateful to Arthur Samuel and two anonymous reviewers for their comments on an earlier version of this manuscript. Correspondence concerning this article should be addressed to P. Case, Center for Complex Systems, Florida Atlantic University, P.O. Box 3091, Boca Raton, FL 33431-0991 (e-mail: case@walt.ccs.fau.edu).

Figure 1. Percent "stay" identification across subjects in the randomized forced-choice task. The acoustic parameter is duration of silent gap inserted between fricative "s" and vocalic "ay." Solid: sequential presentation. Dashed: mixed presentation.

sponse when stimuli are randomized. (3) In a pattern revealing a contrastive effect, the category change from "say" to "stay" occurs in response to a stimulus with a shorter gap when gap duration progressively lengthens than does the category change from "stay" back to "say" when gap duration shortens (enhanced contrast). The last two patterns occurred equally often and much more frequently than did the first pattern (41%, 42%, and 17% of runs, respectively, although the absolute frequencies are of course dependent on the size of the step change in gap duration).

The dependence of speech categorization on recent percepts and the direction of parameter change is strong evidence of nonlinearity and multistability. In what follows, we briefly describe the theoretical model proposed to capture the observed patterns of category change within a unified dynamical account. Then, we describe the model prediction evaluated in the present work.

A Dynamical Model of Categorization

A dynamical system is a system that evolves over time such that its present state always depends in a rule-governed way on previous states. Differential equations offer a mathematical description of the rate of change of states as time passes and parameters alter. Thus, a dynamical system is a system whose behavior is described by differential equations. An attractor can be loosely thought of as the observable behavior of a dynamical system. For more information on dynamical systems that is accessible to the general reader, see Abraham and Shaw (1992), Kelso, Ding, and Schöner, (1992), and van Gelder and Port (1995).

Implementing a dynamical theory means mapping the reproducibly observed states of a system—here the two perceptual categories "say" and "stay"—onto attractors

of a corresponding dynamical model. In this we follow essentially the same strategy as that employed by Haken, Kelso, and Bunz (1985) in studies of coordinated movement. For a single perceptual category, a local model containing a fixed point would be adequate. But the presence of hysteresis and enhanced contrast indicates that more than one stable percept coexists for a given acoustic stimulus, so a nonlinear dynamical model must be found. The percepts studied are parameterized by an acoustic variable, the silent gap duration between "s" and the synthetic syllable "ay." An initial perceptual form is established at either extreme of the parameter range. Sequential changes in parameter values eventually result in a switch to the alternative perceptual form. Notably, the initial category perceived may persist even when the current value of the gap duration favors the alternative response in a randomized stimulus sequence. Although the theoretical description was developed in a particular experimental context, we suspect that it applies across a range of acoustic cues or speech categories and may also generalize to categorization of ambiguous sentences and visual patterns (Ditzinger & Haken, 1989, 1990; Haken, 1990; Kelso et al., 1994).

The results of our previous work were modeled concisely by the following dynamical system (Tuller et al., 1994):

$$\begin{aligned} dx/dt &= -dV(x)/dx \\ &= -k + x - x^3. \end{aligned} \quad (1)$$

Rewriting the differential equation in the form of the potential function

$$V(x) = kx - x^2/2 + x^4/4, \quad (2)$$

where x is a variable characterizing the perceptual form and k is a parameter specifying the direction and degree of tilt for the potential, allows visualization of the behavior of the system as the parameter k is manipulated. Think of Equation 2 as describing the motion of a viscous point mass (a "sticky" ball) moving in the potential landscape $V(x)$ (such as one of those shown in Figure 2). The minima of the potential, the valleys in the landscape, are defined as attractors corresponding to the perceptual categories.

Figure 2 shows how the landscape changes for several values of k . With $k = -1$, only one stable point exists corresponding to a single category (e.g., "say"; Figure 2a). As k increases, the potential landscape tilts but otherwise remains unchanged in terms of the composition of attractor states (Figure 2b). However, when k reaches a critical point $k = -k_c$, a qualitative change in the attractor layout takes place. In other words, a bifurcation occurs. Bifurcations are particularly important for the investigation of complex behavior, because, near critical points, the essential processes governing a behavior's stability, flexibility, and selection can be uncovered (for discussion, see Kelso et al., 1992). The particular change at $k = -k_c$ is a saddle-node bifurcation in which a *saddle* (the point repeller, or maximum, at $x = 0$)

Figure 2. Potential landscape defined by Equation 2 for five values of k (see text).

and a *node* (the point attractor at $x < 0$) are simultaneously created. Thus, where there was once only a single perceptual category there are now two possible categories. When $-k_c < 0 < k_c$ (Figure 2c), both “say” and “stay” are available categories. The co-existence of both categories continues until $k = k_c$, where the attractor corresponding to “say” ceases to exist via a reverse saddle-node bifurcation (where the qualitative change is from two available categories to one), leaving only the stable fixed point corresponding to “stay” (Figure 2d). Further increases in k only serve to deepen the potential minimum corresponding to “stay” (Figure 2e). Thus, the model captures the three observed states of the system: at the smallest values of the acoustic parameter, only “say” is reported; for an intermediate range of parameter values, either “say” or “stay” is reported; and for the largest values of gap duration, only “stay” is reported.

An accurate portrait of any real-world problem must take into account the influence of random disturbances, the sources of which, in a perception experiment, may correspond to factors such as fatigue, attention, boredom, and so on. Mathematically, spontaneous switches among attractive states occur as a result of these fluctuations, modeled as random noise. For a given point attractor, the degree of resistance to the influence of random noise is related to its *stability*, which, in general, depends on the depth and width of the attractor (i.e., its *basin of attraction*). As k is increased successively in Figure 2, the stability of the attractor corresponding to the initial percept decreases (the minimum becomes shallower and flatter), leading to an increase in the likelihood of switching to the alternative percept. This implies that perceptual switching is more likely with repeated presentations of a stimulus near the transition point than with repetition of a stimulus far from the transition point, a prediction confirmed in earlier work (Tuller et al., 1994).

In order to account for the three response patterns observed (critical boundary, hysteresis, and enhanced contrast), the behavior of k must have multiple determinants. One influential factor is the number of repetitions of each perceptual form before and after a boundary. In the Tuller et al. (1994) experiments, most subjects judged that the stimulus changed in the first half of the sequence. Therefore, a run in which gap duration first increased and then decreased tended to have many more

contiguous instances categorized as “stay” than categorized as “say.” In fact, repetitive presentation of a speech stimulus has long been known to shift the location of adjacent phoneme boundaries in a predictable direction (see Darwin, 1976; Eimas & Miller, 1978, for early reviews). Taking this factor explicitly into account, we proposed the following equation describing the behavior of k as a function of the gap duration:

$$k(\lambda) = k_0 + \lambda + \varepsilon/2 + \varepsilon\theta(n - n_c)(\lambda - \lambda_f), \quad (3)$$

where the value of k_0 specifies the percept at the beginning of a run, λ is linearly proportional to the gap duration, λ_f denotes the value of λ at the other extreme from its initial value, and n is the number of perceived stimulus repetitions in a run. The influence of the last term depends on a step function, $\theta(n - n_c)$. Before a critical number of accumulated repetitions (n_c) is reached, $\theta(n - n_c) = 0$. When $n \geq n_c$ (during the second half of each run), $\theta(n - n_c) = 1$. An additional parameter, ε , represents cognitive factors such as learning, linguistic experience, and attention. Note that the introduction of criteria stemming from cognitive processes is not without precedent; attention and previous experience, for example, play a large role in synergetic modeling of perception of ambiguous visual figures (Ditzinger & Haken, 1989, 1990; Haken, 1990) and contribute to factors that determine adaptation level in Helson’s work (Helson, 1964).

When n and/or ε are sufficiently small, the tilt of the potential is only dependent on gap duration and the initial configuration. Figure 3 illustrates how the observed patterns of responding are dependent on the rate of change of k . It shows three regions in the $\varepsilon - \lambda$ plane that correspond to the availability of perceptual categories. The first half of each run, when n is small, is displayed in Figure 3a; the second half of each run, when n is large, is shown in Figure 3b. White regions indicate the set of parameter values for which a stimulus has but a single possible categorization in the represented portion of the run. Shaded regions indicate the set of parameter values for which a stimulus may be categorized as either one form or the other (the bistable region) and thus represent the condition from $-k_c$ (the lower border of each shaded region) to k_c (the upper border of each shaded region). Looking at Figure 3a and reading from bottom to top, consider the condition with initial $k_0 = -1$ (“say”)

Figure 3. (a) Perceptual states in the ϵ - λ plane for increasing λ (small n). (b) Perceptual states in the ϵ - λ plane for decreasing λ (large n). (c) Superposition of transitions between categories in a and b. See text for details. (From "The Nonlinear Dynamics of Speech Categorization," by B. Tuller, P. Case, M. Ding, and J. A. S. Kelso, 1994, *Journal of Experimental Psychology: Human Perception & Performance*, 20, p. 5. Copyright 1994 by the American Psychological Association. Adapted with permission.)

and the gap duration parameter λ increasing. As λ increases, the stimuli are categorized as "say" for any value of ϵ so long as the $\epsilon - \lambda$ coordinate remains below the shaded region. Within the shaded region, the stimuli are categorized as "say" despite the percept's becoming progressively less stable. As λ continues to increase, the

percept switches from "say" to "stay" at the upper boundary of the shaded region (the heavy line), after which "say" is not a possible percept. Note that for different values of ϵ , the switch to a new percept occurs at different durations of silent gap.

In the second half of the run, λ is decreasing and the resulting division of the $\epsilon - \lambda$ plane looks somewhat different (Figure 3b). This portion of the figure should be read from top to bottom, from large gap durations to small ones. As λ decreases, the stimuli are categorized as "stay" for any value of ϵ as long as the $\epsilon - \lambda$ coordinate remains above the shaded region. Again assuming the absence of a perturbing force, the subject continues to categorize the stimuli as "stay" within the bistable (shaded regions) despite the percept's becoming less stable. As λ continues to decrease (gap duration continues to shorten), the lower boundary of the shaded region (the heavy curve) marks the switching from "stay" back to "say." In Figure 3c, we have simply superimposed the boundaries at which switching occurs in Figures 3a and 3b. The line with negative slope represents the category switch in the first half of the run, and the curve with positive slope represents the switch back to the initial category. Their intersection yields a value of ϵ , ϵ_c , for which a critical boundary would be observed in the "say-stay" continuum. For $\epsilon > \epsilon_c$ (e.g., the vertical line to the right of ϵ_c), the system exhibits the enhanced contrast effect. In that region, λ is smaller for the negatively sloped line than for the positively sloped curve. For $\epsilon < \epsilon_c$ (e.g., the vertical line to the left of ϵ_c), the classical hysteresis phenomenon is obtained: λ is larger for the line than the curve.

Model Predictions and Tests

Several predicted effects of the model were evaluated in Tuller et al. (1994) with a strong correspondence apparent between the theory and experimental observations. However, the correspondence is qualitative in that the values of coefficients in Equations 1 and 2 are chosen only for convenience of presentation. Other experimental information (e.g., measures of perceptual fluctuations, local relaxation times of perceptual states) is necessary to establish a precise, quantitative correspondence.

Here we evaluate predicted effects of the θ function on $k(\lambda)$, henceforth k . Specifically, when $\theta(n - n_c) = 1$ in Equation 3 (when the number of perceived repetitions, n , is larger than the critical value, n_c), each step change in gap duration (λ) entails a larger change in k than when $\theta(n - n_c) = 0$. This is illustrated in Figure 4, in which k is plotted against λ . Consider what happens for the "say"- "stay" continuum when gap duration sequentially increases (solid line) and then decreases (dashed line). When λ is at its minimum value (0 msec) at the beginning of the run, k_0 is arbitrarily assigned the value -1 so that k is also negative. As λ increases to its maximum (76 msec in our continuum), k continuously increases and the stimulus with maximum λ is categorized as "stay." Now λ begins to decrease, although the stimuli are still identified as "stay." When $n > n_c$, the θ function (the last term in Equation 3) acts to increase the rate of change

Figure 4. k - λ plot. The solid line represents increasing gap duration, and the dashed line represents decreasing gap duration. There is a dotted reference line at $k = 0$.

of k . Hence the steeper slope of the dashed line in the k - λ plane. The net result is that, as λ decreases back to 0 msec, the absolute value of k corresponding to a given value of λ is larger for the second portion of the run than for the first portion, for all response patterns.

Figure 5 shows solutions of $V(x)$ for expanded k when $n < n_c$ and gap duration is increasing (top row, left to right), and when $n > n_c$ and gap duration is decreasing (bottom row, right to left). Solutions in the top row are for the identical gap duration as the corresponding solutions in the bottom row. Nevertheless, the shape of the minima are different. A comparison of Figures 5a and 5n, or Figures 5b and 5m, reveals that the slope of the potential is steeper for a stimulus presented near the end of

the sequence than for the identical stimulus presented near the beginning. In contrast, the value of k associated with the largest value of λ (the turnaround point) is the same whether that stimulus is presented as the final stimulus of the first half (Figure 5g) of the run or as the first stimulus of the second half of the run (Figure 5h). To reiterate, although Figure 5 shows an instance of enhanced contrast, the relative difference in steepness of the potential at the beginning and end of a run is not dependent on response pattern.

In the present experiment, we exploit the findings that listeners are able to make fine distinctions among members of a given phonetic category and to give reliable judgments about the extent to which a given stimulus constitutes a good exemplar of the category (see Hodgson & Miller, 1992). We use subjects' judgments of goodness of a stimulus to index perceptual stability (the local slope of the attractor), and we compare the ratings of physically identical stimuli that occur in different positions in the sequence. If the model predictions hold, sequential changes in an acoustic parameter should result in a stimulus at the end of a run being judged as a better exemplar of the category than the same stimulus at the beginning of a run. Nonsequential, random changes in the acoustic parameter should not influence judged category goodness.

METHOD

Subjects

The subjects were 16 psychology students, including 10 undergraduates who received credit in their introductory psychology courses and 6 graduate students. All subjects were native speakers of American English and reported normal hearing in both ears.

Stimuli

Each stimulus consisted of a natural, 120-msec /s/ excised from a male utterance of "say," followed by a silent gap of 0–76 msec in 4-msec increments, followed by a 290-msec vocalic syllable created originally by Best et al. (1981) on the Haskins parallel-resonance synthesizer. Thus, there were 20 unique stimuli. The vo-

Figure 5. Potential landscape defined by Equation 2 for several values of k as determined by Equation 3 (see text). This sequence illustrates enhanced contrast with the perceptual shift from "say" to "stay" occurring between panels c and d, and the switch back between panels j and k.

calic syllables were three-formant stylized versions of the vocalic portions from natural, male utterances of "say" and "stay" and had a first formant ($F1$) onset frequency of 230 Hz that rose to 611 Hz over 40 msec, remained constant, and then fell linearly to 304 Hz during the last 90 msec. The second formant ($F2$) was constant at 1840 Hz for the first 150 msec, rose linearly to 2298 Hz over the next 90 msec, and remained constant for the last 50 msec. The third formant ($F3$) began with a linear 40-msec transition from 3196 to 2694 Hz, remained constant for 130 msec, then rose linearly to 3029 Hz over the next 70 msec, and remained constant for the final 50 msec. The $F0$ contour began at 120 Hz and remained constant for 240 msec, after which it fell to 90 Hz during the final 50 msec. All stimuli had a 25-msec rise time and 50-msec fall time. Parameter values for the amplitudes of $F2$ and $F3$ were $4/5$ that of $F1$. The stimuli are identical to those biased toward perception of "stay" when preceded by /s/ used in Tuller et al. (1994).

Procedure

Subjects were assigned to one of two groups, denoted *sequential presentation* and *mixed presentation*. The sequential-presentation group ($N = 8$) received two experimental conditions with runs of 40 stimuli that differed only in the order of stimulus presentation. In Condition 1, stimuli were presented with gap duration increasing from 0 to 76 msec in 4-msec steps. Then the *turnaround* stimulus (76 msec) was repeated and the gap durations for subsequent stimuli decreased in 4-msec steps until the *endpoint* stimulus (0 msec) was reached. In Condition 2, stimuli were presented with gap duration sequentially decreasing from 76 to 0 msec in 4-msec steps, the turnaround stimulus (0 msec) was repeated, and gap durations for subsequent stimuli increased in 4-msec steps until the endpoint stimulus (76 msec). Thus, these subjects received experimental runs that always varied the acoustic parameter of gap duration sequentially.

The mixed-presentation group ($N = 8$) also received two experimental conditions with runs of 40 stimuli, but the acoustic parameter varied sequentially only in the first half of each sequence. The second half of each sequence was altered in such a way that all but the turnaround and endpoint stimuli were randomized. Thus, Condition 1 included runs with gap duration varying in the following way: 0 msec, 4, 8, 12, 16, ... 72, 76, 76, (72 through 4 in random order), 0. Similarly, Condition 2 included runs with gap durations 76 msec, 72, 68, 64, ... 4, 0, 0, (72 through 4 in random order), 76.

Subjects heard 20 of the runs from each condition. In addition, both groups received 20 sets of the 20 stimuli in random order to ensure that the continuum elicited a typical identification function from both subject groups. Each of these randomized sets was divided in half, and these half-sets were interspersed between the experimental runs. All stimuli were blocked so that subjects heard the first 10 stimuli from a randomized set, a run randomly chosen from Conditions 1 or 2, the remaining 10 stimuli from the randomized set and another run randomly chosen from Condition 1 or 2. Thus, there were 100 stimuli in a block. The interstimulus interval (ISI) within a block was 3 sec; the interblock interval was controlled by the subject, but was at least 10 sec. This arrangement of runs embedded in a block was chosen so that subjects could not simply count from the beginning of a run to produce consistent category boundaries and to minimize the simple effect of the stimulus' serial position in a block. The interspersed random stimuli provided a buffer of relatively neutral context between each run. The 20 blocks were presented to subjects in three sessions of about 1 h each.

The subjects' task was to identify each stimulus as "say" or "stay" and then rate from 1 to 7 how good an exemplar of the category the stimulus was. They were given the following instructions for rating the stimuli. "Choose the number 1 only if you really could not tell if the stimulus was 'say' or 'stay.' Choose the number 2 if you thought you heard 'say' or 'stay,' but were not com-

pletely sure. Choose 3 if you were sure of what you heard, but the stimulus was a very poor example. Choose 4 if you were sure, and it was an "okay" example, 5 if you heard a good example of 'say' or 'stay,' 6 if you heard a very good example, and 7 if the word you heard was the best possible example of 'say' or 'stay' given the examples you have heard. Feel free to make comparisons between the stimulus words." Subjects entered their identification and rating responses on an appropriately labeled computer keypad. All stimuli were presented binaurally through headphones at a comfortable listening level (approximately 70 dB SPL). Subjects were tested individually in an IAC sound-attenuated booth.

Prior to the first experimental session on the 1st day, subjects heard 5 tokens each of the stimuli with 0-msec and 76-msec gaps to familiarize them with the "best" possible examples of the computer-synthesized words "say" or "stay," and a practice run of 30 stimuli (10 random, 10 with gap duration increasing from 4 to 76 msec in 8-msec steps, and 10 more random) to familiarize them with the experimental procedure.

RESULTS

Random Presentation

When stimuli were presented to subjects in random fashion, typical identification functions resulted (Figure 1). Across subjects, stimuli with gap durations from 0 to 24 msec were perceived as "say" nearly 100% of the time, and stimuli with gap durations from 52 to 76 msec were perceived as "stay" nearly 100% of the time. The 50% crossover point (often referred to as the category boundary) fell between stimuli with 36- and 40-msec gap durations for both groups. The pattern for individual subjects was similar, with the crossover point ranging from stimuli with gap durations of 28–48 msec.

Model Evaluation

Our working model implies that (1) sequential changes in an acoustic parameter should result in a stimulus at the end of a run being judged as a better exemplar of the category than the same stimulus at the beginning of a run (Figure 5a vs. Figure 5n); (2) the repeated stimuli at the turnaround point in a run should be judged as equivalent exemplars of the category (Figure 5g vs. Figure 5h); and (3) nonsequential, random changes in the acoustic parameter should not influence judged category goodness.

The experimental conditions were assessed by comparing the judged goodness of the two endpoint stimuli in each run (i.e., the two stimuli with 0-msec gap duration in Condition 1, and the two stimuli with 76-msec gap duration in Condition 2), for both sequential and mixed presentation. The judged goodness of the first occurrence of the stimulus was subtracted from the judged goodness of the second occurrence of the same stimulus in the same run. A mean difference was calculated across all 20 runs for each subject in each of the two experimental conditions. Thus, a positive difference would indicate that judged goodness of the second occurrence of the endpoint stimulus was better than the first occurrence. Next, the judged goodness of the two adjacent stimuli at the turnaround point in each run were compared by subtracting the rating for the first occurrence in the run from the second occurrence.

The mean differences in judged goodness of stimuli versus position in sequence are illustrated in Figure 6a. For the sequential-presentation group (the filled symbols), there are obvious differences in judged goodness between the first and second presentation of the endpoint stimuli for both Condition 1 (squares) and Condition 2 (circles) that are in the direction predicted by the model. The second presentation is judged as a better exemplar than the first, yielding positive mean differences. No such differences in judged goodness are observed for mixed presentation (the open symbols) or between the first and second presentation of a stimulus at the turnaround.

These results were confirmed by an analysis of variance (ANOVA) performed with group (sequential vs. mixed) as a between-subjects factor and stimulus position (endpoint vs. turnaround) and condition (1 vs. 2) as within-subjects factors. As expected, there was a significant main effect of group [$F(1,14) = 6.39, p < .05$]. Neither condition nor the interaction of group and condition had a significant effect on judged category goodness ($ps > .7$). The position of the stimulus (endpoint vs. turnaround) affected the judged category goodness [$F(1,14) = 7.31, p < .05$], but this interacted with group [$F(1,14) = 11.50, p < .01$]. No other interactions significantly affected goodness judgments. Thus, the model predictions were confirmed by the data.

We have not yet addressed whether pattern of responding interacts with judged category goodness of the endpoints. In this study, 38% of runs showed hysteresis, 38% showed enhanced contrast, and 24% showed a critical boundary. These percentages are close to those reported in Tuller et al. (1994), which were 41%, 42%, and 17%, respectively. It is important to note that for all three patterns of responding, the mean differences in judged goodness between the second and first presentations of the endpoint stimuli were positive (Wilcoxon matched-pairs signed-ranks tests were significant at the .05 level or better, with $N = 8$; T ranged from 0 to 4). Mean differences at the turnaround points were not significant for any response pattern and hovered around 0.

We turn the reader's attention again to the k - λ plot in Figure 4 and remark that differences in the depth of the potential as indicated by the absolute value of k are observable for less extreme stimuli, not only those at the endpoints. This results because with large n the increase in the potential's tilt, k , for a given change in gap duration λ is dependent on the influence of the step function, $\theta(n - n_c)$. In a 2×2 repeated measures ANOVA, the effects of stimulus position and condition on judgments of category goodness were examined for stimuli with 4-msec and 72-msec gap durations. Note that this analysis is possible only for the sequential-presentation group. Once again, there was a significant main effect of position [$F(1,7) = 10.22, p < .05$], no effect of condition [$F(1,7) = 0.27, p > .1$], and no significant interaction [$F(1,7) = 3.18, p < .1$]. A plot of the mean differences in judged goodness (Figure 6b) illustrates that the predictions of the model were confirmed with differences one

Figure 6. Mean differences in judged goodness versus position in sequence; (a) comparisons between stimuli with 0-msec gaps and between stimuli with 76-msec gaps, (b) 4 msec and 72 msec, (c) 8 msec and 68 msec, and (d) 12 msec and 64 msec. Filled symbols: Sequential presentation; open symbols: mixed presentation. Squares: Condition 1, circles: Condition 2.

step away from the endpoints' being positive and no differences one step away from the turnaround point.

When we step further into the sequence, comparing the first and second presentation of stimuli with gap durations of 8 and 68 msec, there is again a significant main effect of position [$F(1,7) = 10.20, p < .05$] and no significant main effect of condition [$F(1,7) = 0.25, p > .1$]. There is, however, a significant interaction between condition and position [$F(1,7) = 12.00, p < .05$]. When simple main effects are examined, the predicted difference in the endpoints versus the turnaround points holds for Condition 2 [$F(1,21) = 17.24, p < .001$], but not for Condition 1 [$F(1,21) = 0.57, p > .1$]. The mean differences in judged goodness are plotted in Figure 6c.

In order to understand the interaction, we must consider the asymmetry of the continuum and examine results on a subject by subject basis. That the continuum is asymmetric is illustrated by a significant difference in the mean judged goodness of the stimuli with the 0-msec and 76-msec gap durations in the random condition (Wilcoxon matched-pairs signed-ranks test, $T = 0, \alpha = .01$). The stimuli with the 76-msec gap duration were perceived to be better exemplars of "stay" than the stim-

uli with the 0-msec gap duration were perceived to be exemplars of "say." Thus, if one keeps in mind that judged goodness is being used as an index of stability, the Condition 2 runs began with a more stable percept than did the Condition 1 runs. Other things being equal (e.g., rate of change of the tilt parameter), in a sequence that begins with a more stable percept, it will take more step changes in the acoustic parameter to move into a region in which the position effects are obscured.

To some extent, we can determine the bistable region for subjects by looking at their random identification functions. Stimuli that an individual identifies sometimes as "say" and sometimes as "stay" can be said to be in that subject's bistable region (both percepts are available in response to those stimuli). In the random condition when only "say" or only "stay" responses were given to a stimulus, the stimulus was in a monostable region (only a single percept is possible). By this definition, the stimulus with the 8-msec gap duration can be considered to be bistable for 6 of the 8 subjects. The other two subjects showed only "say" responses for all stimuli with gap durations up to 28 and 40 msec, respectively. In contrast, all 8 subjects invariably identified the three stimuli with the longest gap durations (76, 72, and 68 msec) as "stay." That is, evidence of bistability is observed farther from the long-gap end of the continuum than from the short-gap end and is reflected by the interaction between stimulus position and condition. This is supported by the analysis of goodness judgments for the third step into the sequence, comparing stimuli with 12- and 64-msec gap durations. Here the near-endpoint stimuli were in the bistable region in both conditions, and there were no significant main effects of position [$F(1,7) = 0.10$, $p > .1$] or condition [$F(1,7) = 0.88$, $p > .1$], nor did they interact [$F(1,7) = 3.32$, $p > .1$]. The mean differences in judged goodness are plotted in Figure 6d. Thus, the model's predictions were upheld when the endpoint and near-

endpoint stimuli were perceived in an apparently monostable fashion.

One prediction of pattern-forming dynamic systems near instabilities is the presence of enhanced fluctuations as a transition is systematically approached (Haken, 1983a; Kelso, Scholz, & Schöner, 1986). Such fluctuations signal the progressive loss of stability of the potential minimum and are observable near the transition from "say" to "stay" and vice versa. For stimulus runs in which gap duration increased, the transition was defined as the stimulus after the last "say" identification. For stimulus runs in which gap duration decreased, the transition was defined as the stimulus after the last "stay" identification. Mean judged goodness and mean standard deviation of stimulus judgments aligned relative to the transition for each run are shown in Figure 7. Again, only sequential runs and sequential portions of mixed runs are included. The portion of each run in which gap duration increased is shown on the left and the portion in which gap duration decreased is shown on the right. The graphs should be read from negative to positive values for distance from the transition (negative distances denote stimuli that preceded the transition). Note that, as expected, the stimulus at the transition, or just after the transition, was judged to be the worst exemplar of its category and that the standard deviation of judged category goodness rises as judged goodness falls. However, the increased standard deviation is not uniform on either side of the transition. A sign test was used to compare each subject's standard deviation of judged category goodness for the stimulus just before the transition with the standard deviation of the stimulus just after the transition. The pretransition standard deviation was significantly greater than the post-transition standard deviation, regardless of the direction of gap duration change (with $N = 16$, $p < .002$, when gap duration sequentially increased; $p < .033$ for decreasing gap duration). This suggests the loss of stability as the

Figure 7. Mean judged goodness and the mean standard deviation for stimuli aligned to the transition. Gap duration increasing: left panel. Gap duration decreasing: right panel. Negative and positive distance values refer to the number of steps before and after the transition, respectively.

perceptual change is approached and supports the prediction of enhanced fluctuations in perception.

DISCUSSION

In the present experiment, we evaluated the predictions of a nonlinear dynamical model of speech perception. These concern the relative strength of percepts that arise in response to the same physical stimulus in different contexts. The results are in agreement with the theoretical model, but a complete evaluation requires a comparison with alternative explanations. Does the model make unique predictions? Can the observed data be accounted for by other theories? What does the nonlinear dynamical approach contribute to the understanding of basic issues in speech perception?

There are numerous examples of contextual effects on speech perception, including, but not limited to, the anchoring effects of stimulus repetition on category boundaries, selective adaptation effects, and interactions with overall stimulus properties (e.g., rate of speech). Several theories have been invoked to explain those effects, including Helson's (1964) adaptation level theory (see Diehl, Lang, & Parker, 1980, for application to speech), signal detection theory (see, e.g., Braidia et al., 1984; Durlach & Braidia, 1969; Macmillan, 1987), phoneme perception theory (Schouten & van Hossen, 1992; van Hossen & Schouten, 1992), and range-frequency theory (Parducci, 1965; Parducci & Wedell, 1986). Each of these theories on its own can account for some, but not all, of the contextual effects observed in the present study.

Adaptation-level theory makes some of the same predictions by virtue of the fact that it predicts contrastive effects under the specific sequential stimulus presentation conditions of our experiment. (Range-frequency theory makes qualitatively the same predictions as does adaptation level theory under these specific conditions.) In enhanced contrast, the transition between categories occurs, for example, at smaller gap durations in the first (increasing) half of the stimulus sequence than in the second (decreasing) half of the sequence. Thus, the first presentation of the 0-msec gap duration stimulus is closer to the category boundary (transition) in the first half of the stimulus sequence than the second presentation of the same stimulus is to the transition in the second half of the sequence. By inference from selective adaptation studies that show that stimuli farther from the boundary are judged to be better exemplars of a category (see, e.g., Miller, Connine, Schermer, & Kleunder, 1983; Samuel, 1982), the second presentation of the stimulus should be judged to be better than the first. In other words, adaptation level theory predicts our results, given that the subject's response pattern was enhanced contrast. When hysteresis is the observed response pattern, adaptation-level theory and range-frequency theories run into difficulty. Both theories fail to account for the observation of hysteresis in the present experiment, as well as in our earlier work (Tuller et al., 1994). But

given the observation of hysteresis, in which the presentation of an endpoint stimulus in the second half of the sequence occurs closer to the category boundary than does the presentation of the same physical stimulus in the first half of the sequence, then, once again on the basis of selective adaptation results, the first presentation of the endpoint stimulus is predicted to be a *better* exemplar than the second. Our present results show that the second presentation of the same stimulus is judged as a better exemplar of the category regardless of the pattern of responding. Note that, in our framework, enhanced contrast and hysteresis arise as two different consequences of the same dynamic model.

One puzzle concerns why previous research highlights the contrastive effects of stimulus repetition, whereas our own work demonstrates hysteresis and contrast equally often. In fact, our work departs from most studies of adaptation or anchoring in speech perception in at least two ways. One relevant difference is that we separate the effects of systematically leading the subject through the range of stimuli from the effects of stimulus repetition alone. Typical adaptors are a single stimulus repeated several times; anchors are a single stimulus spread through a block but occurring with higher frequency than does the rest of the stimulus set (Sawusch & Pisoni, 1973; see Simon & Studdert-Kennedy, 1978, for a comparison of adaptation and anchoring effects). Even when the adaptor included several syllables within the same phonetic category, they were randomly repeated prior to the test continuum (Cole & Cooper, 1977).

A second difference involves methods of analysis. Studies showing contrastive effects of adaptation and anchors have evaluated the effect of the adaptor or anchor on a statistically defined phoneme boundary based on the averaged responses across trials and subjects. Here we focus on the temporal evolution of the perceptual process in individual runs. It is notable that when the results of randomized presentation were examined for local sequential effects (Tuller et al., 1994) by computing the conditional probability of judging each stimulus as a member of the same category as that of the immediately preceding stimulus and data were averaged across subjects, no systematic contrast or assimilative effects were evident. These data differed from those previously reported (see, e.g., Diehl, Elman, & McCusker, 1978; Diehl, Kluender, & Parker, 1985; Diehl et al., 1980; Macmillan, Goldberg, & Braidia, 1988), most likely because in our analysis the biasing context was drawn from the entire range of stimuli, not a restricted set.

Rather than employ procedures designed to eliminate possible dynamic effects, we study these in their own right, because we believe them to reflect intrinsic properties of the speech perception system. The experimental paradigm and the theory that underlies it are dynamic in the sense that the ISI between successive presentations of stimuli is sufficiently small so that the contextual influence of the previous percept is carried from the past into the present. More specifically, in terms of the model, this means that the system state clings to its ini-

tial attractor, although random fluctuations may "push it over the hump." The paradigm is no longer dynamical if the ISI is chosen to be very large, thereby eliminating the context effect.

This time dependence is also evident in the signal detection model of Durlach and Braida (1969), which distinguishes three sources of variance in listening tasks: sensory, trace, and context variance. In trace mode, each stimulus is compared with the perceptual residue of the previous one. However, for identification tasks, such as in the present experiment, listeners are typically thought to operate in context mode, in which each signal is compared with an internalized standard that is based on the overall stimulus context. Variability associated with the context mode is proportional to stimulus range. The revised model of context coding (Braida et al., 1984) takes into account data showing that stimuli at the edge of the stimulus range are perceived better than would be expected according to Durlach and Braida's original signal detection model. They proposed that such edge effects were due to edge stimuli acting as perceptual anchors. According to this version of the model, when listeners are in context mode, stimuli are compared with well-remembered perceptual anchors. Stimuli near perceptual anchors are most likely to be accurately identified. (See Macmillan, 1987, and Macmillan et al., 1988, for applications to categorization of speech.) On the basis of this account, we would expect stimuli at the edge of the stimulus range (our 0-msec and 76-msec stimuli) to be accurately identified and would expect there to be no difference in how stimuli are perceived when they are presented at the beginning and end of a stimulus sequence. Nevertheless, systematic differences were observed. It might also be argued that subjects in the present experiment were instructed to compare stimulus words, so that the task was not strictly identification. Thus, listeners may also have been subject to trace variance. However, trace effects are inversely proportional to ISI (Macmillan, 1987), and the ISI used here was outside the range of ISI typically described as invoking trace variance.

The phoneme perception theory of van Hesson and Schouten (1992) adds to the signal detection model a long-term-memory or phoneme-labeling component like that found in the dual process theory of Fujisaki and Kawashima (1971). Van Hesson and Schouten propose that the context-coding and phoneme-labeling processes are of the same kind, but operate on different time scales. The basic assumption is that the phoneme-labeling process occurs very rapidly but fades slowly, while context coding (which is assumed to be a sublabeling process dependent on stimulus range) takes longer to build, but decays faster. The incorporation of the long-term-memory component, and the time dependence associated both with it and with the context-coding process, provides a better fit than does the trace-context theory of Macmillan et al. (1988) to experimental data collected in both identification and discrimination tasks and is consistent with our approach. In fact, when we used a methodology that destroyed the sequential nature of the acoustic pa-

rameter change (i.e., that minimized time dependencies), the contextual dependence of the judged category goodness disappeared.

A key issue in dynamic systems is the notion of loss of stability as a mechanism for behavioral (in this case, perceptual) change. The experimental question is whether the observed perceptual changes are in fact due to loss of stability. This point was recognized by Diehl et al. (1980), who reported that in order for adaptation-level theory to provide a good quantitative fit of adaptation and contrast effects in speech an additional parameter was necessary that would index the inherent stability of the category membership of a stimulus. In the present model, we have accomplished the same goal by using a somewhat different tack. We did not include a parameter that independently set stability; rather, the relative stability was inherent in the depth and width of the basin of attraction. Judgments of category goodness served as a behavioral consequence, or index, of perceptual stability. It is intriguing to consider that the strong "pull" of prototypical adaptors in selective adaptation studies also reflects the stability of prototypes relative to nonprototypical adaptors.

In physical systems (Haken, 1983b) and certain biological systems (e.g., Kelso, 1984; Kelso et al., 1986), stability can be measured by examining the variance or standard deviation of collective variables that characterize the behavior around attractive states; that is, the more stable the attractor, the less variability observed for a given stochastic force. In our previous work (Tuller et al., 1994), we provided indirect evidence for loss of stability as a means for perceptual change. There, we investigated the relative stability of percepts by presenting runs in which the acoustic control parameter, silent gap duration, changed sequentially and then remained fixed for multiple presentations. Stimulus repetition was used to probe the region of bistability: Spontaneous switching could occur only if more than one stable state was available, the principle being that as a category loses stability, the shallower minimum in the potential has a less restraining influence on fluctuations so that changes in category occur earlier with repetition than with single presentation of a stimulus. Indeed, we observed that an increase in the number of stimulus repetitions maximized the likelihood of earlier perceptual switching. In the experiment reported here, we used a very different index of stability, judged goodness, and observed an increase in variability as subjects approach consistent identification of the alternate category. Thus, the data provide converging evidence that the perceived change in category results from loss of stability of the perceptual state. It may be that earlier investigations showing an increase in identification reaction times near category boundaries should be reevaluated in terms of instabilities (Massaro, 1987; Pisoni & Tash, 1974; Repp, 1981).

Speech perception is highly context dependent. With its stability properties and time dependencies, there appears to be an exceptionally fluid, continual "remapping" of category structure. This remapping has been nicely demonstrated in several recent studies by Miller

and various colleagues, investigating contextual effects on the internal structure of speech categories. For example, Miller and Volaitis (1989) reported that a change in speaking rate not only shifted the category boundary in a /bi/-/pi/ continuum, but also shifted the range of stimuli judged to be good exemplars of a category. Hodgson and Miller (1992) investigated the internal structure of categories by using two "say"- "stay" continua with 230-Hz or 430-Hz *F1* onset frequencies. They observed not only the expected trading relations with different *F1* onsets (the boundary was shifted toward longer values of gap duration with the larger *F1* onset frequency; cf. Best et al., 1981), but also the region of best exemplars shifted toward longer gap durations. That is, a comprehensive alteration in mapping between acoustic property and phonetic category occurred both within and between categories. They concluded that theories of speech perception must account not only for the graded internal structure of categories, but also for the way in which the graded structure is determined by the multiple acoustic properties that result from any given articulatory act.

We take seriously Volaitis and Miller's (1992) prescription for the requirements of an adequate model of speech perception. First, one must account for the finding that phonetic categories do not comprise equally salient stimuli between category boundaries. We accomplish this in that for each value of λ in a given context there is a particular value of k ; that is, the depth and local slope of the potential is different for each stimulus presented, even when stimuli are categorized as the same. Second, a model of speech perception must allow contextual variables to alter the internal structure of the category, shifting both the range of stimuli perceived as belonging to the category and the range of stimuli perceived to be the category's best exemplars. In our model, the variable, λ , is linearly proportional to the manipulated acoustic parameter, gap duration. It is, of course, more realistic for λ to be a function (itself nonlinear?) of some complex of acoustic cues. As one aspect of λ is manipulated with respect to another, the slope of the potential function and the location of the stable minima are changed. Contextual variables and the relationship among acoustic cues affect the stability of perception. Third, an adequate model of speech perception must account for the way in which a contextual variable alters production and the way in which the variable restructures the phonetic category in perception. While our model does not specifically address this issue, the theories and methods of dynamic systems have been fruitful in understanding both perception and action systems. For example, Tuller and Kelso (1990) illustrated the usefulness of the basic concepts of self-organization, such as control parameters, bifurcations, attractors, multistability, instability, and hysteresis, in a study of speech production. Subjects were required to say /ip/ or /pi/ repetitively, at ever increasing speaking rates (the control parameter) while glottal and lip movements were monitored. In general, the observed interarticulator phase relationships varied little within limited regions of change

in the rate parameter. However, at critical values of speaking rate, the relative phase of articulator movements changed qualitatively. These transition points corresponded exactly to the perceptual shift in syllable affiliation of the consonant (see also Tuller & Kelso, 1991; for a discussion of perceptual consequences of dynamic perturbations during speech production, see Munhall, Löfqvist, & Kelso, 1994). Thus, the dynamic approach not only offers a real possibility to establish a link between speech perception and production, but provides complementary (if not alternative) strategies to understanding their relation (see Kelso, 1995, for elaboration).

In conclusion, the previously proposed model makes specific predictions, confirmed by the data presented here, regarding the relative stability of percepts that occur in response to the same physical stimulus in different contexts. The model meets, or has the potential to meet, multiple criteria for an adequate model of speech perception, including the ability to account for contextual variables that alter the internal structure of perceptual categories. More generally, the results of our investigations into contextual effects on categorization indicate that both contrastive and assimilative effects can arise as different consequences of the same dynamic process.

REFERENCES

- ABRAHAM, R. H., & SHAW, C. D. (1992). *Dynamics: The geometry of behavior*. Redwood City, CA: Addison-Wesley.
- BEST, C. T., MORRONGIELLO, B., & ROBSON, R. (1981). Perceptual equivalence of acoustic cues in speech and nonspeech perception. *Perception & Psychophysics*, **29**, 191-211.
- BRAIDA, L. D., DURLACH, N. I., LIM, J. S., BERLINER, J. E., RABINOWITZ, W. M., & PURKS, S. R. (1984). Intensity perception: XIII. Perceptual anchor model of context coding. *Journal of the Acoustical Society of America*, **76**, 722-731.
- COLE, R. A., & COOPER, W. E. (1977). Properties of friction analyzers of [j]. *Journal of the Acoustical Society of America*, **62**, 177-182.
- DARWIN, C. (1976). The perception of speech. In E. C. M. Friedman (Ed.), *Handbook of perception* (Vol. 1, pp. 175-226). New York: Academic Press.
- DIEHL, R. L., ELMAN, J. L., & MCCUSKER, S. B. (1978). Contrast effects on stop consonant identification. *Journal of Experimental Psychology: Human Perception & Performance*, **4**, 599-609.
- DIEHL, R. L., KLUENDER, K. R., & PARKER, E. M. (1985). Are selective adaptation and contrast effects really distinct? *Journal of Experimental Psychology: Human Perception & Performance*, **11**, 209-220.
- DIEHL, R. L., LANG, M., & PARKER, E. M. (1980). A further parallel between selective adaptation and contrast. *Journal of Experimental Psychology: Human Perception & Performance*, **6**, 24-44.
- DITZINGER, T., & HAKEN, H. (1989). Oscillations in the perception of ambiguous patterns: A model based on synergetics. *Biological Cybernetics*, **61**, 279-287.
- DITZINGER, T., & HAKEN, H. (1990). The impact of fluctuations on the recognition of ambiguous patterns. *Biological Cybernetics*, **63**, 453-456.
- DURLACH, N. I., & BRAIDA, L. D. (1969). Intensity perception: I. Preliminary theory of intensity resolution. *Journal of the Acoustical Society of America*, **46**, 372-383.
- EIMAS, P., & MILLER, J. (1978). Effects of selective adaptation on the perception of speech and visual patterns: Evidence for feature detectors. In R. Walk & H. Pick (Eds.), *Perception and experience* (pp. 307-345). New York: Plenum.
- FUJISAKI, H., & KAWASHIMA, T. (1971). A model of the mechanisms for speech perception—quantitative analysis of categorical effects in dis-

- crimination. *Annual Report of the Engineering Research Institute, Faculty of Engineering, University of Tokyo*, **30**, 59-68.
- HAKEN, H. (1983a). *Advanced synergetics: Instability hierarchies of self-organizing systems and devices*. Berlin: Springer-Verlag.
- HAKEN, H. (1983b). *Synergetics, an introduction: Non-equilibrium phase transitions and self-organization in physics, chemistry and biology*. Berlin: Springer-Verlag.
- HAKEN, H. (1990). Synergetics as a tool for the conceptualization and mathematization of cognition and behavior—How far can we go? In H. Haken & M. Stadler (Eds.), *Synergetics of cognition* (pp. 2-31). Berlin: Springer-Verlag.
- HAKEN, H., KELSO, J. A. S., & BUNZ, H. (1985). A theoretical model of phase transitions in human hand movements. *Biological Cybernetics*, **51**, 347-356.
- HELSON, H. (1964). *Adaptation level theory: An experimental and systematic approach to behavior*. New York: Harper & Row.
- HODGSON, P., & MILLER, J. L. (1992, October). *Internal phonetic category structure depends on multiple acoustic properties: Evidence for within-category trading relations*. Paper presented at the meeting of the Acoustical Society of America, New Orleans.
- KELSO, J. A. S. (1984). Phase transitions and critical behavior in human bimanual coordination. *American Journal of Physiology: Regulatory, Integrative & Comparative Physiology*, **15**, R1000-R1004.
- KELSO, J. A. S. (1995). *Dynamic patterns: The self-organization of brain and behavior*. Cambridge, MA: MIT Press.
- KELSO, J. A. S., CASE, P., HOLROYD, T., HORVATH, E., RĄCZASZEK, J., TULLER, B., & DING, M. (1994). Multistability and metastability in perceptual and brain dynamics. In M. Stadler & P. Kruse (Eds.), *Ambiguity in mind and nature* (pp. 159-184). Berlin: Springer-Verlag.
- KELSO, J. A. S., DING, M., & SCHÖNER, G. (1992). Dynamic pattern formation: A primer. In A. Baskin & J. Mittenthal (Eds.), *Principles of organization in organisms* (pp. 397-439). Redwood City, CA: Addison-Wesley.
- KELSO, J. A. S., SCHOLZ, J. P., & SCHÖNER, G. S. (1986). Non-equilibrium phase transitions in coordinated biological motion: Critical fluctuations. *Physics Letters A*, **118**, 279-284.
- LIBERMAN, A. M., HARRIS, K. S., HOFFMAN, H. S., & GRIFFITH, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, **54**, 358-368.
- MACMILLAN, N. A. (1987). Beyond the categorical/continuous distinction: A psychophysical approach to processing modes. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 53-85). Cambridge: Cambridge University Press.
- MACMILLAN, N. A., GOLDBERG, R. F., & BRAIDA, L. D. (1988). Resolution for speech sounds: Basic sensitivity and context memory on vowel and consonant continua. *Journal of the Acoustical Society of America*, **84**, 1262-1280.
- MASSARO, D. (1987). Categorical partition: A fuzzy-logical model of categorization behavior. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 254-283). Cambridge: Cambridge University Press.
- MILLER, J., CONNINE, C., SCHERMER, T., & KLUENDER, K. (1983). A possible auditory basis for internal structure of phonetic categories. *Journal of the Acoustical Society of America*, **73**, 2124-2133.
- MILLER, J. L., & VOLAITIS, L. E. (1989). Effect of speaking rate on the perceptual structure of a phonetic category. *Perception & Psychophysics*, **46**, 505-512.
- MUNHALL, K. G., LÖFQVIST, A., & KELSO, J. A. S. (1994). Lip-larynx coordination in speech: Effects of mechanical perturbations to the lower lip. *Journal of the Acoustical Society of America*, **95**, 3605-3616.
- PARDUCCI, A. (1965). Category judgment: A range-frequency model. *Psychological Review*, **72**, 407-418.
- PARDUCCI, A., & WEDELL, D. H. (1986). The category effect with rating scales: Number of categories, number of stimuli, and method of presentation. *Journal of Experimental Psychology: Human Perception & Performance*, **12**, 496-519.
- PISONI, D. B., & TASH, J. (1974). Reaction time to comparisons within and across phonetic categories. *Perception & Psychophysics*, **15**, 285-290.
- REPP, B. H. (1981). Perceptual equivalence of two kinds of ambiguous speech stimuli. *Bulletin of the Psychonomic Society*, **18**, 12-14.
- REPP, B. H., & LIBERMAN, A. (1987). Phonetic category boundaries are flexible. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 89-112). Cambridge: Cambridge University Press.
- SAMUEL, A. G. (1982). Phonetic prototypes. *Perception & Psychophysics*, **31**, 307-314.
- SAWUSCH, J. R., & PISONI, D. B. (1973). Category boundaries for speech and nonspeech sounds (Abstract). *Journal of the Acoustical Society of America*, **54**, 76.
- SCHOUTEN, M. E. H., & VAN HESSEN, A. J. (1992). Modeling phoneme perception: I. Categorical perception. *Journal of the Acoustical Society of America*, **92**, 1841-1855.
- SIMON, H., & STUDDERT-KENNEDY, M. (1978). Selective anchoring and adaptation of phonetic and nonphonetic continua. *Journal of the Acoustical Society of America*, **64**, 1338-1357.
- TULLER, B., CASE, P., DING, M., & KELSO, J. A. S. (1994). The nonlinear dynamics of speech categorization. *Journal of Experimental Psychology: Human Perception & Performance*, **20**, 3-16.
- TULLER, B., & KELSO, J. A. S. (1990). Phase transitions in speech production and their perceptual consequences. In M. Jeannerod (Ed.), *Attention and performance XIII* (pp. 429-452). Hillsdale, NJ: Erlbaum.
- TULLER, B., & KELSO, J. A. S. (1991). The production and perception of syllable structure. *Journal of Speech & Hearing Research*, **34**, 501-508.
- VAN GELDER, T., & PORT, R. (1995). *Mind as motion: Explorations in the dynamics of cognition*. Cambridge, MA: MIT Press.
- VAN HESSEN, A. J., & SCHOUTEN, M. E. H. (1992). Modeling phoneme perception: II. A model of stop consonant discrimination. *Journal of the Acoustical Society of America*, **92**, 1856-1868.
- VOLAITIS, L. E., & MILLER, J. L. (1992). Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories. *Journal of the Acoustical Society of America*, **92**, 723-725.

(Manuscript received July 21, 1994;
revision accepted for publication April 13, 1995.)