CHAPTER 8

Categorization and Learning in Speech Perception as Dynamical Processes

Betty Tuller

Center for Complex Systems and Brain Sciences
Florida Atlantic University
Boca Raton, FL 33431
U. S. A.
E-mail: tuller@ccs.fau.edu

The perception and production of speech unfolds in time. Although this seems like an obvious and perhaps trivial statement, this defining quality is not well captured by linguistic descriptions of speech, most of which are fundamentally static. Over the past few decades, both the theoretical and mathematical foundations for understanding organized behavior that emerges in time have been more fully developed and have infiltrated the study of many different human behaviors. In general, nonlinearity is a hallmark characteristic of these behaviors. That is, small changes in context or constituents can produce large behavioral effects and large changes in context or constituents might, in other conditions, produce little or no behavioral effect. The conditions that reveal the nonlinearities are often exactly those conditions that are excluded from experiments since they are more difficult to analyze and understand.

One nonlinear aspect of speech perception that has been the subject of a large number of studies is the phenomenon known as categorical perception. Within certain ranges of an acoustic parameter it is extremely difficult to discriminate between different stimuli that are labeled as the same speech segment. At the same time, stimuli with the same-size acoustic difference but in a different part of the parameter range are easily discriminated (e.g. Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman, Harris, Hoffman, & Griffith, 1957). As an example, consider the words "say" and "stay". When a short silent gap is introduced between the "s" noise and the vowel in "say" (e.g., from 0-20 ms), listeners continue to perceive the word "say." Similarly, listeners perceive the word "stay" when the silent gap after the "s" in the original stimulus is as long as 60 ms to 80 ms. But when the gap ranges from 30-50 ms, the same absolute difference in gap duration as in the previous two examples (i.e., 20 ms), listeners perceive an abrupt shift in the stimulus from "say" to "stay" (Best, Morongiello, & Robson, 1981). Importantly, when listeners identify two stimuli as belonging to the same phonetic category, they often have great difficulty discriminating between them. As the "category boundary" nears, stimuli are more easily discriminated from each other. Obviously, stimuli that are identified as different words or syllables are also easily discriminated. This means that there is acoustic variability but phonetic/perceptual stability in some ranges of the acoustic parameter (here gap duration) but perceptual change accompanies the same degree of acoustic change for other values of the acoustic parameter. In other words, the relationship between acoustics and phonetic perception does not change in a linear fashion.

The perceptual boundaries between categories, or "critical points," are not hard-wired by neurophysiology, or set indelibly by one's native language, but adjust flexibly with factors such as phonetic context, the acoustic information available, speaking rate, speaker, and linguistic experience (see Repp & Liberman, 1987, for review). This is not simply a laboratory demonstration: Listeners recognize the same word produced by different speakers (males, females, speakers of different ages) and by the same speaker in markedly different linguistic and intentional contexts, even when the listener has had no prior experience with the other individual's speech patterns. Thus, perceptual stability coexists with perceptual flexibility.

About a decade ago, Pam Case, Mingzhou Ding, Scott Kelso and I considered seriously the ideas that speech categorization is inherently nonlinear, and that the nonlinearity can serve as a window into the dynamics of speech perception (i.e., the equations of motion that characterize the intrinsic organization of speech perception). My charge by the organizers of the NSF workshop "Nonlinear Methods in Psychology" was to re-cap that work as an example of how these ideas can help guide empirical studies to allow a deeper understanding of the phenomenon under study. In exploring the process of speech perception in a non-traditional way, two things are extremely important. First, it is important not to ignore the decades of previous research on how people perceive speech. Any alternative theoretical views should be compatible with that body of work. Second, theory and experiment are related in a mutually informative fashion. The investigator's theoretical viewpoint guides not only *what* she or he chooses to examine, but also *how* it is examined. In turn, experimental results guide theory development, which then suggests the next empirical step. In the present case, since the focus is on evaluating whether the identification of speech sounds is itself characterized by what has been termed *perceptual dynamics* (characterized by multistability, loss of stability, flexibility, etc.; cf. Kelso, 1994a, 1995), the methodology chosen for the experiment must be one that allows for the possibility to see those signature characteristics. In what follows, I give a brief description of the main features of nonlinear dynamical systems relevant to this enterprise and describe how these features offered strategic guidance for the specific experiments and the theoretical model developed.

The first step was to conceive of perceptual space as a dynamical system with context, experience, and learning (among other things) as processes that can modify this dynamical system. Briefly, a dynamical system is one that evolves over time such that its present state always depends in some rule-governed way on previous states. Differential equations or maps (equations that dictate how a system evolves in discrete time steps) of relevant variables offer a mathematical description of the system's behavior as time passes and parameters change. Typically, one observes the stable behaviors of a system, referred to as its *attractors*. The attractor layout, or set of possible behaviors of a system, may change over time in such a way that observed behaviors change gradually or abruptly. Abrupt, or qualitative, changes (called *phase transitions* or *bifurcations*) may be thought of as the spontaneous emergence of new forms of organization (a self-organized pattern formation process) under specific boundary constraints (e.g., Haken, 1977; Nicolis & Prigogine, 1977). In a speech perception experiment, qualitative change in categorization of the stimuli allows a clear differentiation between patterns; there is no ambiguity as to what are the stable patterns for a given listener. Note that the qualitative change (here the shift in categorization) is informationally meaningful (Kelso, 1994). Although in any experimental situation there are many variables likely to be changing, the key is to discover the ones that bring about this qualitative categorical change. As Kelso (1995) has pointed out, situations where qualitative change occurs are also regions of dynamic instability and dynamic instability is the generic mechanism underlying self-organized pattern formation (Haken, 1977; Nicolis & Prigogine, 1977). Without the dynamic instability, no change in pattern would occur. In turn, if one can see evidence of growing dynamic instability, then one can study the emergence of the new pattern. We will return to the idea of dynamic instability when we describe the experiments evaluating speech categorization as a dynamical phenomenon.

Although this description of qualitative pattern change as some parameter varies bears a strong similarity to the results of speech categorization tasks, the similarity may be only superficial. Empirical work on speech categorization, in order to maintain the independence of treatment levels required by most parametric statistical techniques, typically presents the stimuli to listeners in random order. Such experiments thus describe the location of a statistically defined phoneme boundary (most often, the point corresponding to the 50% crossover of the response function for a two-category set; see Ganong & Zatorre, 1980, for a comparison of different methods for defining boundary location). Unfortunately, this traditional methodology is far from optimal for revealing the dynamical characteristics being evaluated, because the randomization of stimuli destroys the footprints of any underlying dynamical process that may govern the transition between speech sounds. So one's theoretical viewpoint *must* influence experimentation from the initial design stage.

The strategy in our experiments was to use a stimulus continuum for which categorical perception has often been demonstrated but to vary the acoustic parameter sequentially, i.e., as a *control parameter*. A control parameter is one that, when the appropriate range of values is used, takes the subject from one perceived categorization to another. For some behaviors, finding control parameters is non-trivial. However, the literature on categorical perception gives us many plausible control parameters for different speech categorizations. In what follows, I will review the observed dynamical effects and delineate some of the factors responsible. A model of the results was proposed and is discussed, and unique predictions of the model tested. Lastly, I will describe how viewing the speech perception process as a nonlinear dynamical system forces, as a natural extension, a re-examination of the process that occurs when learning to hear non-native phonemic distinctions. Our experiments demonstrate the fruitfulness of the approach and reveal that speech perception and perceptual learning in speech are characterized by rich underlying dynamics.

In 1994, Tuller, Case, Ding, and Kelso examined speech categorization when an acoustic parameter—the length of the silent gap between a natural "s" and a synthetic "ay"—was varied in a stepwise fashion. We used this particular stimulus continuum because it had already been shown that listeners perceive "say" at short silent gaps but they perceive "stay" at long silent gaps (e.g., Best et al., 1981). Thus, the gap duration after the "s" was a possible control parameter by which we could explore the mechanism of switching between categorizations. However, a major difference between our experiment and those of others was that we presented the stimuli *in order*. That is, gap duration either increased systematically from 0-76 ms, then back to 0 ms, in 4-ms steps, or decreased from 76 ms to no gap, then back to 76 ms in 4 ms steps. There were 5 trials of each of these two sequences. We also randomized the stimuli and presented 10 randomizations to the listeners. The subject's task was to indicate whether they perceived the word "say" or the word "stay" by pressing appropriately labeled keys on a computer keyboard. First, we determined that the randomized stimuli resulted in the same perceptual identification function as reported previously in the literature. This ensured that our stimuli (and listeners) were equivalent to those used by others. Because the point at which categorization shifts as a function of the direction of changes in gap duration is considered a theoretically important juncture, the next analysis focused on that point.

Logically, there are only three possible patterns of switching: (1) A subject will switch between "say" and "stay" at the same gap duration regardless of direction of gap change (a critical boundary); (2) A subject's percept will change at a larger gap duration as gap increases than when gap decreases (an effect know as *hysteresis* or *assimilation*); or (3) A subject's percept will change at a larger gap duration when gap decreases than when gap increases (a contrastive effect). All three patterns were observed, with critical boundary being much less frequent than hysteresis or contrast, which occurred equally often. Thus, the perceptual changes in this speech identification task show quite complicated dynamics when a relevant acoustic parameter is sequentially varied. Closer analysis revealed that the incidence of hysteresis and contrast was not simply random fluctuation around a critical boundary, because their relative frequency changed in predicted ways over the course of the experiment. These patterns of change reveal that dynamic instability is playing a role in perceptual switching, thereby linking phonemic categorization to self-organized pattern formation.

How do you begin to connect experimental data to a generic dynamical model? Quite simplistically, since we have two reproducibly observed states—here the two categorizations "say" and "stay"—we identify the categorizations with attractors, or stable states in perception. We use differential equations to define systems with attractor properties that fit the observed experimental data. Differential equations allow us to model quantities that change continuously in time. We can find stable solutions of the differential equations by finding equilibrium points, values of x for which the derivative $dx/dt=0$ (see Equation 8.1; by definition, if the derivative of some variable is zero, that means the variable is unchanging, which is what it means for that value to correspond to a stable state). Trajectories (solutions to the differential equation) may be "attracted" to an equilibrium point or "repelled." We call the first case a *stable attractor* (also called a *sink*) and we call the second case an *unstable attractor* (also called a *source* or *repeller*).

If listeners perceived only a single perceptual category, a theoretical model of a single attractor, a fixed-point, would be adequate. A situation in which two states, or categories, occur requires that the model contain at least two stable attractors that change with the control parameter. In our case, the model must be able to account for the fact that at some gap durations a listener perceives only "say" and for other gap durations the listener perceives only "stay." The presence of hysteresis and contrast is also informative, indicating that more than one stable percept can coexist for a given acoustic stimulus—either "say" or "stay" might be perceived. In this case the stimulus is *bistable*—the two attractors must coexist for some range of the control parameter.

These results were modeled concisely by the following dynamical system (Tuller et al., 1994), written as a differential equation:

$$dx/dt = -dV(x)/dx = -k + x - x^3 \qquad [8.1]$$

Differential equations may be rewritten in the form of a potential function (Equation 8.2), in which the attractors are geometrically obvious when the potential is plotted. Here *x* is a variable characterizing the perceptual form and *k* is a parameter specifying the direction and degree of tilt for the potential. This allows visualization of the behavior of the system as the parameter *k* is manipulated.

$$V(x) = kx - x^2/2 + x^4/4 \qquad [8.2]$$

Think of Equation 8.2 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 8.1). The minima of the potential, the valleys in the landscape, are the attractors corresponding to the two perceptual categories.

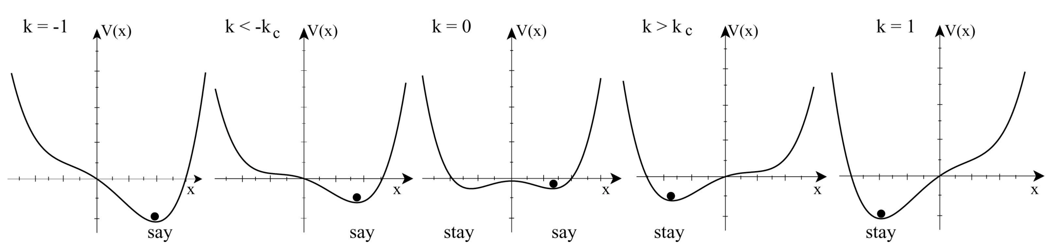

Figure 8.1. Potential landscape defined by Equation 8.2 for five values of *k* (adapted from Case et al., 1995).

Figure 8.1 shows how the landscape changes for several values of *k*. With *k* = *minimum* only one stable point exists corresponding to a single category (e.g., "say"). As k increases, the potential landscape tilts but otherwise remains unchanged. 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. The particular change at $k = -k_c$ is a saddle-node bifurcation in which a "saddle" (the point repeller, or maximum, at $x = 0$) 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. This bistability, 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." Further increases in k only serve to deepen the potential minimum corresponding to "stay." 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" are 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. In the present work, we considered factors such as fatigue, attention, and boredom to correspond to random disturbances because we could not measure the changes in those factors over time. Further experimental work may elaborate whether these factors are indeed random or predictable modifiers of perceptual space. 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 8.1, 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 away from the transition point, a prediction confirmed in Tuller et al. (1994).

In order to account for the three response patterns observed (critical boundary, hysteresis, and contrast), the behavior of k must have multiple determinants. One influential factor suggested by earlier research is the number of repetitions perceived from each category. 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, and 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:

$$k(\lambda)= k_0 + \lambda + \varepsilon /2 + \varepsilon\theta\, (n-n_c)\, (\lambda-\lambda_f), \qquad [8.3]$$

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 final value of λ (i.e., 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, $\theta(n-n_c) = 0$. That is, in the first half of each run, the tilt of the potential is only dependent on gap duration and the initial configuration. When $n \geq n_c$ (during the second half of each run) $\theta(n-n_c) = 1$. This means that each step change in gap duration λ will produce a larger change in tilt k than it did in the first half of the run. An additional parameter, ε, represents cognitive factors such as learning, linguistic experience, and attention. Note that the importance of cognitive processes is well-established, for example, attention and previous experience play a large role in synergetic modeling of perception of ambiguous visual figures (Haken, 1990; Ditzinger & Haken, 1989, 1990) and contribute to factors that determine adaptation level in Helson's work (Helson, 1964).

Although the additional term was needed to incorporate contrast effects into the same model that described hysteresis and a critical boundary, it gave rise to unexpected predictions. For example, if the subject is presented with a run with gap duration first systematically increasing (from 0-76 ms) then systematically decreasing (from 76 ms back to 0 ms), the percept is predicted to be more stable—the potential would have a locally steeper slope—when the same stimulus appeared as the last item in the run than as the first item in the run. This is because the rate of change of tilt of the potential is faster in the second half of the run for the same amount of acoustic change. This prediction is unexpected given the literature on selective adaptation effects in speech. In selective adaptation, a standard identification task is first used to locate the "category boundary," or point of subjective equality, for the test continuum. Next, the subjects listen to the stimulus from one end of the continuum presented many times over. After a second identification test with the original stimulus continuum, the position of the perceived category boundary moves towards the repeated stimulus. For example, in a [ba]-[pa] continuum varying in the lag of voicing onset after the initial consonant release burst, if the stimulus with the longest voicing lag is repeatedly presented after the first identification test, listeners then require a longer voicing lag for a stimulus to be perceived as a [pa] (Eimas & Corbit, 1973)—in our terms, perception of [pa] has destabilized. Somewhat counterintuitively, our model predicts that when a word is perceived many times over, its stability will increase.

This prediction was confirmed by experiment (Case, Tuller, Ding, & Kelso, 1995). In that work, we used the same "say"-"stay" stimulus continuum but asked listeners not only to categorize the stimulus as either "say" or "stay" but also to rate how good an exemplar of the category the stimulus was. The goodness rating was used as an index of the stability of the percept (the local steepness of the potential function). As predicted, regardless of whether the stimuli were presented with gap duration between the "s" and the "ay" first increasing from 0-76 ms and then decreasing back to 0 ms, or in the opposite direction, 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 (Figure 8.2). One crucial difference between the work of Case et al. (1995) and the earlier work on selective adaptation concerns the repeated stimulus. In the former, the stimuli were changing systematically, albeit at a subcategory level; in the latter, the identical stimulus (typically an

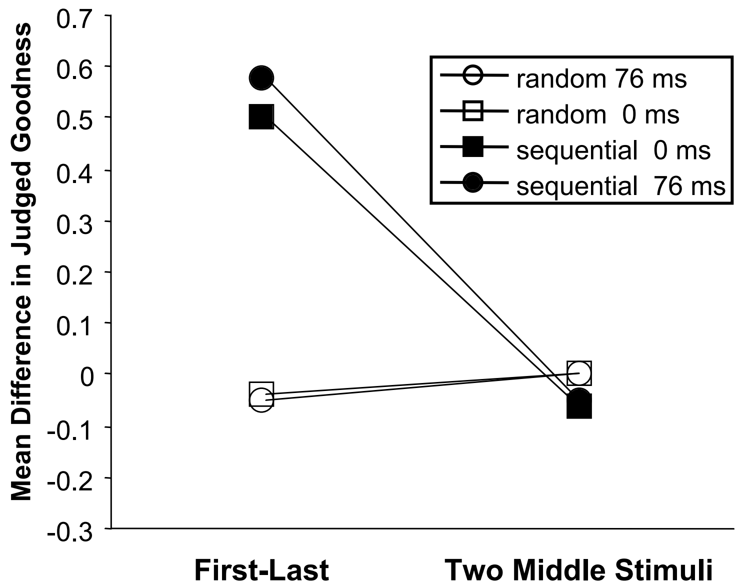

Figure 8.2. Mean differences in judged goodness versus position in sequence as a function of sequential vs. random stimulus order. When stimuli are presented sequentially (solid symbols), the last stimulus presented is judged as a better exemplar than the same stimulus when presented first in the sequence. This occurs for both 0 ms (square) and 76 ms (circle) gap stimuli and does not occur with random stimulus orders intervening (open symbols) or when the same stimuli are the "turnaround" stimuli in the middle of the trial (adapted from Case et al., 1995).

end-point stimulus) was repeated. In fact, when Case and colleagues presented stimuli with an intervening set of stimuli with randomly changing gap durations, no differences in judged goodness were observed. This result confirmed one prediction of speech categorization as a context-sensitive, pattern-forming system.

Another difference between this empirical confirmation of the model's predictions and the literature on selective adaptation motivated additional research. The model implies that the temporal evolution of the alternative forms, and hence switching between them, depends on how the stimuli move through perceptual space. This was supported by Case et al. (1995), described above, at least for the judged goodness of the stimuli as members of the identified category. Thus, systematic change in an acoustic control parameter, and not solely the number of stimulus repetitions, is crucial. This was tested directly by presenting subjects with a single “say”-“stay” trial with gap duration either increasing or decreasing (again, in 4-ms steps between 0-76 ms silent gap). The second trial was adjusted for individual subject responses to the first trial. If, for example, a subject heard a switch from "say" to "stay" on the 6th stimulus in the first trial, then in the second trial stimulus #1 was presented 5 times, then stimulus #6 was presented, then the trial continued to the end, with each successive stimulus presented once. Selective adaptation leads one to expect that repeating the initial stimulus in trial 2 should cause listeners to switch earlier, or at the same stimulus, as in trial 1 (*contrast* or *critical boundary* should increase in observed frequency). Similarly, if the preponderance of hysteresis observed previously reflects only a response perseveration, then the incidence of *critical boundary* should increase markedly because both trials present the same number of instances of the initial category. Identical predictions are made by Helson’s (1964) Adaptation Level Theory, which holds that all stimulus inputs in a given domain are pooled and their running average determines the level of stimulation to which the person is adapted. Alternatively, if the underlying nonlinear dynamic model has validity, then subcategorical sequential acoustic change, not simply perceived repetition, enhances hysteresis. Results confirmed overwhelmingly that only sequential acoustic change increases the frequency of hysteresis

(Figure 8.3), a result that was later shown to generalize to the perception of directional pitch (Giangrande, Tuller, & Kelso, 2003).

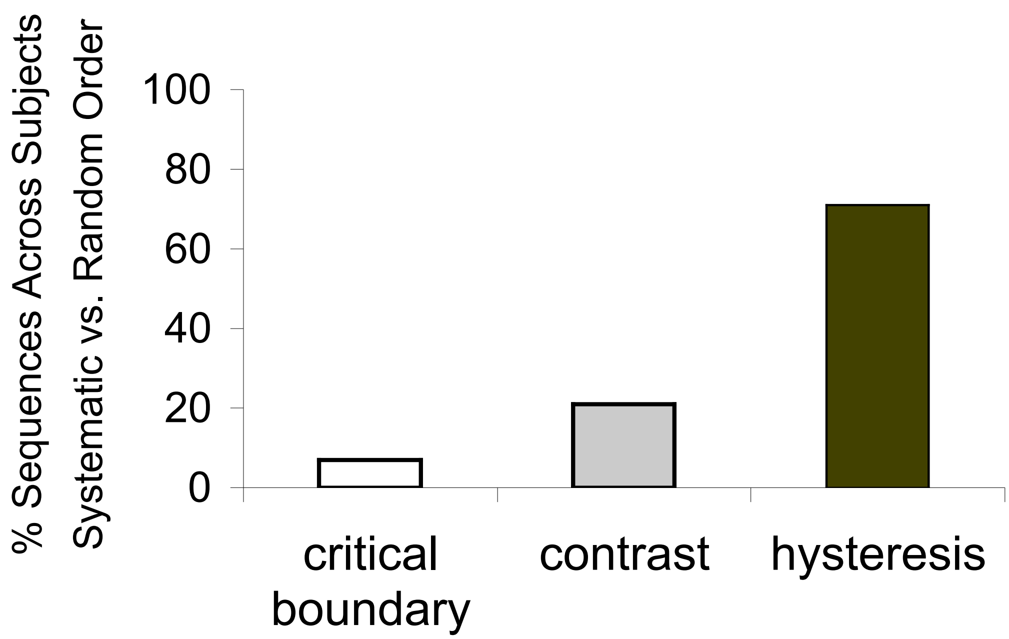

Figure 8.3. Comparison of switching behavior in sequences that contain systematic acoustic change with matched sequences that instead repeat the end-point category (see text). Percent of sequence pairs perceived as switching at the same stimulus (critical boundary; white bar), systematic stimulus change switching earlier than random change (contrast; gray bar), and systematic stimulus change switching later than random change (hysteresis; black bar).

Yet another test of the model's predictions began to address the role of learning and experience. Recall that enhanced experience (of which stimulus repetition is one example) causes the potential to change more rapidly. Minimizing learning and experience should lead to a majority of hysteresis response patterns; contrast should occur much less often. To evaluate this prediction, we presented subjects with a single run of the “say”-“stay” continuum with gap duration first increasing from 0-76 ms then decreasing back to 0 ms. Another group of subjects was presented with a single run of stimuli that began at 76 ms gap duration, decreased in 4-ms steps to no gap, then increased back to 76 ms gap duration. The task was to identify each stimulus as "say" or "stay." A subject's pattern of responding (hysteresis, critical boundary, or contrast) was determined by comparing the gap duration at which the perceptual switch occurred in the increasing vs. decreasing portion of a run. Results confirm that when experience with the stimuli is minimized, the proportion of hysteretic responses is far greater than either contrast or critical boundary. In fact, hysteresis is over 3 times more prevalent than any other response pattern and is independent of the direction of change in gap duration. When the first trial for each subject from Tuller et al. (1994) and Case et al. (1995) is examined, results are statistically identical to those obtained when subjects were presented with only a single trial (Figure 8.4).

Obviously, these experiments consider only a very restricted definition of “phonological learning” in adults. Typically, when adults attempt to learn new speech sounds, they do so in the context of the phonology of their native language. From the perspective we have been taking, it makes sense to think of perceptual space as a dynamical system that is modified by learning. In other words, learning a new phonological category (when a range of acoustic objects acquires a common meaning) is viewed as the creation of an attractor that modifies the existing dynamics. This allows us to predict how learning will proceed, depending on how the stimuli are initially perceived by the individual. In non-speech perceptuomotor tasks, evidence that

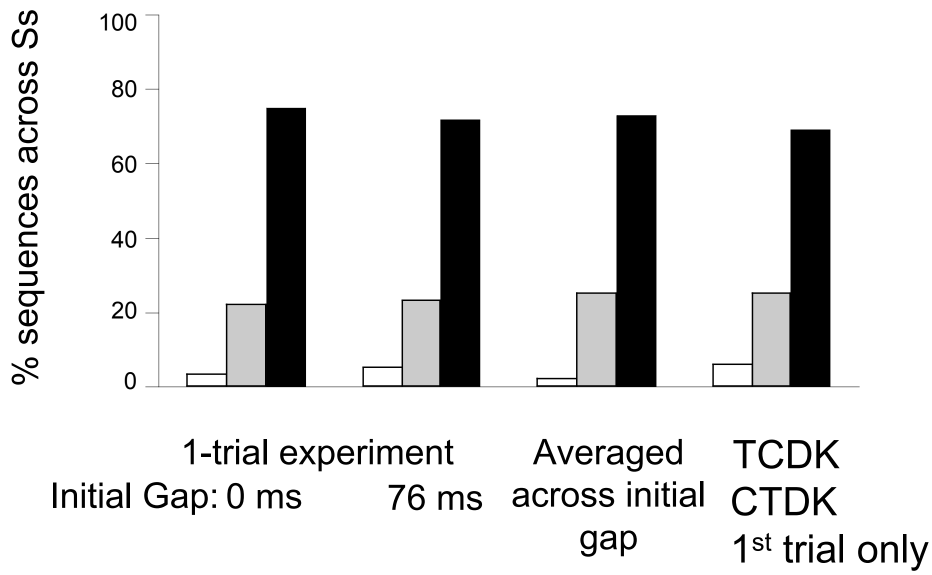

Figure 8.4. Percent of sequences perceived as having a critical boundary (white bar), contrast (gray bar), and hysteresis (black bar) when for only one trial per subject, or the first trial per subject. TCDK: Tuller et al. (1994). CTDK: Case et al. (1995).

learning consists of the interaction between pre-existing constraints that the subject brings into the learning situation and the behavior to be learned has been provided by Schöner and Kelso (1988; see also Schöner, Zanone, & Kelso, 1992). In their model, behavioral information (such as the task to be learned) acts as a parameter of the attractor dynamics, attracting behavior toward the required behavior. When the former does not correspond to a stable attractor of the existing, intrinsic dynamics, learning is predicted to take the form of a phase transition: A new behavioral attractor is found that alters the entire dynamics. When the required task is close to, or coincides with, an existing stable pattern, cooperative mechanisms ensure that learning will proceed rapidly and smoothly (Zanone & Kelso, 1992; 1994; 1997; see also Kelso, 1990).

How might these ideas impact upon the acquisition of new phonological categories that a person has never used? If a listener initially can perceive a non-native sound as "different" from a native one, although perhaps still acceptable as an exemplar of the native category, the existing perceptual landscape cooperates with the sound to be learned. Operationally, the rate of change of the landscape to include the sound to be learned, the progressive stabilization of the new sound, should be relatively smooth and fast. In contrast, if a listener initially perceives the non-native sound as indistinguishable from a native one, then learning to recognize the non-native sound competes with the existing perceptual organization. In this case, the strength of the attraction of the to-be-learned sound increases until a qualitative change (a bifurcation, or phase transition) reflects the emergence of a new attractor. The rate of change of the perceptual space to the new sound should be slower than when the initial perceptual landscape cooperates with the new sound. In addition, because this competition entails destabilization of the existing attractor, the bifurcation should be marked by high variability.

In order to test these ideas, it is necessary to modify the standard experimental techniques used in phonological learning tasks in two ways. First, it is not sufficiently informative simply to note whether learning occurs with a particular stimulus set and training régime. Observations of the changes in each listener's behavior as learning proceeds must supplement measures of whether the trained distinction was finally learned to some criterion. Second, the focus of analysis must be the individual, not the language. As an example, consider Iverson and Kuhl's (1996) investigation of native English speakers' perception of English /r/ and /l/ in which multidimensional scaling analyses of individual listener's similarity ratings of stimulus pairs revealed that the warping of perceptual space corresponded best to the listener's own identification patterns. Similarly, Aaltonen, Eerola, Hellström, Uusipaikka, and Lang (1997) showed individual differences in mismatch negativity EEG patterns depending on how the subject categorized the stimulus sequence. In other words, perceptual learning as a result of language training must be assessed relative to the individual's perceptual space as it exists before training begins. To do this, appropriate probes, or maps, of the latter should be conducted prior to, and during, the learning process.

In a doctoral thesis that embodied these attributes, Case (1996) used the voiced Hindi dental stop consonant /d̲/, which is acoustically similar to the American English alveolar stop consonant /d/, as the category to be learned. The major articulatory distinction between these two sounds is in place of articulation—in /d̲/ the tongue tip is placed against the upper front teeth, and in /d/, the tongue tip is against the alveolar ridge. There is no phonemic contrast between the dental and alveolar place of articulation in either Hindi or American English, although it is contrastive in at least a half dozen languages (including Malayalam and several Australian and African languages; Jongman, Blumstein, & Lahiri, 1985).

Here I will concentrate on the following questions: What are the dynamics of the learning process itself? Does the form that learning takes depend on the relationship between the sounds to be learned and how the individual initially perceives them? What are the effects of learning a new speech sound on an acoustically/articulatorily close native speech sound? That is, does an individual's phonetic system reorganize during learning by modifying native categories (e.g., Flege, 1995)?

To answer these questions, we used a "perceptual mapping" procedure that included three different tasks (identification, judged goodness, and difference ratings). These tasks together allow a more complete assessment of each listener's perceptual space than use of any of the tasks alone. Each of the tasks taps somewhat different aspects of speech perception. Identification tasks encourage phonetic coding, and a variable stimulus context that includes different speakers, utterances, and phonetic contexts facilitates robust category formation with training (Lively, Logan, & Pisoni, 1993; Pisoni & Lively, 1995). The judged goodness task examines the internal structure of a category in a way that an identification task obscures, allowing the listener to determine how good an exemplar of a category a given stimulus is and focusing attention on differences among stimuli. Data from the difference-rating task allow one to investigate the internal structure of one or more categories simultaneously. Incorporating the results of all three tasks gives a fuller picture of how a given listener perceives the stimuli.

A group of monolingual American English listeners first completed the three-task perceptual mapping procedure and then participated in a 15-session training program distributed over a three-week period. Their progress was monitored throughout training. Following training, the perceptual mapping procedure was repeated.

Pre-training/post-training comparisons as well as daily assessments during the training process were performed to assess whether learning occurred and, if so, to reveal its dynamics. Persistence of learning was evaluated by follow-up testing administered a few weeks after the training was completed. This methodology stems from the scanning probes of the dynamics employed during the learning process by Zanone and Kelso (1992, 1997) in order to understand how, in their case, pre-existing coordination tendencies were modified by practicing a new skill.

The training stimuli, a list of /CV/ syllables and /αCV/ disyllables, were produced by four native speakers of Hindi (H) and two native speakers of American English (AE). The consonant was either /d̲/ or /d/ and the vowels were those in "hot," "heat," "hoot," and "hut." Hindi speakers were instructed in the production of the alveolar stop, and AE speakers were instructed in the production of the dental stop. Three native speakers of AE rated all intended alveolar productions and three native speakers of H rated all intended dental productions. Only productions judged to be acceptable by all native listeners were used in training. The final training set was acoustically diverse in that it included 3 tokens each of the 16 different syllables (8 dental, 8 alveolar) from four H speakers and two AE speakers.

The test stimuli were a synthetic continuum of eleven syllables with an initial stop consonant followed by the vowel /α/. The consonant spanned a range from the Hindi dental /d̲/ to the American English alveolar /d/ by manipulating the second (F2) and third (F3) formant onset frequencies. Hindi listeners judged stimuli from the dental end of the continuum to be better exemplars of their native category than stimuli from the alveolar end of the continuum.

Monolingual speakers of American English (AE) participated in two pre-training sessions of about one hour each. In the first session, they performed the judged goodness and identification tasks. In the second session, they performed the difference-rating task. For the judged goodness procedure, subjects were presented with a randomized set of ten tokens each of the eleven unique synthetic stimuli. The task was to rate from 1 to 7 (poorest to best) how good an exemplar of /d/ the stimulus was.

For the identification task, subjects were presented with a differently randomized set of ten tokens each of the eleven stimuli. Subjects were told that stimuli would be either a synthesized version of an American English alveolar /d/ or a Hindi dental /d̲/. Differences in how the two sounds are produced were described and examples of the endpoint stimuli from the continuum representing the two sounds were presented. The two-alternative forced-choice task was to identify the stimulus as either alveolar or dental.

In the difference rating task subjects heard all possible pairs of stimuli from a 6-stimulus subset of the continuum (stimuli 1, 3, 5, 7, 9, and 11). Pairs were rated on a scale from 1 to 7, with 1 being "exactly the same" and 7 being "most different."

After the initial perceptual mapping subjects participated in 15 training sessions within a 3-week period, a second perceptual mapping just after training, and another mapping at least two weeks later. Each daily training session consisted of (in order) an initial free exploration period, a two-alternative forced-choice identification task (with feedback) for a training set of 48 natural speech stimuli randomly chosen from the full set of 288 natural speech stimuli, the difference rating test, an identification task with feedback for a different 48-item subset of the natural speech stimuli, and a second difference rating test with a new randomization of stimulus pairs. If subjects had not been paid for participating I doubt anyone would have completed the experiment!

Although every subject showed some improvement in differentiating dental from alveolar stop consonants in natural speech, in what follows, I will discuss two subjects' learning patterns in order to address the questions posed above.

In the pre-training identification task with voiced stimuli, our first learner showed some ability to identify the four extreme dental-end stimuli as dental (Figure 8.5). Nevertheless, he still rated all stimuli as relatively good members of the alveolar category (Figure 8.6). These results are intriguing in that stimuli consistently identified as dental were still judged as relatively good alveolars. This underscores not only the poverty of using only a single measure of an individual's phonetic perception but also the flexibility of perception.

In both the post-training and follow-up identification tasks, the identification functions partition the stimuli into two clear categories with more stimuli now being identified as dental (Figure 8.5). In contrast to the pre-training mapping, however, stimuli on the dental end of the continuum are now judged to be poor exemplars of the alveolar category and the stimulus judged as the "best" alveolar moves toward the alveolar end of the continuum (Figure 8.6).

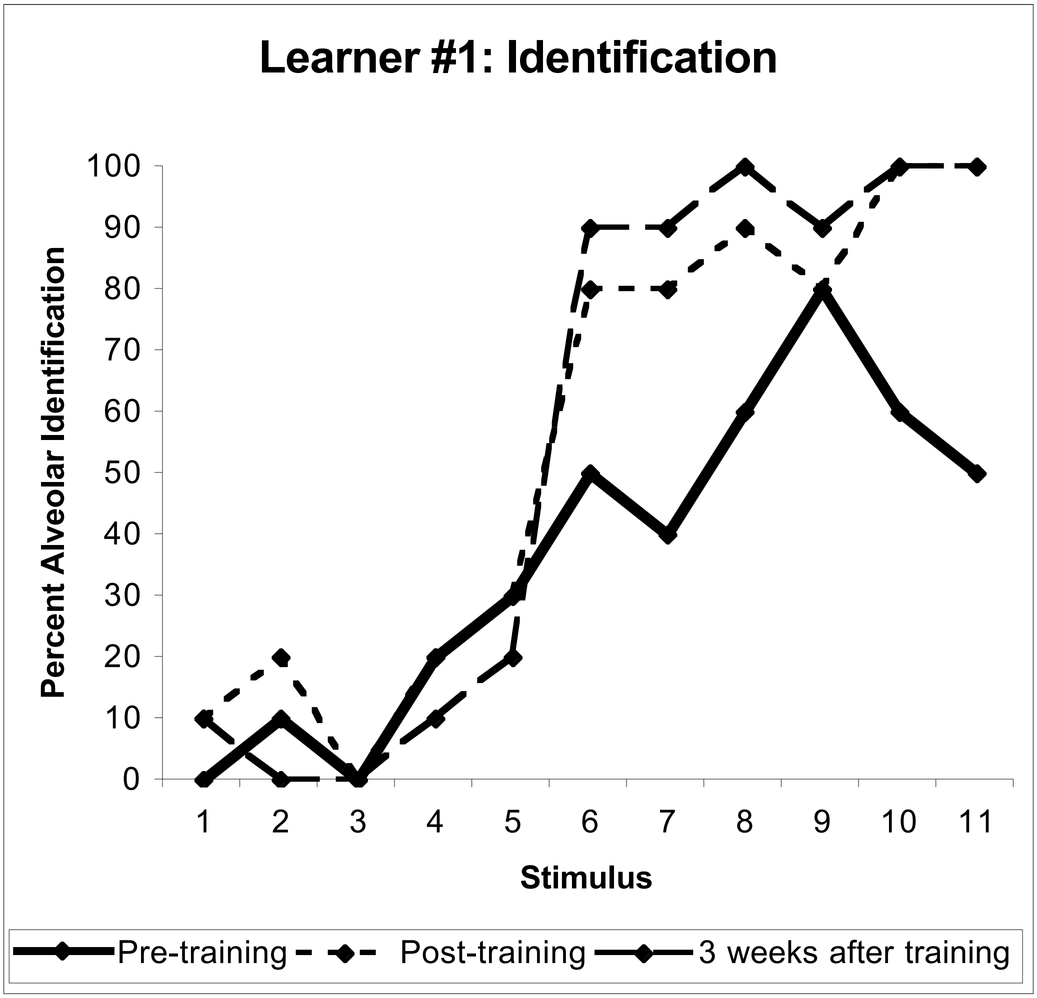

Figure 8.5. Identification functions pre-training (solid line), post-training (dotted line), and three weeks after training (dashed line) for learner #1.

Multidimensional scaling (MDS) analyses based on the difference ratings were also calculated. MDS is a technique used to uncover and visualize proximities in a low dimensional space and is strongly related to methods such as principal component analysis and cluster analysis. Although in many perceptual studies order of presentation of stimuli in a pair is presumed to have no effect

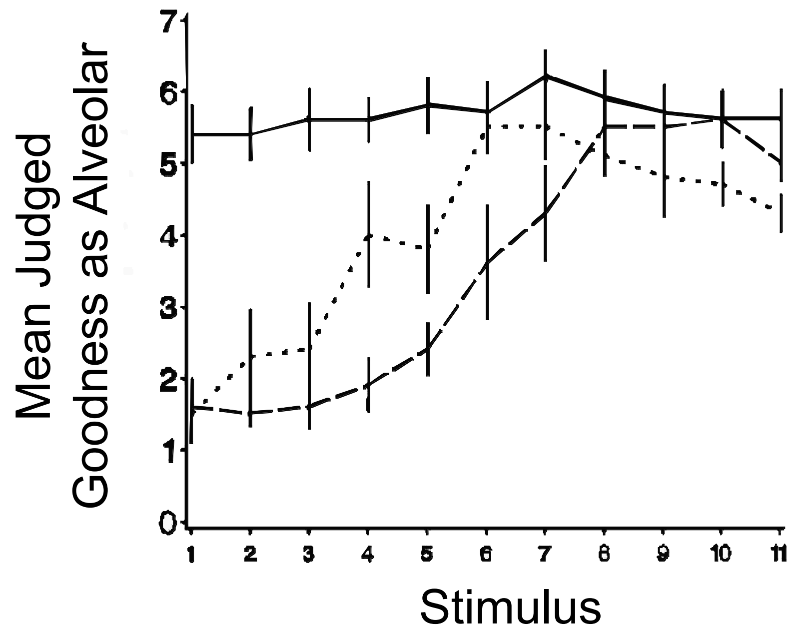

Figure 8.6. Mean judged goodness (error bars indicate one standard deviation) as an exemplar of the alveolar /d/. Pre-training (solid line), post-training (dotted line), and three weeks after training (dashed line) for Learner #1.

(Schiffman, Reynolds, & Young, 1981), our earlier data suggested that order of pair elements might indeed influence difference ratings (in other words, the initial condition, or initial categorization, matters). In the pre-training data, when the first stimulus in a pair is identified as the subject's native category, stimuli that are acoustically closest to the best exemplar are attracted or pulled in; dental-end stimuli cluster separately from the alveolar-end stimuli. When the acoustically more dental stimulus is presented first, there is little if any evidence of stimulus grouping before training. In the post-training and follow-up testing, the dental-first pairs also show an attractive effect, although the effect is still weaker than that observed for the pairs in which the native sound, the alveolar, is presented first. When the day-to-day variability of the MDS solutions is calculated, total variability is relatively low from the beginning of training and quickly decreases over the first six days, remaining low thereafter. The initially higher variability in the total is exclusively due to the degree of clustering across the alveolar-first pairs (Figure 8.7).

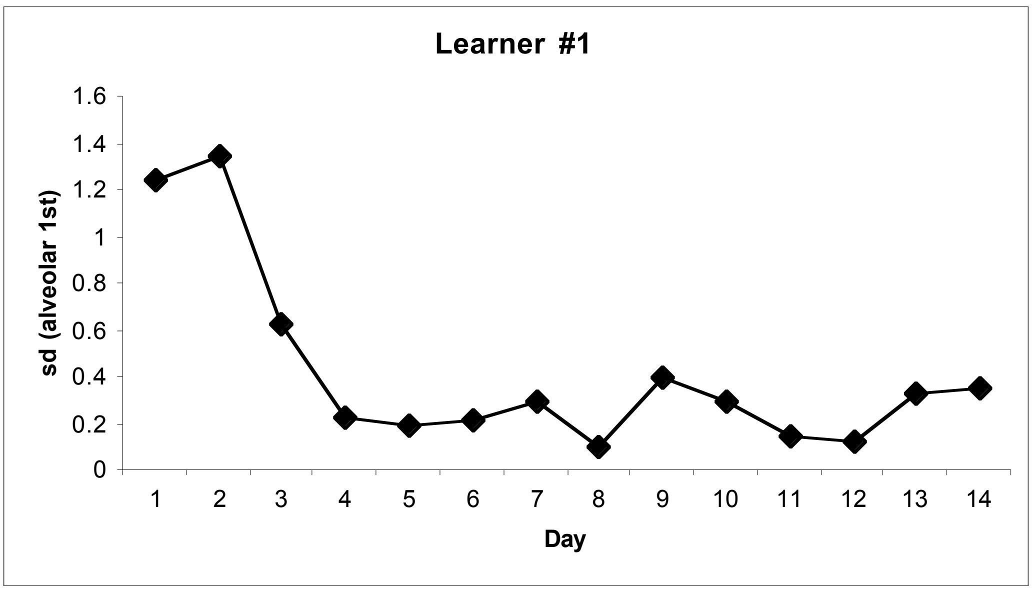

Figure 8.7. Total variability in the MDS analysis, as a function of day of training.

Our second learner showed a very different initial perceptual mapping from learner #1, and a markedly different pattern of learning over time. Pre-training, only stimuli 7 and 8 are identified at levels different from chance (both as alveolar; Figure 8.8) and stimulus 8 is judged as the "best" alveolar (although all stimuli were judged as acceptable members of the alveolar category; Figure 8.9). After training and in follow-up testing, this subject's identification functions showed clear categorization of the stimuli into alveolar and dental, with stimuli on the alveolar end of the continuum now judged to be better exemplars of the alveolar category than stimuli from the dental end. Stimulus 11, judged the best alveolar after training, was also judged a better alveolar than before training (Figures 8.8 and 8.9).

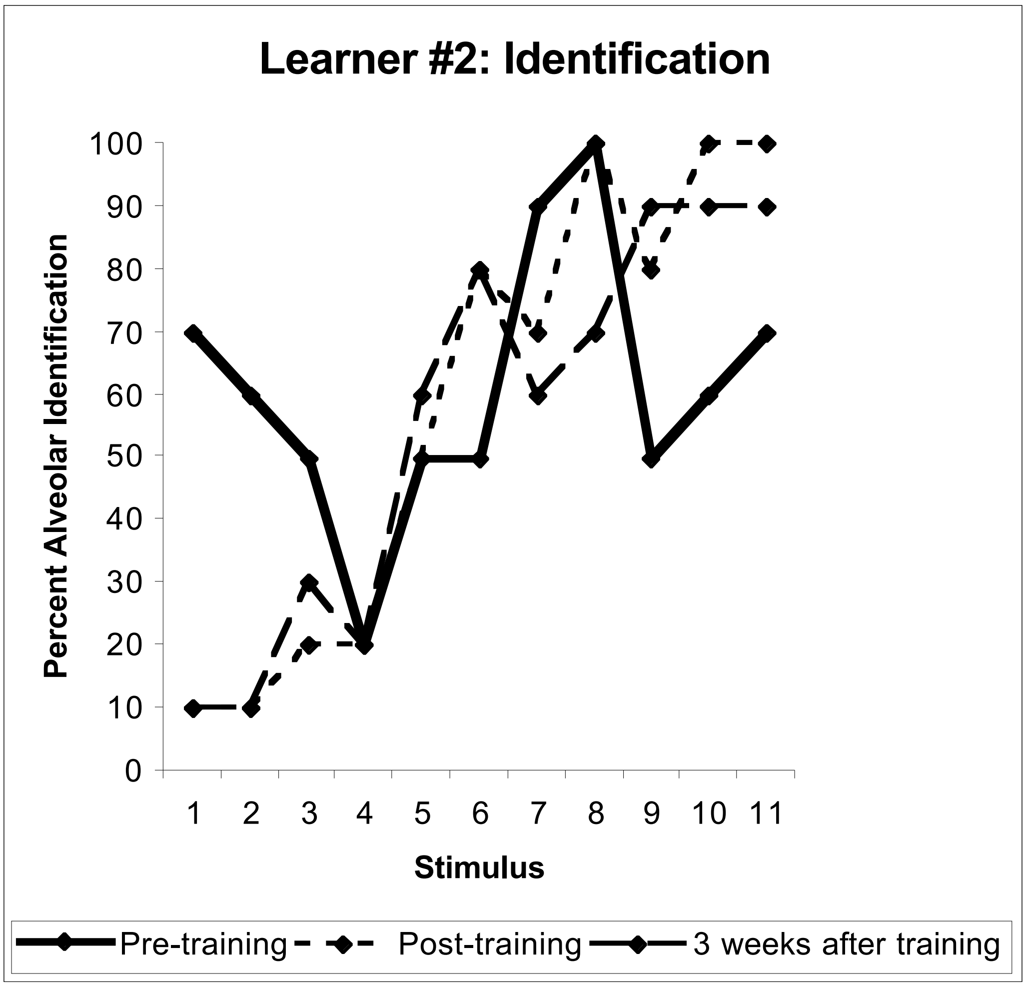

Figure 8.8. Identification functions pre-training (solid line), post-training (dotted line), and three weeks after training (dashed line) for Learner #2.

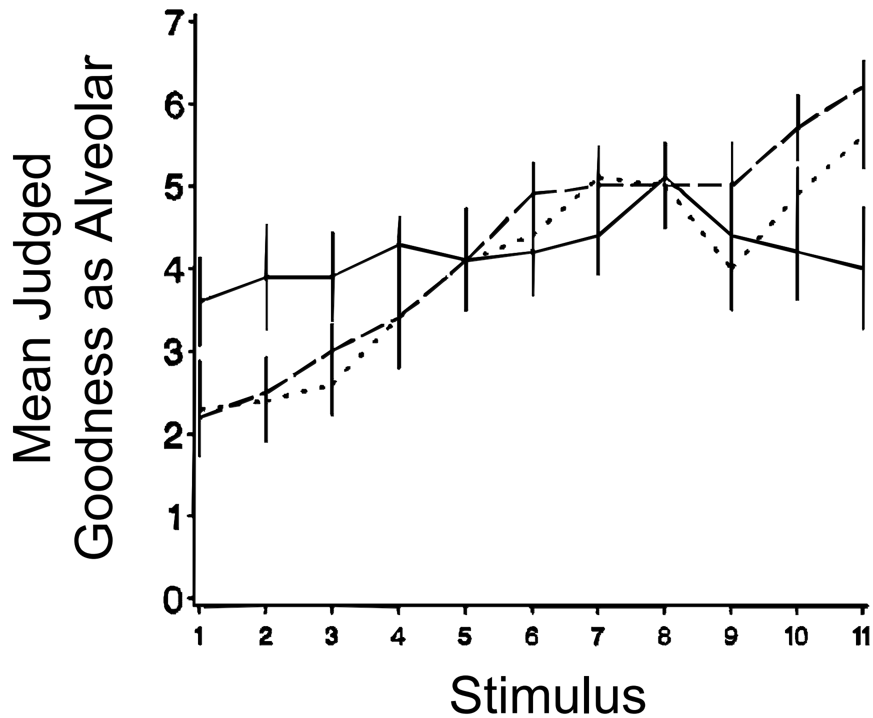

Figure 8.9. Mean judged goodness (error bars indicate one standard deviation)as an exemplar of the alveolar /d/. Pre-training (solid line), post-training (dotted line), and three weeks after training (dashed line) for Learner #2.

The MDS analyses based on difference ratings (taking order into account) revealed that the pre-training solution does not respect acoustic ordering, consistent with the initial identification results. By the time of the post-training evaluation, difference ratings of the alveolar-first pairs showed a tight clustering of stimuli into two groups corresponding to alveolars and dentals; dental-first pairs also grouped, although somewhat more weakly. Grouping of stimuli was tighter in the follow-up as well, with less of an order effect. The total variability in the MDS solutions is shown in Figure 8.10. Total variability was initially much higher than for learner #1 and showed a steady decline until, around Day 5, an increase in variability occurred through Day 9. This increase preceded a sharp drop in total variability at Day 10 to levels equivalent to those observed for learner #1. Note that the peak in variability in judging the alveolar-first pairs may be interpreted as a *destabilization* of the attractor corresponding to the alveolar category.

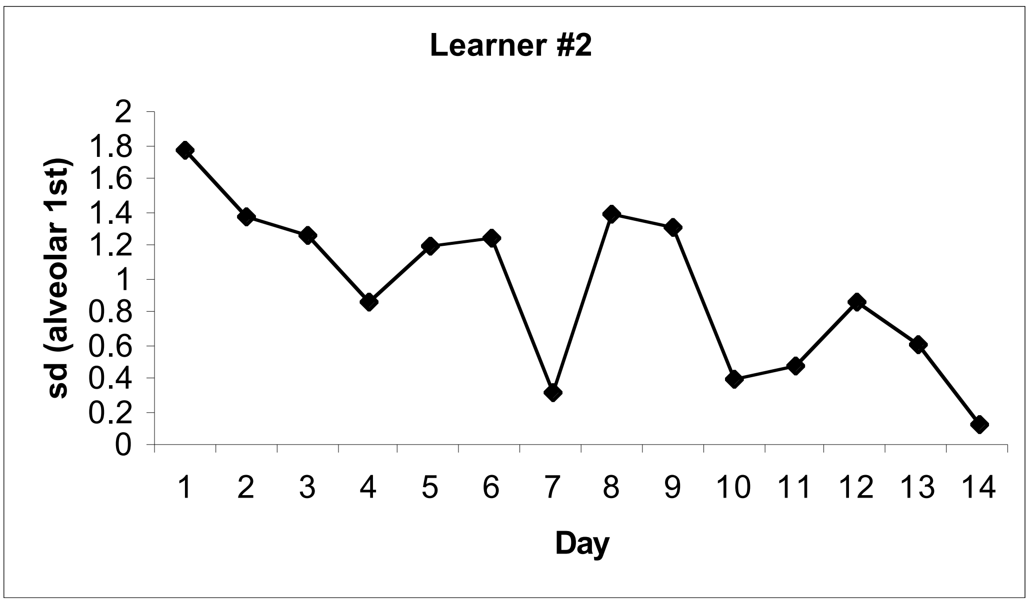

Figure 8.10. Total variability in the MDS analysis, as a function of day of training.

To summarize, learner #1 showed an initial ability to distinguish some of the dental-end stimuli from the alveolar, even though they were still acceptable as alveolars. The pre-, post-, and follow-up test results all indicate a smooth and rapid learning process occurring over the first six days of training and the decrease in variability in his MDS profile was smooth and fast. This is congruent with the initial prediction: If a listener initially can perceive a non-native sound as "different" from a native one, the existing perceptual landscape cooperates with the sound-to-be-learned and learning should be relatively smooth and fast. This pattern is consistent with the idea of progressively stabilizing an already existing stable pattern.

Learner #2 showed little evidence for an initial ability to hear dental-end stimuli as different from alveolar-end stimuli. Variability of the MDS solutions also began at a level nearly three times greater than initial variability for learner #1, and the rate of contraction of the stimuli into groups was slower than for learner #1. After the variability began to decrease, it reversed direction and peaked again just prior to reliable clustering of the MDS solutions. This local increase in variability occurred almost exclusively in alveolar-first pairs and can be considered analogous to critical fluctuations that often precede bifurcations (Schöner, Haken, & Kelso, 1986). Again, these results were congruent with predictions: If a listener initially perceives the non-native sound as indistinguishable from a native one, then learning to recognize the non-native sound competes with the existing perceptual organization. This process is slower than when the initial perceptual landscape cooperates with the new sound, and because this competition entails destabilization of the existing attractor, the bifurcation is marked by high variability.

One aspect of the data that has not yet been highlighted is that learning the non-native category modified perception of the native one (cf. Flege, 1992, 1995), especially for listeners who did not initially parse the stimulus continuum. After learning, not only did the stimulus judged as the best alveolar exemplar shift away from the dental group, but the best exemplar was also a better exemplar post-training than pre-training. Thus the pre-existing phonological organization is malleable. Learning does not entail simply an addition of a new category but in fact changes the existing attractor layout (see also Sancier & Fowler, 1997).

In the cognitive, behavioral, and brain sciences, large strides have been made in understanding pattern formation using the concepts of self-organization and the mathematical tools of nonlinear dynamical systems (e.g., see Haken & Stadler, 1990, for a variety of different contributions in this context; Kelso, 1995). Explicitly dynamical investigations of speech include attempts to identify phonological units with dynamically specified gestures (Browman & Goldstein, 1986, 1989, 1992; Kelso, Saltzman, & Tuller, 1986; Kelso, Tuller, & Harris, 1983), to construct a topology of vowels (Wildgen, 1990) and consonants (Petitot-Cocorda, 1985) in terms of a landscape of attractors and repellers within an articulatory or acoustic space, and to model the phonological system of artificial languages as a self-organized solution of talker-based and listener-based constraints (Lindblom, MacNeilage, & Studdert-Kennedy, 1983). In our own work (Tuller et al., 1994; Case et al., 1995; Tuller, 2003), we demonstrated that changes in perception that occur as the acoustic signal is altered are indicative of a pattern-formation process in perception. A model of the results was proposed and unique predictions of the model were tested and confirmed.

The approach also provides a theoretically motivated way to understand the process of learning to perceive non-native speech sounds (and perhaps the emergence of categories in development). Fundamental to this approach is a methodological stance: Instead of studying features of objectively existing prototypes (either as abstract linguistic entities or as stored multiple exemplars) in a group of listeners, focus on the interaction of an individual perceiver with speech stimuli in context. In this way, we have observed changing patterns of categorization that parallel those observed in perceptuo-motor learning (Kelso, 1990; Kelso & Zanone, in press; Schöner, Zanone, & Kelso, 1992; Zanone & Kelso, 1992, 1994, 1997) and are consistent with the notion that reliably categorizing a new speech sound depends on whether the new category cooperates or competes with an individual's initial perceptual capabilities and that learning serves to reorganize the perceptual space.

In summary, I have described a program of research in which the tenets of dynamical systems and empirical research on speech are mutually informative and directive. In this, I have followed the basic strategy identified by Kelso (in press), but applied to the study of speech perception. This strategy entails (1) Choosing a level of analysis and description that captures the behavior you are studying. (So if I'm interested in how people learn to change their perceptual categorization of speech, it would not be fruitful to choose to describe the behavior in terms of the phasing of harmonics in the signal.); (2) Prune away complications so that the essence of your question remains foremost in the experimental design; (3) Focus on finding the conditions that yield qualitative changes in behavior. Qualitative change allows one to define the perceptual categories clearly as well as to exploit the *patterns* of change as a key to the mechanisms underlying pattern formation (e.g., dynamic instability); and (4) Explore both the coordinative and the component levels as well as the relation between them. How one defines the coordinative level and "one level down" depends on the experimenter's insights into step (1)—choosing the level of description. This last step, deriving the coordinative level dynamics from the usually nonlinear coupling among individual components, is as yet the weakest link in understanding the self-organizing nature of speech dynamics.

Finally, the empirical and modeling strategy described here is both speech-specific and generalizable. The approach has also been fruitfully applied to the verbal transformation effect (Ditzinger, Tuller, Haken, & Kelso, 1997; Ditzinger, Tuller, & Kelso, 1997) and more recently, auditory streaming (Almonte, Jirsa, Large, & Tuller, submitted). It also shares much with studies of the effects of attention on behavioral patterns (e.g., Temprado, Zanone, Monno, & Laurent, 1999), and with studies of learning from behavioral, theoretical, and neurophysiological perspectives (Jantzen, Fuchs, Mayville, & Kelso, 2001; Kelso & Zanone, in press; Kelso, 1995; Schöner, Zanone, & Kelso, 1992; Sporns & Edelman, 1993; Zanone & Kelso, 1992, 1994, 1997). More recently, neural correlates of the stability and change of behavioral coordination have been uncovered using several methods that reveal brain function, such as high density SQuID, multichannel EEG, and functional MRI and PET (Daffertshofer, Peper, & Beek, 2000; Frank, Daffertshofer, Peper, Beek, & Haken, 2000; Fuchs, Jirsa, & Kelso, 2000; Fuchs, Kelso, & Haken, 1992; Fuchs, Mayville, Cheyne, Weinberg, Deecke, & Kelso, 2000; Kelso, Bressler, Buchanan, DeGuzman, Ding, Fuchs, & Holroyd, 1992; Kelso, Fuchs, Holroyd, Lancaster, Cheyne, & Weinberg, 1998; Mayville, Bressler, Fuchs, & Kelso, 1999; Mayville, Fuchs, Ding, Cheyne, Deecke, & Kelso, 2001; Meyer-Lindenberg, Ziemann, Hajak, Cohen, & Berman, 2002; Ullen, Ehrsson, & Forssberg, 2000; Wallenstein, Kelso, & Bressler, 1995). Behavioral investigations have been spurred by, and have spawned, theoretical work at the neural level (Fuchs & Jirsa, 2000; Haken, Kelso, & Bunz, 1985; Jirsa, Fink, Foo, & Kelso, 2000; Jirsa, Friedrich, Haken, & Kelso, 1994; Jirsa & Haken, 1996, 1997; Schöner, Haken, & Kelso, 1986; Schöner, Jiang, & Kelso, 1990; Treffner & Turvey, 1996) that is rapidly becoming more neurobiologically grounded (Frank et al., 2000; Fuchs et al., 2000; Jirsa, Fuchs, & Kelso, 1998; Jirsa & Haken, 1997).

Despite this wealth of information concerning the dynamics of behavior, the specific boundary conditions and control parameters that establish the context for speech phenomena, the coordinative and component levels that makes sense in speech, are specific to speech and must be identified within the speech context. "Dynamics" in and of itself will not give us the answers—it must be fleshed out for each system under study with conceptual content and implementation via experiment, simulation, modeling, and theory development.

REFERENCES

Aaltonen, O., Eerola, O., Hellstrom, A., Uusipaikka, E., & Lang, A. H. (1997). Perceptual magnet effect in the light of behavioral and psychophysiological data. *Journal of the Acoustical Society of America, 101*, 1090-1105.

Almonte, F., Jirsa, V. K., Large, E., & Tuller, B. (submitted). Neural model of streaming in rhythm perception.

Best, C. T, Morongiello, B., & Robson, R. (1981). Perceptual equivalence of acoustic cues in speech and nonspeech perception. *Perception and Psychophysics, 29*, 191-211.

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

Browman, C., & Goldstein, L. (1989). Articulatory gestures as phonological units*. Phonology, 62*, 210-251.

Browman, C., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica, 49*, 155-180.

Case, P. (1996). *Learning to hear new speech sounds: A dynamical approach.* Unpublished doctoral dissertation, Florida Atlantic University, Boca Raton, FL.

Case, P., Tuller, B., Ding, M., & Kelso, J. A. S. (1995). Evaluation of a dynamical model of speech perception. *Perception and Psychophysics, 57*, 977-988.

Daffertshofer, A., Peper, C. E., & Beek, P. J. (2000) Power analysis of event-related encephalographic signals. *Physics Letters A*, *266*, 290-302.

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.

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.

Ditzinger, T., Tuller, B., & Kelso, J. A. S. (1997). Temporal patterning in an auditory illusion: The verbal transformation effect. *Biological Cybernetics, 77*, 23-30.

Ditzinger, T., Tuller, B., Kelso, J. A. S., & Haken, H. (1997). A synergetic model for the verbal transformation effect. *Biological Cybernetics, 77*, 31-40.

Eimas, P. D., & Corbit, J. D. (1973). Selective adaptation of linguistic feature detectors. *Cognitive Psychology, 4*, 99-109.

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.

Flege, J. E. (1992). Speech learning in a second language. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: Models, Research, Implications* (pp. 565-604). Timonium, MD: York Press.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* (pp. 233-277). Baltimore, MD: York Press.

Frank, T. D., Daffertshofer, A., Peper, C. E., Beek, P. J., & Haken, H. (2000). Towards a comprehensive theory of brain activity: Coupled oscillator systems under external forces. *Physica D*, *144*, 62-86.

Fuchs, A., & Jirsa, V. K. (2000). The HKB model revisited: How varying the degree of symmetry controls dynamics. *Human Movement Science*, *19*, 425-449.

Fuchs, A, & Kelso, J. A. S. (1994). A theoretical note on models of interlimb coordination. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 1088-1097.

Fuchs, A., Jirsa, V. K., & Kelso, J. A. S. (2000). Theory of the relation between human brain activity (MEG) and hand movements. *NeuroImage*, *11,* 359-369.

Fuchs, A., Kelso, J. A. S., & Haken, H. (1992). Phase transitions in the human brain: Spatial mode dynamics. *International Journal of Bifurcation and Chaos, 2,* 917-939.

Fuchs, A. Mayville, J., Cheyne, D., Weinberg, H., Deecke, L., & Kelso, J. A. S. (2000). Spatiotemporal analysis of neuromagnetic events underlying the emergence of coordinative instabilities. *NeuroImage*, *12,* 71-84.

Ganong, W. F., & Zatorre, R. J. (1980). Measuring phoneme boundaries in four ways. *Journal of the Acoustical Society of America*, *68*, 431-439.

Giangrande, J., Tuller, B., & Kelso, J. A. S. (2003) Perceptual dynamics of circular pitch. *Music Perception*, *20*, 241-262.

Haken, H., & Stadler, M. (1990). *Synergetics of cognition*. Berlin: Springer-Verlag.

Haken, H. (1977). *Synergetics*, *an introduction: Non-equilibrium phase transitions and self-organization in physics, chemistry, and biology*. Berlin: Springer.

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 and Row.

Iverson, P., & Kuhl, P. K. (1996). Influences of phonetic identification and category goodness on American listeners' perception of /r/ and /l/. *Journal of the Acoustical Society of America*, *99*, 1130-1140.

Jantzen, K. J., Fuchs, A. Mayville, J. M., & Kelso, J. A. S. (2001). Neuromagnetic activity in alpha and beta bands reflects learning-induced increases in coordinative stability. *Clinical Neurophysiology, 112, 1685-1697.*

Jirsa, V. K., Fink, P. W., Foo, P., & Kelso, J. A. S. (2000). Parametric stabilization of biological coordination: A theoretical model. *Journal of Biological Physics*, *26*, 85-112.

Jirsa, V. K., Friedrich, R., Haken, H. & Kelso, J. A. S. (1994). A theoretical model of phase transitions in the human brain. *Biological Cybernetics, 71,* 27-35.

Jirsa, V. K., Fuchs, A., & Kelso, J. A. S. (1998). Neural field theory connecting cortical and behavioral dynamics: Bimanual coordination. *Neural Computation, 10,* 2019-2045.

Jirsa, V. K., & Haken, H. (1996). Field theory of electromagnetic brain activity. *Physical Review Letters, 77*, 960-963.

Jirsa V. K., & Haken, H. (1997). A derivation of a macroscopic field theory of the brain from the quasi-microscopic neural dynamics. *Physica D, 99,* 503-526.

Jongman, A. Blumstein, S. E., & Lahiri, A. (1985). Acoustic properties for dental and alveolar stop consonants: A cross-language study. *Journal of Phonetics*, *13*, 235-251.

Kelso, J. A. S. (1990). Phase transitions: Foundations of behavior. In H. Haken & M. Stadler (Eds.), *Synergetics of Cognition* (pp. 249-268). Berlin: Springer.

Kelso, J. A. S. (1994a). Elementary coordination dynamics. In S. Swinnen, H. Heuer, J. Massion, & P. Casaer (Eds.), *Interlimb Coordination: Neural, Dynamical, and Cognitive Constraints* (pp.301-318). San Diego: Academic Press.

Kelso, J. A. S. (1994b). The informational character of self-organized coordination dynamics. *Human Movement Science, 13*, 393-413.

Kelso, J. A. S. (1995). *Dynamic patterns: The self-organization of brain and behavior.* Cambridge: MIT Press. [Paperback edition, 1997].

Kelso, J. A. S., Bressler, S. L., Buchanan, S., DeGuzman, G. C., Ding, M., Fuchs, A., & Holroyd, T. (1992). A phase transition in human brain and behavior. *Physics Letters A, 169,* 134-144.

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). Santa Fe, NM: Addison-Wesley Publishing Co.

Kelso, J. A. S., Fuchs, A., Holroyd, T., Lancaster, R., Cheyne, D., & Weinberg, H. (1998). Dynamic cortical activity in the human brain reveals motor equivalence. *Nature, 392,* 814-818.

Kelso, J. A. S., Saltzman, E., & Tuller, B. (1986). The dynamical perspective on speech production: Data and theory. *Journal of Phonetics, 14*, 29-60.

Kelso, J. A. S., Tuller, B., & Harris, K. (1983). Converging evidence for the role of relative timing in speech. *Journal of Experimental Psychology: Human Perception and Performance*, *9*, 829-835.

Kelso, J. A. S., & Zanone, P. G. (in press). Coordination dynamics of learning and generalization across different effector systems. *Journal of Experimental Psychology: Human Perception & Performance*.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*, 431-461.

Liberman, A. M., Harris, K. S., Hoffman, H., & Griffith, B. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*, 358-368.

Lindblom, B., MacNeilage, P., & Studdert-Kennedy, M. (1983). Self-organization processes and the explanation of phonological universals. In B. Butterworth, B. Comrie, & O. Dahl (Eds.) *Explanations of Linguistic Universals* (pp. 181-203). Berlin: Molton.

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/: II. Their role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America*, *94*, 1242-1255.

Mayville, J. M., Bressler, S. L., Fuchs, A., & Kelso, J. A. S. (1999). Spatiotemporal reorganization of electrical activity in the human brain associated with a phase transition in rhythmic auditory-motor coordination. *Experimental Brain Research*, *127*, 371-381.

Mayville, J. M., Fuchs, A., Ding, M., Cheyne, D., Deecke, L., & Kelso, J. A. S. (2001) Event-related changes in neuromagnetic activity associated with syncopation and synchronization tasks. *Human Brain Mapping*, *14*, 65-80.

Meyer-Lindenberg, A., Ziemann, U., Hajak, G., Cohen, L. & Berman, K.F. (2002). Transitions between dynamical states of differing stability in the human brain. *Proceedings of the National Academy of Science, 99,* 10948-10953.

Nicolis, G., & Prigogine, I. (1977). *Self-organization in nonequilibrium systems*. New York: Wiley.

Petitot-Cocorda, J. (1985) *Les catastrophes de la parole. De Roman Jakobson à René Thom.* Paris: Maloine.

Pisoni, D. B., & Lively, S. E. (1995). Variability and invariance in speech perception: A new look at some old problems in perceptual learning. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* (pp. 433-459). Baltimore, MD: York Press.

Repp, B. H., & Liberman, A. M. (1987). Phonetic categories are flexible. In S. Harnad (Ed.), *Categorical Perception* (pp. 89-112). Cambridge, UK: Cambridge University Press.

Sancier, M., & Fowler, C. A. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics*, 25, 421-436.

Schiffman, S. S., Reynolds, M. L., & Young, F. W. (1981). *Introduction to multidimensional scaling*. New York: Academic Press.

Schöner, G., & Kelso, J. A. S. (1988). A synergetic theory of environmentally-specified and learned patterns of movement coordination. I. Relative phase dynamics. *Biological Cybernetics, 58,* 71-80.

Schöner, G., Haken, H. & Kelso, J. A. S. (1986). A stochastic theory of phase transitions in human hand movement. *Biological Cybernetics*, *53*, 442-452.

Schöner, G., Jiang, W.-Y., & Kelso, J. A. S. (1990). A synergetic theory of quadrupedal gaits and gait transitions. *Journal of Theoretical Biology, 142*, 359-391.

Schöner, G., Zanone, P. G., & Kelso, J. A. S. (1992). Learning as change of coordination dynamics: Theory and experiment. *Journal of Motor Behavior, 24*, 29-48.

Sporns, O., & Edelman, G. M. (1993). Solving Bernstein's problem: A proposal for the development of coordinated movement by selection. *Child Development, 64,* 960-981.

Temprado, J. J., Zanone, P. G., Monno, A., & Laurent, M. (1999). Attentional load associated with performing and stabilizing preferred bimanual patterns. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 1595-1608.

Treffner, P. J., & Turvey, M. T. (1996). Symmetry, broken symmetry, and the dynamics of bimanual coordination. *Experimental Brain Research, 107,* 463-478.

Tuller, B. (2003). Computational models in speech perception. *Journal of Phonetics*, *31*, 503-507.

Tuller, B., Case, P., Ding, M., & Kelso, J. A. S. (1994). The nonlinear dynamics of speech categorization. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 1-14.

Ullen, F., Ehrsson, H. H., & Forssberg, H. (2000). Brain areas activated during bimanual tapping of different rhythmical patterns in humans. *Society for Neuroscience Abstracts*, *26*, 458.

Wallenstein, G. V., Kelso, J. A. S., & Bressler, S. L. (1995). Phase transitions in spatiotemporal patterns of brain activity and behavior. *Physica D, 84,* 626-634.

Wildgen, W. (1990). Basic principles of self-organization in language. In H. Haken & M. Stadler (Eds.), *Synergetics of Cognition* (pp. 415-426). Berlin: Springer-Verlag.

Zanone, P. G., & Kelso, J. A. S. (1992). The evolution of behavioral attractors with learning: Nonequilibrium phase transitions. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 403-421.

Zanone, P. G., & Kelso, J. A. S. (1994). The coordination dynamics of learning: Theoretical structure and experimental agenda. In S. P. Swinnen, H. Heuer, J. Massion, & P. Casaer (Eds.), *Interlimb Coordination: Neural, Dynamical, and Cognitive Constraints* (pp. 461-490). San Diego: Academic Press.

Zanone, P. G., & Kelso, J. A. S. (1997). The coordination dynamics of learning and transfer: A multilevel study. *Journal of Experimental Psychology: Human Perception and Performance*, *23*, 1454-1481.