Language Dynamics: Sound Categorization

Betty Tuller
Florida Atlantic University, Complex Systems and Brain Sciences, 777 Glades Road, Boca Raton FL 33435
National Science Foundation, 4201 Wilson Blvd, Suite 995, Arlington VA 22230
btuller@nsf.gov

Abstract

Speech perception depends largely on information in an acoustic stream that is inherently dynamic: it changes constantly and fades rapidly. Transitions occur on many levels, from those between sound segments to syntactic adjustments. The present paper describes a program of research that is based on the idea that speech sound categorization occurs via transitions between attractors in a dynamical system. The dynamical system is structured by the native phonological system, as it is instantiated in the individual listener. Phonological learning is conceptualized as the emergence of attractors in the existing dynamical system. Nonlinear dynamic models, constrained by empirical data, are used to inform further research.

Introduction

The history of research on speech perception is replete with examples of nonlinearities, or threshold phenomena, relating acoustics to perception. These nonlinearities are essential in that they allow stable communication despite variation in the acoustic signal across speakers, emphasis, background noise, etc. Furthermore, the range of acoustic signals perceived as equivalent is much larger for speech sounds than for nonspeech. For example the first sound in "kit" and "key" are usually perceived as the same, but the initial noise bursts extracted from "kit" and "key" are clearly different.

Also, the minimal acoustic change in speech segments that can be detected is not equal throughout the range of the acoustic parameter manipulated. When speech stimuli are categorized as the same linguistic segment, people have a great deal of difficulty discriminating between them. As the acoustic parameter continues to change, the detection of the change tends to be abrupt not continuous. This means that there is a perceptual warping of the physical acoustic signal that is conditioned by one’s linguistic categories. This restricted ability to discriminate between stimuli that are identified as the same speech sound is termed categorical perception (Liberman, Harris, Hoffman, and Griffith, 1957). The phenomenon has been replicated many times with a wide range of stimuli in different contexts. The boundaries, or thresholds between categories are flexible, adjusting with factors such as individual speaker differences, different listeners, phonetic context, speaking rate, and linguistic experience (see Repp, 1984, for a review).

A form of categorical perception occurs constantly outside the laboratory, as when different speakers produce the “same” word or when a speaker says the “same” word quickly or slowly. This means that speech perception cannot be a simple concatenation of sound elements to yield syllables, syllables to yield words, or words to yield sentences. The interdependency across scales reveals a complex system with nonlinearly interacting elements that somehow allow veridical communication.

The preponderance of nonlinearities and threshold phenomena in speech can be viewed not as indicating immutable boundaries but rather as transitions between categories. The focus of research then becomes the conditions that allow the emergence of another category, or state, in a self-organized, non-directive fashion. Behavior around the transitions may reveal a pattern formation process with its own perceptual dynamics (multistability, loss of stability, etc.). The aim here is to characterize the time-dependent behavior of the speech system in terms of its (nonlinear) dynamics, especially the stability and change of perceptual forms. The framework of dynamical systems theory is used to uncover fundamental principles that may be common to many domains, including the perception of speech.
Nonlinearities in perception as a window into the underlying dynamics.

Can dynamics or adaptive systems theory reconcile the highly context-dependent sensitivity of speech perception with the characterization of language as having a limited number of stable states (meaningful sounds in the language)? In a dynamical view, the stable states are not purely symbolic representations but themselves have a dynamic. They are viewed as attractors, which allow the system to perform the discretization of perceptual space associated with abstract perceptual categories.

Early experiments aimed at discovering whether there is a dynamic to speech categorization used synthesized speech stimuli in which a silent gap was inserted between the “s”-noise and the “ay” in the word say (Case, Tuller, Ding & Kelso, 1995; Tuller, Case, Ding & Kelso, 1994). When gap durations are short (or absent), monolingual speakers of American English identify the word as say; with long silent gaps, they identify the word as stay. Listeners heard the entire set of stimuli presented in order. For example, one sequence started with the stimulus with 0ms silent gap. Each successive stimulus had a 4ms longer gap until it reached 76ms. Then the silent gap after the "s" decreased in 4ms steps until 0ms was reached. Listeners were required to identify each stimulus as either say or stay. The point where each listener switched from one word to the other as the silent gap (the control parameter) increased, and where they switched when it decreased, was determined for each trial. Hysteresis (the tendency for the listener's response to persist across the sequence of stimuli) and contrast (the listener quickly switches to the alternate percept and does not hold on to the initial categorization) were the predominant patterns of perception, although hysteresis was initially far more frequent than contrast. The presence of hysteresis and contrast implies bistability in perception: two percepts are possible for the same acoustic token.

We also evaluated the variability around the switch point by adding “noise” in the form of repetition of stimuli. If the transition is dynamic and the time scales of the parameter change are appropriate, perceptual instability should increase as the transition to another percept approaches. Perception should also become more stable just after the switch. Regardless of whether the dominant pattern was hysteresis or contrast, noise increased the likelihood of a perceptual switch only in the bistable region. Moreover, perceptual variability peaked just before the transition and dropped sharply thereafter.

These results have been modeled as a simple nonlinear dynamical system with two potential stable states (Case et al., 1995; Tuller et al., 1994). Phonological categories are considered equivalent to attractors (stable behaviors of the system) and switching between phonological categories means changes in the relative stability of the attractors. Thus, there exist ranges of parameter variation within which the perceptual form remains relatively stable (i.e., resistant to change despite acoustic variation or noise). In other ranges, however, even small variations in the acoustic parameter can result in large (nonlinear) changes in categorization of the input. These perceptual changes are enhanced in the presence of noise. At threshold values, which are sensitive to context, history, linguistic factors, etc., the existing attractor(s) lose stability and the observed behaviors may change gradually or abruptly as other attractors dominate. This model, although far from realistically elaborated, has predicted perceptual categorization behavior by both American and French listeners (Nguyen, Lancia, Bergounioux, Wauquier-Gravelines & Tuller, 2005; Nguyen, Wauquier-Gravelines, Lancia & Tuller, 2007) that was not expected on prior theoretical grounds.

### Complex adaptive systems and language learning.

Does this theoretical framework, and the special status of transitions, illuminate how new attractors form? Is learning to perceive a speech sound that is not in one’s native language akin to the emergence of a new attractor? How does the process differ as a function of initial conditions (an individual’s initial perceptual capabilities)? What is the response of the existing population of attractors? What are the dynamics of learning over time?

Learning a second language as an adult is typically much more difficult than in childhood (Flege, 1995; Iverson et al., 2003; Kuhl, 2000). This is not surprising given that the neuroplasticity typical of early development is not usually maintained into adulthood (Long, 1990; Pallier et al., 2003; Scovel, 2000). One of the barriers to adult second language learning is that speech sounds in the target language sound not occur in one’s native language and are not reliably perceived. A well-known example of this is Japanese listeners trying to hear the distinction between "r" and "l" in English. However, some distinctions are easier for adults to learn than others. In general, ease of learning a distinction between a native sound and a nonnative one is inversely related to their similarity (Flege et al., 2003). Kuhl and her colleagues (e.g. Kuhl & Iverson, 1995) have suggested that prototypical sounds in the native language function as “perceptual magnets,” causing acoustically similar tokens to be perceived as members of the same linguistic category. All else being equal, when nonnative sounds are attracted or drawn to the same native prototype, they are less discriminable from each other and from the category prototype. Conversely, when nonnative sounds are attracted to different native prototypes, they are perceptually distinct.

The hypothesis here is that the theoretical and
empirical approach to changes in categorization of one's own language sounds may also reveal how people learn to perceive a speech sound that is not in their native language. Both situations may entail nonlinear changes in categorization in a complex adaptive system. In our experiments, monolingual American English listeners tried learning to perceive the difference between the voiced alveolar stop consonant /d/ and the voiced dental stop consonant /ð/ (e.g., Tuller, Jantzen, & Jirsa, 2008). The major articulatory distinction between these two sounds is in place of articulation— in /d/ the tongue tip is placed against the alveolar ridge, while in /ð/, the tongue tip is against the upper front teeth. There is no phonemic contrast between these sounds in English (that is, they act as equivalent), although both occur as phonemes in other languages, for example Malayalam. The subjects trained on words containing either /d/ or /ð/, produced by native Malayalam speakers for about one hour each day for 15 days. During each training session, they performed a 2-alternative forced choice identification task (was it the English /d/ or the "foreign" /ð/? with feedback.

A synthesized continuum of 11 stimuli that ranged from the native /d/ to the foreign /ð/ was constructed and used to test whether people had learned the distinction between sounds and how the process of learning the distinction unfolded over the 15 days. Subjects rated the difference between nearest neighbor synthetic stimuli before the first training session, during each training session, and after the last training session. The difference-ratings were used to determine the warping (if any) of the linear changes in acoustic composition of the stimuli into a "perceptual space" and whether the arrangement of stimuli in perceptual space changed as learning progressed. A multidimensional scaling (MDS) analysis on the different rating matrix was used to assess the structure of perceptual space. Before training, all subjects perceive the stimulus relationships in ways that do not follow the physical acoustic order and no clear grouping of stimuli occurs. Over time and training, two patterns of learning emerge.

In one learning pattern, the stimuli partition over training days into two distinct groups corresponding to the native and foreign categories. Over time, members within each group become closer to each other in perceptual space. As in categorical perception, within-category distinctions are small and are non-linearly related to acoustic differences (often a large acoustic shift is necessary before there is a perceptible change). Cross-category distinctions are perceptually larger than would be expected on the basis of the acoustics alone. The day-to-day variability in the MDS is low for stimulus pairs with the acoustically more foreign stimulus presented first, and it continues to decline over training sessions. When the acoustically more native stimulus is presented first, the MDS variability increases over the first few days of training and peaks just prior to the split into two perceptual groups. In other words, when a token that is a better exemplar of one's native language category is presented as the first member of the stimulus pair, it acts as an attractor, pulling the non-native token closer in perceptual space. In order for tokens to be perceived reliably as members of the foreign category, there must be a bifurcation in perceptual space, a process that typically exhibits an increase in variability prior to settling into the new, multi-attractor space. In short, this learning pattern suggests that true phonological learning has occurred. The mechanism is the loss of stability in the existing category attractor and a transition or bifurcation as the new attractor emerges.

In the second learning pattern, the stimuli progressively align in perceptual space according to their acoustic composition. In other words, within-category stimuli are no more like each other perceptually than predicted by the linear acoustic differences. The day-to-day variability in MDS does not show any effect of stimulus order and does not show evidence of a bifurcation. The most likely explanation is that these subjects label stimuli by applying a criterion threshold, determined on the basis of small acoustic differences to which they get more sensitive with training.

Results converge on the idea that there are two modes of learning, one mode based on learning a phonological category and the other based on increasing sensitivity to acoustic differences. Only the phonological learners act consistent with the description of a self-organization of perceptual space in which a bifurcation creates an additional attractor corresponding to a newly learned category. A possible dynamical model that can incorporate both patterns of learning has been proposed and predictions of the model have been defined that are empirically testable (Tuller et al., 2008). The model’s behavior is constrained by the observed data. It assumes that there are two basic modes of category learning that are in competition with each other (since no subject was observed to exhibit both modes of learning). Each mode is described by a time-dependent variable measuring either the degree to which phonology structures the percept over time or the degree to which the acoustics structures the percept over time. These two time-dependent variables define the phase space for category learning. Depending on the portion of phase space in which the learner begins, and the phase space to which the learner evolves, the process of learning can be smooth or can exhibit a bifurcation (reflected in the increasing variability before the tipping point).

Conclusion

In summary, dynamical systems and complex
adaptive systems provide a theoretically motivated way to understand how listeners identify tokens of a speech category, shift between categories, and learn new categories. Fundamental to this approach is the focus on relative stability of attractors within the speech system. Acoustic properties, frequency of occurrence of perceived categories, trajectory of speech sounds in the acoustic space, training, native language, and so on, all can affect the relative stability of categories and in turn, switching behavior. From this stems the wide range of phenomena observed, such as hysteresis, contrast, and bifurcations that are hallmarks of adaptive systems. These theoretical and methodological tools are being used to identify fundamental principles of categorization in language across many scales, linking the analysis of sound change in individuals to the changing structure of languages over historical time (Lightfoot, 2006).

References