Learning to discriminate English /r/ and /l/ in adulthood: Behavioral and modeling studies

James L. MCCLELLAND, Stanford University

Abstract

I describe a body of work undertaken to explore the effect of experience on the perception of speech sounds. The work is undertaken within the context of my overall theoretical perspective, in which language is viewed as reflecting the influence of graded rather than categorical constraints and in which experience gradually shapes the way we respond to the sounds we experience, including sounds in our native language and sounds in foreign languages. Focusing on the case of Japanese adults learning the contrast between English /r/ and /l/, I present a neural network simulation model that captures how experience shapes perception during native language exposure during childhood. I go on to describe predictions from the model about the effects of different training regimes on learning the English /r/-/l/ contrast after acquiring natural Japanese spoken language categories. Then, I present the results of several experiments, partially supporting the model's predictions but also demonstrating some limitations Japanese adults face in learning the English /r/-/l/ contrast that are not fully captured by the model. Overall, the work suggests that perceptual change is possible in adulthood, with some limitations that remain to be addressed in future work.

1. Introduction

My research is grounded in an approach to understanding language and cognition in which graded or continuous rather than categorical constraints shape processing, and in which linguistic objects, such as features, phonemes, syllables, and words are viewed as essentially continuous rather than discrete in nature. Papers describing this view, which I outlined in my lecture to the Japanese Society for Language Sciences in 2013, are available elsewhere (Bybee & McClelland, 2005; McClelland & Bybee, 2007; McClelland, 2014). In the present paper, presented in a symposium talk at the same JSLS meeting in 2013, I provide a more detailed review of work by my group on the role of experience in the perception of speech sounds, focusing on the contrast between the English /r/ and /l/ phonemes. The work illustrates the graded and continuous nature of speech categories and the gradual cumulative effects of experience in shaping these categories in early life. The work also addresses the mechanisms that support change and provides some evidence relevant to one explanation for the reasons why speech perception often becomes resistant to change in adulthood.

Studies in Language Sciences: Journal of the Japanese Society for Language Sciences, 13, 32−52.
2014, Kaitakusha, Tokyo, Japan.

2. Explanations for the loss of sensitivity to non-native speech contrasts in adulthood: The entrenched attractor theory

As is widely known, acquisition of certain aspects of language appears to progress most easily in early childhood (Lenneberg, 1967). For example, children appear to be sensitive to most of the phonemic contrasts made in the world's languages early in life, but begin to lose sensitivity to contrasts they are not exposed to very soon (Kuhl, Williams, Lacerda, Stevens & Lindblom, 1992). Extensive exposure during the early years may reverse this loss, but adults often appear less flexible, with change occurring gradually if at all (Flege, 1995).

The basis of these effects is not fully understood. Explanations range from a biological switch that shuts off or reduces plasticity after puberty (Lenneberg, 1967), to influences exerted by native speech categories on perception of non-native sounds (Best & Tyler, 2007; Kuhl et al., 1992; Flege, 2003), to yet other explanations based on differences in extent of immersion in non-native language contexts; the young may immerse themselves the most, while the old may interact primarily with those who speak their mother tongue, even when living in a country where few speak their language (Flege & Liu, 2001).

My exploration of these issues arose out of an interest in the possibility that influences exerted by native speech categories might be an important contributing factor, based on neural network models that implement a distorting influence of speech categories on perception (Anderson, Silverstein, Ritz & Jones, 1977, Vallabha & McClelland, 2007), producing what Kuhl (1991) has described as *perceptual magnet effects.* I call this view the *entrenched attractor theory*, using the word *attractor* common in neural network and dynamical systems research for essentially the same idea Kuhl describes with the phrase *perceptual magnet.* I continue to believe there is considerable validity to this view — it appears to provide a basis for understanding quite a bit of what we know about how experience and training can influence perceptual abilities, and the present article will primarily focus on explicating this idea. However, for reasons that I will explain at the end of this article, it no longer seems that this idea alone is sufficient to explain the difficulties adult learners face when trying to master the speech contrasts in a non-native language, at least in the case of the difficult English /r/-/l/ contrast for native Japanese speakers.

3. A network model implementing the entrenched attractor theory

I begin by describing a neural network model that implements the entrenched attractor theory as a possible partial explanation for the reduction in plasticity in adulthood. Precursors to the model were described in a series of papers (McClelland, Thomas, McCandliss, & Fiez*,* 1999; McClelland, 2006), and it was most fully realized in an implemented computational model in Vallabha & McClelland (2007).

Figure 1 provides an illustration of the model and of the small part of phonological space on which it was trained. It attempts to capture the initial acquisition of relevant Japanese speech categories, and the subsequent effect of these on learning to distinguish the English phonemes /r/ and /l/.

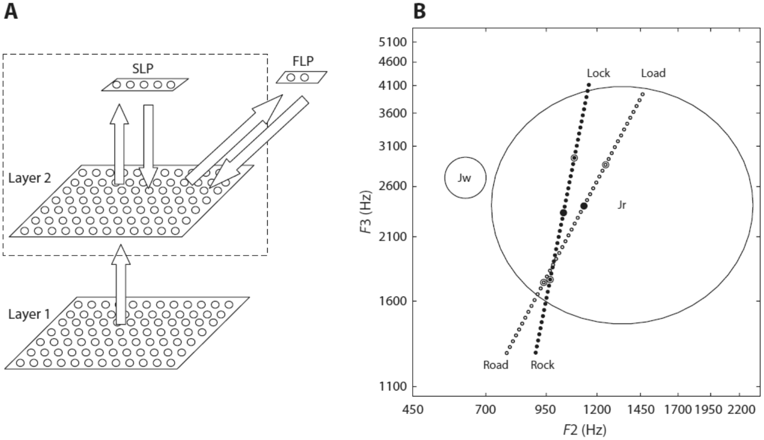

Figure 1. (A) Attractor network model used in simulations of first and second language learning by Vallabha & McClelland (2007). Circles represent neuron-like processing units, and arrows indicate connections between units (where an arrow is shown, there is a separate connection from each unit at the sending end of the arrow to each unit at the receiving end. Bi-directional connections implement attractors in the network that produce perceptual magnet effects. Within layers (not including L1) each unit excites itself and inhibits all other units (see Figure 2). (B) Relation of natural speech training stimuli from the studies of McCandliss et al. (2002) to the Japanese /ɾ/ and /w/ attractors. The attractors are represented by the circles labeled Jr and Jw. The stimulus continua labeled with English words (Lock, Rock; Road, Load) correspond to those shown in Figure 4 below. From Vallabha & McClelland (2007), Success and failure of new speech category learning in adulthood: Consequences of learned Hebbian attractors in topographic maps. *Cognitive, Affective and Behavioral Neuroscience,* Fig. 4, p. 60.

The network is initially trained on examples of the Japanese tap phoneme /ɾ/ and the Japanese /w/ phoneme, corresponding to the sounds at the beginning of the words *ryokan* and *wasabi*, respectively. The intention is to capture the experience we imagine a native Japanese speaker would have with inputs in this space during acquisition of his or her native language. Each experience of a /ɾ/ or /w/ input is treated as a bump of activity in the two-dimensional input space shown in Figure 1B; here one dimension corresponds to the onset frequency of the second formant and the other dimension corresponds to the onset frequency of the third formant. The circles drawn on the diagram represent the range of inputs in this space that the model treats as corresponding to the /ɾ/ and /w/ attractors: After learning, the processes operating in the model distort inputs falling within either of the two circles, ‘pulling’ them toward the center of the circle. I now describe how the model operates.

The input layer of the neural network provides a layer of neuron-like processing units, represented in the figure by small circles, whose activations are driven by phonological inputs. Thus, for example, an input right at the middle of the /ɾ/ region will activate input units near the center of the input layer of the network. These input units then have connections onto the units in the second layer of the network. These projections are pre-specified so that an input at a given position produces a ‘Gaussian bump’ of activity at the corresponding location in layer 2. However, the activations of these units in layer 2 are also affected by learned bidirectional connections between layer 2 and the slow-learning pool (SLP), a pool of units that gradually learns about the distributions of inputs in its native language, and implements the attractors in the network.

To illustrate how this model implements learned attractors that affect the perception of speech sounds and that produce a perceptual magnet effect, we present a simplified version of the model in Figure 2. In this version, instead of the two-dimensional input space of the full model, the input space has only a single dimension, which could be visualized as a line through the input space in Figure 1. As Figure 2 illustrates, the model contains within-layer as well as between layer connections. Within L2 and L3, each unit has a tendency to excite itself and a tendency to inhibit all of the other units in the same layer. When an input is presented to the network, activations are updated gradually over time, and are affected by both the fixed and the learned connections. Activations in L1 are considered fixed for the duration of a given stimulus presentation. Activations in L2 and L3 are updated over a series of 30 small time steps (each corresponding to approximately 15 milliseconds of real time), simulating a perceptual process thought to unfold over about one-half of a second. The ‘percept’ in this model corresponds to the resulting pattern on Layer 2 at the end of the simulation.

The connections between L2 and L3 are initialized to very small random values. These random weights create a situation in which distinct patterns on L2 tend to activate distinct units on L3, and the inhibition between L3 units accentuates this, so that different inputs effectively recruit different units at L3 to represent them as distinct perceptual categories. At first, however, the strengths of the connections between L2 and L3 are so small that the circuit of activation from L2 to L3 and back exerts only a very small influence on activations in L2. However, after every presentation of an input, these connection weights are adjusted, so that attractors reflecting the inputs to the network gradually develop, and come to strongly influence perception.

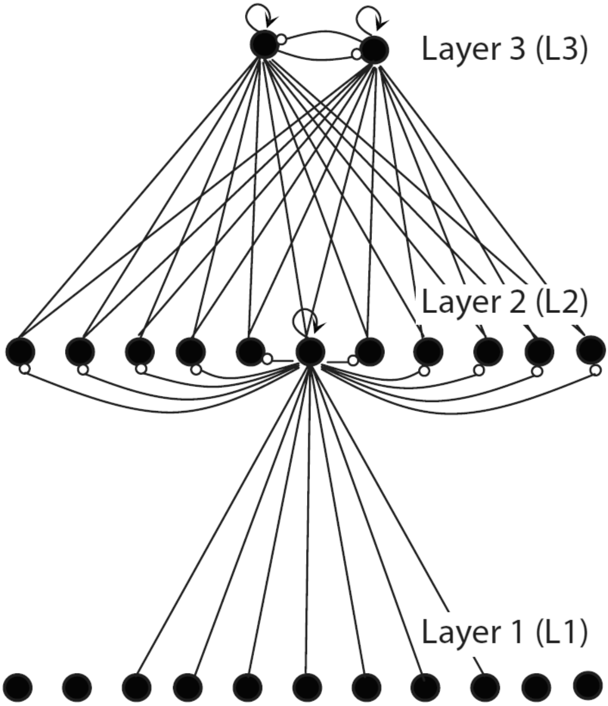

Figure 2. The simplified one-dimensional attractor neural network model used to illustrate the basic properties of attractor networks. All units in layer 2 receive connections from units in layer one from a restricted region of layer 1, as shown for a single layer 2 unit. Connections between units in layers 2 and 3 go in both directions. Each unit within a layer excites itself and inhibits all other units. From Vallabha & McClelland (2007), Success and failure of new speech category learning in adulthood: Consequences of learned Hebbian attractors in topographic maps. *Cognitive, Affective and Behavioral Neuroscience,* Fig 1a, p. 57.

For the simplified model in Figure 2, training experiences were distributed according to two normal distributions centered at two different points in the input space of the network, spanned by a set of eighty input units. For simplicity in illustrating how a single attractor works, we consider here attractors that are spaced very far apart (one centered on unit 28, the other on unit 51), though in Vallabha & McClelland (2007) we also considered more closely spaced attractors.

We now introduce how learning occurs in the network, as a result of changes in connection strengths, according to a simple associative (or *Hebbian*) learning rule. According to Donald Hebb (1949), as paraphrased by many later writers, 'neurons that fire together wire together'. In the model, what this means is that strengths of connections between units that are active at the end of the settling process are increased according to the rule:

$$\Delta w_{ij} = \varepsilon a_i a_j$$

In this expression Δw_{ij} refers to the change in the connection weight to unit i from unit j, a_i and a_j correspond to the activations of units i and j respectively, and ε is a learning rate constant that controls the rate of learning. In keeping with our belief that attractors become entrenched gradually over the course of learning, we keep the learning rate very small, so that each experience leads to only a very small change in the strengths of the connections. Each time the connection weights are updated, a constraint is imposed on the sum of the set of connection weights coming into each L3 unit from L2 and on the set of connection weights coming out of each L3 unit to L2, so that the sum of the weights in each set cannot exceed a fixed maximum value (see Vallabha & McClelland, 2007, for details).

In Figure 3 we show how activations on Layer 2 are 'warped' during the activation process that simulates perception, once experience driven by Hebbian learning has created an attractor centered on input unit 28. Panels A-E of the figure illustrate what happens to inputs centered on units 24, 26, 27, 29, and 33, respectively. In each case, the curves shown start out as low bumps centered on the actual input, but as time passes during settling, activation grows overall, and the bump is pulled toward the center of the attractor. The amount of pull is relatively weak for an input centered at 24, away from the center of the attractor; here the overall activation is also fairly weak, because the attractor is only weakly activated, so activity is due almost completely to the bottom up input alone. The influence of the attractor grows stronger for points closer to the center of the attractor (B and C), producing both greater activation and more pull toward the center of the attractor. For an input at the center of the attractor, activation is very strong, and is centered on the input because it falls right at the center of the attractor (panel D). For a point to the right of the center (panel E) the pull is back toward the middle of the attractor. The effect of this is to make inputs falling within the same attractor more similar to each other, and points between the two attractors less similar to each other, modeling the perceptual magnet effect (Kuhl, 1991) and its dependence on language-specific experience. The gradual learning in the network captures the gradual strengthening of the attractor, corresponding to gradual entrenchment as experience accumulates.

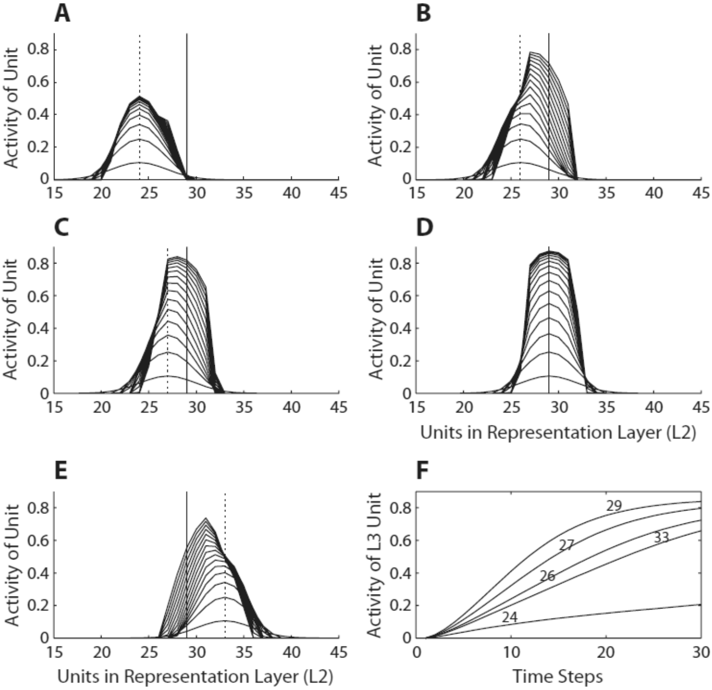

Figure 3. The attractor dynamics of the simplified network after learning. (A) The development of L2 activity for an input centered at location 24 (vertical dotted line). The center of the attractor is at location 29 (vertical solid line). The rising curves indicate L2 activity at successive points in time during settling. (B–E). The L2 activity for inputs at locations 26, 27, 29, and 33. (F) The activity of the winning L3 unit for the same set of inputs. The inset number is the location of the input. From Vallabha & McClelland (2007), Success and failure of new speech category learning in adulthood: Consequences of learned Hebbian attractors in topographic maps. *Cognitive, Affective and Behavioral Neuroscience,* Fig 2, p. 58.

We now return to the full model in Figure 1A and the two-dimensional input space shown in Figure 1B. For our main simulation, we used this network, pretrained, as previously stated, to have attractors for the Japanese /ɾ/ and /w/ phonemes centered on the points labeled Jr and Jw in Figure 1B, according to processes like those operating in the simpler network. The /ɾ/ attractor was made to be very strong because its influence is thought to be very important in perception of the English phonemes /r/ and /l/, found at the onsets of the words *rock* and *lock,* respectively.[1]

Central to our theory, similar to the theories of Kuhl (1991) and Flege (2003), when Japanese adults perceive examples of English /r/ or /l/, the representation of this sound is pulled toward the center for the Japanese /ɾ/ attractor, making the percepts of these two different English sounds very similar to each other. Building on these ideas, we propose that Hebbian learning may reinforce or strengthen this tendency, since both the English /r/ and /l/ inputs activate the same /ɾ/ attractor. Thus, synaptic plasticity in the form of Hebbian learning may still be working in Japanese adults, but it may tend to increase the tendency of the Japanese-trained speakers to continue to hear the English /r/ and /l/ sounds as the same. Our experimental work, described in the next section, was designed to test predictions based on these ideas.

4. Learning to distinguish /r/ and /l/ using natural speech stimuli with and without feedback

We hypothesized that if training could be undertaken using exaggerated tokens of English /r/ and /l/, such that some of these tokens extended outside of the /ɾ/ attractor, the sounds outside the attractor could result in the formation of a new attractor, and this would then facilitate new learning. Because learning is thought to occur by a Hebbian process, no feedback would be necessary, according to this theory, to establish this new attractor. Once this attractor began to be established, we could then gradually make the stimuli more and more similar to each other. Accordingly, in McCandliss et al. (2002), we took natural tokens of English spoken words “rock” and “lock”, and used speech manipulation software to create a continuum of sounds interpolating between and extrapolating beyond the natural speech tokens. The dots between the words Rock and Lock in Figure 1B represent the full set of stimuli in terms of their values on F2 and F3. We did the same for a second continuum using the words “road” and “load” — the dots between these words in the figure represent the full set of these stimuli. Figure 4 presents both of these continua, along with labeling data from a set of native English speakers, showing a fairly sharp transition from stimuli perceived as /l/ at the left end of each continuum to stimuli perceived as /r/ at the other.

---

[1] In reality, the attractors correspond to hyperspheres in a larger multidimensional space. The two dimensional space shown in the Figure can be viewed as a plane drawn through this space, such that it cuts through the center of the Jr hypersphere but just the edge of the Jw hypersphere.

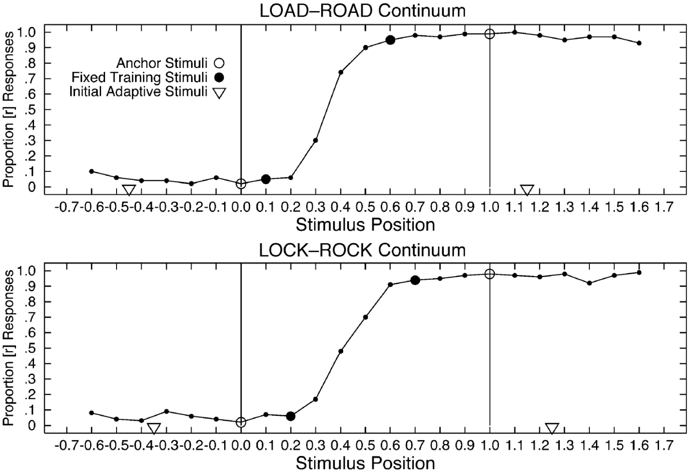

Figure 4. Mean categorization functions of 12 native English speakers for synthesized speech stimuli from each of the two continua used in the experiments of McCandliss et al. (2002). The x-axis represents the position of each stimulus in relation to the anchor stimuli (open circles) for each continuum. Percentages of trials eliciting /r/ responses are plotted on the y-axis for each stimulus. Large empty circles represent the anchor stimuli resynthesized from the recorded base stimuli. Stimuli used for the fixed training condition are indicated with large filled circles. Triangles indicate the positions of the initial stimuli used in the adaptive training condition. From McCandliss, B. D., Fiez, J. A., Protopapas, A., Conway, M., & McClelland, J. L. (2002), Success and failure in teaching the [r]-[l] contrast to Japanese adults: Predictions of a Hebbian model of plasticity and stabilization in spoken language perception. *Cognitive, Affective and Behavioral Neuroscience,* Fig. 1. p. 92.

Native Japanese speakers living in the US who performed poorly (though on average, a bit above chance) discriminating the natural "road-load" and "rock-lock" tokens on a pre-test were assigned to four different training conditions (N = 8 per condition). Within each group, half were trained with stimuli from *road-load* and half were trained with stimuli from *rock-lock.* During training, each participant heard a sound on each trial, and had to press one button to classify it as /r/ or a second button to classify it as /l/.

The four groups differed in the training they received. Two groups received adaptive training: In this procedure, the stimuli used initially were exaggerated examples from the training continuum (see stimuli marked 'Initial Adaptive Stimuli' in the figure). If the participant responded correctly on eight trials in a row, stimuli closer together were selected; after each error, stimuli were moved farther apart unless the stimuli were already at the extremes of the continua. Two other groups received a fixed training regime. In this regime, the stimuli used were easily discriminated by native English speakers but were difficult for Japanese adults (see stimuli marked ‘Fixed Training Stimuli’ in the figure). One of the two adaptive groups and one of the two fixed groups received feedback indicating whether their response was correct on each trial; the other groups received no feedback.

Adaptive training was successful in promoting learning, with or without feedback, as predicted by the model, and captured in the simulations described below. In addition, fixed training with feedback was also successful in promoting learning — indeed, learning occurred most quickly and robustly in this condition. We did not predict this result from the ideas described above, but in the simulations we will describe below, we incorporate a way for feedback to influence learning. Fixed training without feedback resulted in very little learning over three days of training, but there were signs of some learning over three additional training sessions for some participants.

I now describe how Vallabha and McClelland (2007) extended the model described so far to capture the full pattern of results, building on the background training described above to establish attractors for the Japanese /w/ and /ɾ/ categories. Learning within the training experiment context was treated as relying initially on a ‘Fast learning pathway’ (FLP in Figure 1A), which interacts with Layer 2 of the rest of the network as shown. This layer is in principle no different from the SLP, except that it relies on a higher learning rate. The addition of this pathway was principled, based on prior theoretical and experimental research: The idea that the brain contains fast and slow learning systems is central to the complementary learning systems theory my colleagues and I have developed as a general theory of human learning (McClelland, McNaughton, & O’Reilly, 1995). This view is also consistent with a large body of evidence on the effect of brain lesions to the medial temporal lobes, thought to be the locus of the fast learning system in the brain. Damage to the MTL produces a dramatic impact on new learning, but leaves long-standing perceptual, linguistic, and semantic knowledge intact, consistent with the partitioning of our network into SLP and FLP systems. The FLP in the network is viewed as a simplified implementation of the fast learning system, allowing for initial learning in new contexts without directly over-writing existing learned connections built up in the SLP.

Connection weights in the FLP were initialized to capture the approximate average pattern of pre-learning performance of subjects in McCandless et al. Training was simulated by following the protocol of the experiment, allowing inputs to be processed using both the SLP and the FLP. Pre-existing connections within the SLP create perceptual warping effects in the model, as described above; for simplicity, the connections in the SLP are treated as fixed on the time-scale of the training experiment. Activations of the units at the top of the FLP are the basis of responses to the inputs presented in the experiment — one is assumed to correspond to the /r/ response and one to the /l/ response. The response unit with the greater activation tends to be chosen as the network's response, with the consequence that the category boundary corresponds to the point on the stimulus continuum where the activation curve for the R unit exceeds the activation of the L unit. The actual response selected is subject to noise, so that larger differences in response unit activation lead to sharper category boundaries. The initialization described above already provided a slight, weak tendency to activate one of the units in the FLP more strongly for extreme /l/-like stimuli and the other more strongly for extreme /r/-like stimuli, and these tendencies are further changed as a result of training. This is illustrated in Figure 5, which shows how training within the experiment results in learning, affecting activations of the /r/ and /l/ response units in different ways in each of the four conditions of the experiment. The two curves at the bottom of each panel show how the different stimuli initially drove activations in the L and R response units in the FLP.

Learning without feedback is based on the simple Hebbian learning rule previously described, applied to the connections between the L2 units and the units in the FLP, but with a larger learning rate. In the lower right panel of the figure, we can see how learning progresses in the Adaptive-No-Feedback condition. One can see that in this condition, the learning effect is at first somewhat restricted to the extreme stimuli, because these are the stimuli that the network is exposed to during the early phases of adaptive training. As the connections, and therefore the activations, grow stronger, they also spread to adjacent, more similar stimuli. Although response choices themselves are not shown, the network comes to respond in an approximately English-like way, showing a relatively sharp identification boundary. Interestingly, however, the position of the network's boundary between the English /r/ and /l/ categories is shifted because of the strong pre-existing /ɾ/ attractor, which strongly affects processing of the stimuli on the English continuum, pulling those near the middle of the range between the natural English /r/ and /l/ stimuli toward the /l/ end of the continuum, close to the center of the /ɾ/ category as shown in Figure 1. (The position of the category boundary at the end of training can be deduced from the cross-over point of the response activation curves in Figure 5. This point falls at about .6 on the stimulus index scale, whereas the native English boundary, also used as the boundary between the /l/ and /r/ training stimuli shown in Figure 4, lies closer to the /l/ end of the stimulus continuum). The English /r/ stimuli, which fall less close to the center of the /ɾ/ attractor (as shown in Figure 1B) must be fairly extreme to escape assimilation to this category. This effect is seen in the perceptual identification functions exhibited both by the human participants and by the neural network (see Vallabha &

McClelland, 2007, for figures showing these functions).

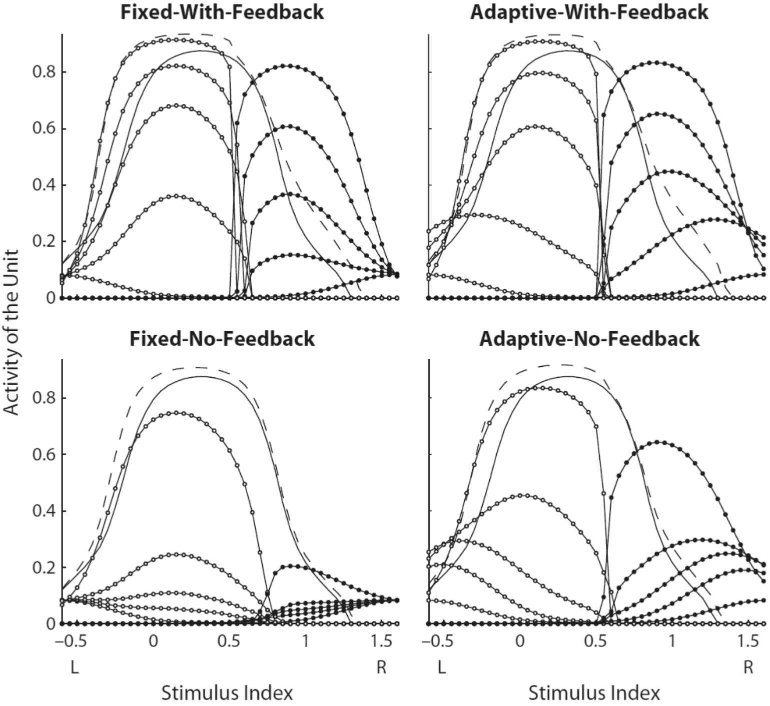

Figure 5. The activity of the FLP units for stimuli on the lock-rock continuum at four different stages of the training, showing the response of the L unit (empty circles), the R unit (filled circles), as well as response of the Japanese tap unit before R/L training (unmarked solid line), and after R/L training (dashed line). The position of the network's category boundary between /r/ and /l/ after training corresponds to the point along the stimulus axis where the tallest open- and filled-circle curves cross. In all cases, this is shifted toward /r/, relative to the actual English category boundary, located at about .3 on the stimulus axis. From Vallabha & McClelland (2007), Success and failure of new speech category learning in adulthood: Consequences of learned Hebbian attractors in topographic maps. *Cognitive, Affective and Behavioral Neuroscience,* Fig. 5, p. 63.

Before we turn to the effect of feedback, it is interesting to note that there is gradual learning even without feedback in the fixed, no-feedback training condition, as shown in the bottom left panel of the figure. Because of the strong influence of the /ɾ/ attractor, the new attractor learned in the FLP for the /r/ category is very weak, and the category boundary is shifted even farther toward the /r/ end of the continuum. However, if the simulation is allowed to run for more training trials, the new attractor gradually grows stronger, exhibiting the increased accuracy exhibited by participants who continued in the fixed, no-feedback condition.

We now turn to a consideration of how feedback might influence learning. We considered a range of possible models, since there are a range of alternative neural network learning rules that can incorporate feedback. The approach that seemed to capture the data best was one in which the feedback simply provided a small additional boost of activation for the unit corresponding to the correct alternative and a small decrement in activation for the unit corresponding to the incorrect alternative. These adjustments in the units' activations were sufficient to allow Hebbian learning to work well in the fixed, no-feedback condition. Learning proceeded very quickly in the simulation, capturing the rapid learning shown by participants in the corresponding condition of the experiment. In the Experiment, feedback helped learning in the adaptive, with feedback condition as well, but learning was quickest in the fixed, with feedback condition. We also observed this in the model; the reason is that in the adaptive condition, training only slowly moves from easy to difficult stimuli, so that participants end up with less exposure to relatively difficult stimuli in this condition.

In summary, the attractor network model implemented by Vallabha & McClelland successfully captures most aspects of the McCandliss et al. (2002) data. The combination of the experimental data and the good fit provided by the model led us to be optimistic, not only that we understood something about second-language learning, but also that we could help Japanese adults acquire a difficult speech discrimination. However, our subsequent work tempered this optimism, as we will now discuss.

5. Learning to distinguish /r/ and /l/ using F3: limited progress

Subsequent to the work above, my collaborators and I became interested in determining whether we could teach Japanese adults to rely on the same perceptual cues that native English speakers use to discriminate /r/ and /l/. Yamada and Tohkura (1990) found that Japanese adults who can discriminate to a degree between the English /r/ and /l/ sounds do so primarily on the basis of cues different from the F3 cue used by native English speakers. Accordingly we constructed synthetic training stimuli like those in Figure 6, and varied the frequency of the F3 formant, extrapolating beyond and interpolating within the range characteristic of native English F3 transitions for /ra/ and /la/ (see figure caption for details). Although native English speakers showed the expected pattern of performance with these stimuli, we were unable to train Japanese adults with these materials. After experimentation we determined that Japanese adults could learn to correctly label the /r/ and /l/ versions of these sounds if all formants other than F3 were eliminated, although once the other formants were removed, the stimuli no longer sound like /r/ and /l/ speech sounds. Starting with these initially discriminable stimuli, Ingvalson et al. (2012) developed an adaptive training procedure starting with 0 amplitude for all formants other than F3, and gradually adapting the amplitude of these formants following the adaptive policy of McCandliss et al. Some participants were trained with feedback and some were trained without feedback. In this study, this variable made little difference.

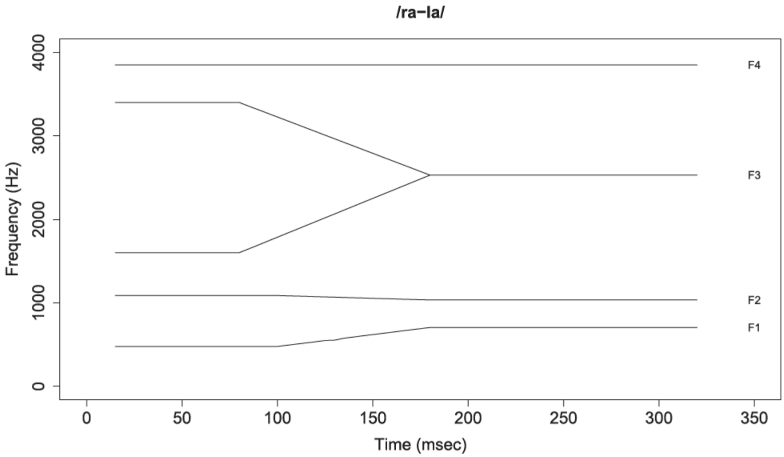

Figure 6. Schematic diagram showing the formant structure of the synthetic /ra/ and /la/ training stimuli used by Ingvalson, Holt, & McClelland (2012). F1, F2, and F4 were the same for all stimuli. Two different F3 transitions are shown. The one that rises from 1600 Hz to 2500 Hz was used as the base formant for /ra/; the one that falls from 3400 to 2500 was used as the base formant for /la/. From Ingvalson, Holt, & McClelland (2012), Predicting native English-like performance by native Japanese speakers. *Journal of Phonetics, 39*, 571−584. Fig. 1, p. 259.

Although all participants improved over the course of eight sessions in the training task, only a few showed improvement in discrimination and labeling on a post test, even though the post test materials were drawn from the same continuum as the training stimuli. This finding may support the view that for many participants, at least, reliance on the F3 cue during training was only achieved by perceiving the stimuli in a non-speech mode. On the positive side, however, a small number of participants did show improvement in discrimination and labeling in a post test, and several of these participants also showed improvement in identifying natural /r/-/l/ stimuli within minimal word pairs across a range of speakers and phonetic contexts. Thus, for a subset of participants, training focused on the F3 contrast did lead to improvement in the discrimination of natural tokens of English /r/ and /l/ stimuli, supporting an important role for reliance on the F3 transition as a cue supporting discrimination of these phonomes.

One limitation of this and most other training studies is their relatively limited duration, and perhaps also the limited range of stimulus materials employed in training. Earlier training studies found that a robust, generalizable gain in both perception and production of English /r/ and /l/ phonemes could be obtained by a training regime that used both a wide range of speakers and contexts for the /r-l/ contrast, including position in word and specific phonological context (Bradlow et al., 1999). However, such training studies still leave native Japanese speakers well short of native English levels of proficiency with the /r-l/ contrast, and other work suggests that the difficulty lies in the failure of these protocols to help native Japanese speakers to learn to rely on F3 (Iverson, Hazan, & Bannister, 2005). One might hope, based on the limited success of our F3 training study, that a more extensive training regime based on manipulation of the F3 cue might lead to a lasting and more extensive gain. Further research should certainly be undertaken to explore this question.

6. Perception and production of /r/ and /l/ from long-term exposure to English: Gradual improvement with no change in reliance on F3

Another approach to understanding plasticity of speech perception and production in adulthood is to explore the effect of long-term, natural exposure to English. Accordingly, Ingvalson, McClelland and Holt (2011) assessed English spoken language perception and production by native Japanese speakers living in the United States, focusing on length of residency and degree of immersion in English language contexts as potential predictors of perception and production of the English /r/ and /l/ phonemes. Our thought in undertaking this research was to explore whether, somehow, long-term immersion in a native English speaking context could lead to improved reliance on F3, even among Japanese native speakers whose immersion began in adulthood. Thus, we focused on individuals who arrived in the US as adults, and we assessed (1) global measures of English spoken comprehension and production, (2) measures of perception and production of natural English /r/ and /l/ phonemes, and (3) measures specifically designed to assess the degree of utilization of the F3 cue, both in perception and production.

The participants were all individuals who first moved to the US after age 18, and individuals attending an international school taught in English before that age were excluded. We recruited groups of participants who had lived in the US for less than 2 years (N=15), 2–5 years (N=15), and more than 10 years (N=25). We used general perception and production tests, tests focused on natural perception and production of /r/ and /l/, and tests assessing use of F3 as a basis for perceiving and producing distinctions between /r/ and /l/, based on measures developed by Yamada and Tohkura (1990) for perception and Lotto, Sato and Diehl (2004) for production. Factors considered included Age of Arrival in the US, Length of Residency, and various measures of degree of immersion in natural English language contexts. We found little influence of age of arrival within the adult range sampled by the study, consistent with the idea that, whatever processes lead to entrenchment with respect to second language learning, their influence has leveled off by about age 20. As expected based on the simple principle that long-term exposure can lead to gradual change, Length of Residency and Immersion measures were associated with improved naturalness and intelligibility both on overall measures of English speech comprehension and production. Measures of discrimination of natural /r-l/ minimal pairs as well as measures of intelligibility of speaker's natural productions of /r/ and /l/ also showed effects of extent of exposure in adulthood.

Critically, there was little evidence of change in reliance on F3 either in production or perception as a function of exposure factors (length of residency and immersion). Although there was considerable variation among individuals in their ability to rely on the F3 contrast both on perception and in production, there was no sign of a correlation between reliance on F3 and either length of residency or immersion, either in perception or in production.

Overall, the results of this study and other studies paint a picture in which gradual change in the mechanisms of perception and production occurs as a function of exposure, even in adulthood, and I believe that the entrenched attractor model could be extended to capture these effects. However, the model in its present form does not explain the finding from our study that these changes do not involve increased reliance on F3. The model could be modefied in various ways that could allow it to capture the absence of reliance on F3, and in this way it would be possible to produce a revised model that could be fit to the experimental data, but such a change would still leave the question: *why* do Japanese adults have such difficulty learning to rely on F3, when they are able to learn to rely on other cues to the /r-l/ contrast?

7. Discussion and future directions

Our work shows that there is plasticity in adult phonological systems, albeit with some limits, and further explorations of training regimes that can maximize this plasticity seems warranted (see Lim & Holt, 2011, for one promising direction this research might take). The work also shows that the entrenched attractor model that we and others have explored, when implemented in a neural network, can provide a good account of the development of perceptual categories in early life, their influence on perception of non-native speech sounds in adulthood, and the effects of four different variations of training on the acquisition of new perceptual categories for difficult non-native contrasts. More generally, our work, together with work from other groups (Logan et al., 1991; Bradlow et al., 1999; Iverson et al., 2005), indicates that native Japanese speakers can benefit from training and exposure to English, both in terms of perception and production of /r/ and /l/ stimuli. However, as pleasing as these results are, they leave us short of a full account of the difficulty native Japanese speakers face in learning English speech categories.

Our limited progress in teaching Japanese adults to rely on the F3 cue in Ingvalson et al. (2012), together with the pattern of individual differences we found in ability to use the F3 cue in Ingvalson et al. (2011), suggests that use of F3 is not impossible for some Japanese adults, even though improvement in the use of F3 appears to be very difficult. Further work is necessary to determine why some native Japanese speakers had less difficulty with the /r-l/ contrast than others. Early exposure to English before arrival in the US may improve native Japanese speakers' mastery of the English /r/-/l/ contrast (Larson-Hall, 2008), and might be responsible for some Japanese adults' ability to rely on F3, but there are several other possibilities, including those based on biological differences as well as others that are more experience-based.

To me, the most interesting approach is to consider whether there is some deeper difference between English and Japanese phonological patterns that create the difficulty Japanese adults have with English /r/ and /l/ stimuli. In this context, it is interesting to note that Japanese speakers are sensitive to the differences in F3 transitions in Japanese consonants /d/ and /g/ (Mann, 1986). In these consonants, the transitions are more abrupt than they are in English /r/ and /l/ stimuli, and it has been obseved that Japanese listeners often hear English syllables like "ra" or "la" as if they were two sylables which might be written in English as "oo-ra". It may be then that the gradual transitions that are characteristic of English are very unlike the more rapid transitions found in Japanese speech sounds, including the /ɾ/ sound. This sound is perceived by native English speakers as a 'd' sound, which has a more rapid transition than English /l/ or /r/. As things stand now, our attractor models do not really capture the dynamic aspects of speech and so do not yet allow us to explore such issues. I do hope that someday we will be able to explore such possibilities.

References

Anderson, J. A., Silverstein, J. W., Ritz, S. A., & Jones, R. S. (1977). Distinctive features, categorical perception, and probability learning: some applications of a neural model. *Psychological Review, 84*(5), 413.

Best, C. T. & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In M. J. Munro & O.-S. Bohn (Eds.), *Second language speech learning: The role of language experience in speech perception and production* (pp. 13–34). Amsterdam: John Benjamins.

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. (1999). Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception and Psychophysics, 61*, 977–985.

Bybee, J. & McClelland, J. L. (2005). Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition. *The Linguistic Review, 22*(2–4), 381–410.

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). Timonium, MD: York Press.

Flege, J. E. (2003). Assessing constraints on second language segmental production and perception. In N. O. Schiller & A. Meyer (Eds.), *Phonetics and phonology in language comprehension and production, differences and similarities* (pp. 319–355). Berlin: Mouton de Gruyter.

Flege, J. E. & Liu, S. (2001). The effect of experience on adults' acquisition of a second language. *Studies in Second Language Acquisition, 23*, 527–552.

Ingvalson, E. M., Holt, L. L., & McClelland, J. L. (2012). Can native Japanese listeners learn to differentiate /r-l/ on the basis of F3 onset frequency? *Bilingualism: Language and Cognition, 15*(2), 255–274.

Ingvalson, E. M., McClelland, J. L., & Holt, L. L. (2011). Predicting native English-like performance by native Japanese speakers. *Journal of Phonetics, 39*, 571–584. doi: 10.1016/j.wocn.2011.03.003.

Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r–l/ to Japanese adults. *Journal of the Acoustical Society of America, 118*, 3267–3278.

Kuhl, P. K. (1991). Human adults and human infants show a 'perceptual magnet effect' for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics, 50*, 93–107.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science, 255*, 606–608.

Larson-Hall, J. (2008). Weighing the benefits of studying a foreign language at a younger starting age in a minimal input situation. *Second Language Research, 24*, 35–63

Lenneberg, E. H. (1967). *Biological foundations of language*. New York: John Wiley & Sons.

Lim, S. J. & Holt, L. L. (2011). Learning foreign sounds in an alien world: Videogame training improves non-native speech categorization. *Cognitive Science, 35*, 1390–1405. PMC3166392.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America, 89*, 874–885.

Lotto, A. J., Sato, M., & Diehl, R. L. (2004). Mapping the task for the second language learner: Case of Japanese acquisition of /r/ and /l/. In J. Slifka, S. Manuel, & M. Matthies (Eds.), *From sound to sense: 50+ years of discoveries in speech communication* (pp. 1–6). Cambridge: MIT Press.

Mann, V. A. (1986). Distinguishing universal and language-dependent levels of speech perception: Evidence from Japanese listeners' perception of English "l" and "r". *Cognition, 24*, 169–196.

McCandliss, B. D., Fiez, J. A., Protopapas, A., Conway, M., & McClelland, J. L. (2002). Success and failure in teaching the [r]-[l] contrast to Japanese adults: Predictions of a Hebbian model of plasticity and stabilization in spoken language perception. *Cognitive, Affective and Behavioral Neuroscience, 2*(2), 89–108.

McClelland, J. L. (in press). Capturing gradience, continuous change, and quasi-regularity in sound, word, phrase, and meaning. In B. MacWhinney & W. O'Grady (Eds.), *The handbook of language emergence* (pp. 53–80). Hoboken, NJ: Wiley-Blackwell.

McClelland, J. L. (2006). How far can you go with Hebbian learning, and when does it lead you astray? In Munakata, Y. & Johnson, M. H. *Processes of change in brain and cognitive development: Attention and performance XXI* (pp. 33–69). Oxford: Oxford University Press.

McClelland, J. L. & Bybee, J. (2007). Gradience of gradience: A reply to Jackendoff. *The Linguistic Review, 24*, 437–455.

McClelland, J. L., Fiez, J. A., & McCandliss, B. D. (2002). Teaching the /r/-/l/ discrimination to Japanese adults: behavioral and neural aspects. *Physiology & Behavior, 77*, 657–662.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review, 102*, 419–457.

McClelland, J., Thomas, A., McCandliss, B., & Fiez, J. (1999). Understanding failures of learning: Hebbian learning, competition for representational space, and some preliminary experimental data. In J. Reggia, E. Ruppin, & D. Glanzman (Eds.), *Progress in brain research. Disorders of brain, behavior and cognition: The neurocomputational perspective* (Vol. 121, pp. 75–80). Amsterdam: Elsevier.

Vallabha, G. K. & McClelland, J. L. (2007). Success and failure of new speech category learning in adulthood: Consequences of learned Hebbian attractors in topographic maps. *Cognitive, Affective and Behavioral Neuroscience, 7*, 53–73.

Yamada, R. A. & Tohkura, Y. (1990). *Perception and production of syllable-initial English /r/ and /l/ by native speakers of Japanese.* Proceedings of the 1990 International Conference on Spoken Language Processing (pp. 757–760). Kobe, Japan.

成人による英語のrとlの区別の学習：行動研究とモデルリング研究

ジェームス・L・マクレランド（スタンフォード大学）

要旨

本稿では、言語音声の知覚に経験がどのように影響するかを調査した一連の研究について紹介する。これらの研究は、言語は絶対的というよりも段階的な制約によって影響を受け、また経験が、母語・外国語を問わず、自分が聞く言語音声に反応する方法を徐々に作り上げていくという筆者の全般的な理論的視点にもとづいて行われてきた。日本人の成人によるrとlの区別の学習という事例に焦点を置き、まず幼児期に母語に触れることにより経験がどのように知覚を形成していくかをとらえたニューラル・ネットワーク・シミュレーションのモデルを紹介する。次に、自然な日本語の音声カテゴリーを習得したあとで英語のrとlの区別を学習する上で、様々な練習法にどのような効果があるかについて、そのモデルに基づく予測を論じる。さらに、モデルの予測を部分的には支持するものの、日本人の成人がrとlの区別を学習する上で、モデルからは充分に予測できないような限界があることを示すいくつかの実験結果を提示する。全体として、これらの研究は、成人した後でも知覚の変化は可能であるが、今後の研究で明らかにしていくべきいくつかの限界があることを示唆している。