

Learning Phonemes Without Minimal Pairs

Jessica Maye and LouAnn Gerken
University of Arizona

1. Introduction

The question addressed by this research is how humans acquire the internalized, mental categories that reflect the phonemes¹ of their language. That humans have such categories is evidenced by the fact that, for example, English speakers immediately recognize *pear* and *bear* as being different words. This immediate realization is independent of the fact that these two words have different *meanings*, since English speakers will also report that nonsense words like *bove* and *pove* are different words – despite the fact that neither word means anything.

Phonemic categories, once acquired, have an influence on the perception of speech sounds. Although adults can easily perceive acoustic distinctions that differentiate the categories of their own language, they often have difficulty perceiving non-native phoneme contrasts. For example, the English distinction between /r/ and /l/ is difficult for native Japanese speakers to perceive (Miyawaki et al. 1975), and the Hindi distinction between /t/ and /ʈ/ (the latter sound is retroflex, pronounced with the tongue curled backwards) is difficult for native English speakers to perceive (Werker et al. 1981).

Infants are born with the ability to categorize speech sounds (Eimas et al. 1971), but these categories are initially universal, not based on the phonemes of a particular language. For instance, Japanese-learning infants can initially discriminate English /r/ and /l/ (Tsushima et al. 1994), and English-learning infants can initially discriminate Hindi /t/ and /ʈ/ (Werker et al. 1981). Over the course of the first year of life, though, the infant gains experience with the native language, and begins to only discriminate contrasts that represent a phonemic distinction in the native language (Werker & Tees 1984). Through our research, we would like to understand how these phoneme-related effects on speech perception arise in the language learner.

2.0 The Acquisition of Phoneme Categories

Two types of hypotheses have been proposed to explain how these phoneme-based categories are acquired, which can be roughly characterized as *minimal pair-based learning* versus *distribution-based learning*. The minimal pair-based hypothesis is that infants begin to attend to native language phonemes when they learn that a phonetic distinction can differentiate the

meanings of two words. For example, if an infant learns that the words *pear* and *bear* consistently refer to different items, they will begin to attend to the /p~/b/ voicing distinction, since it can differentiate between two meanings. Under this hypothesis, it is crucial for an infant to actually know the meanings of words that form minimal pairs, in order to acquire phoneme categories.

The distribution-based hypothesis is that infant speech perception is shaped by the native language sounds *directly*, on the basis of the distribution of phonetic exemplars. For example, in a language in which /p/ and /b/ are different phonemes, there will be considerable phonetic variation in the actual exemplars of the two phonemes, with some overlap between the two categories. However, exemplars of a particular phoneme will presumably cluster together along one or more acoustic dimensions, and these clusters can be used to differentiate the categories which are used contrastively in a language. Under this hypothesis, speech perception abilities are automatically shaped by the statistical distribution of exemplars that an infant is exposed to, and word-learning is not a necessary prerequisite.

2.1 The Minimal Pair Hypothesis

In response to the early studies on infant speech perception, MacKain (1982) argued that in order to be shaped by the ambient language, an infant must experience members of different phonetic categories as representing a contrast. That is, the infant must be aware that they are experiencing sounds that are contrastive in the language, before their perceptual system will be influenced by the contrast in question. This argument amounts to a requirement that infants know whether a given phonetic distinction can result in a meaning distinction in their language; in essence, it proposes minimal pair-based learning.

A similar suggestion was made by Werker & Tees (1984), when they found that the shift from language-general to language-specific speech perception occurs during the latter half of the first year. Since it is also during the latter half of the first year that infants begin to understand and produce words, these researchers suggested that the shift might occur because of the infant's developing lexicon.

Despite the appeal of the minimal pair-based hypothesis, there is emerging evidence that it cannot be true. One piece of evidence comes from the timeline of word learning. If the reorganization of speech perception is *based* on the knowledge of minimally contrastive words, no reorganization should be evident before a child knows any minimal pairs. Since the perceptual reorganization has been shown to begin between 8-10 months for consonants (Werker & Tees 1984), and as early as 6 months for vowels (Kuhl 1991, Polka & Werker 1994), the minimal pair-based hypothesis predicts that infants have learned minimal pairs that contrast vowels by 6 months, and pairs that contrast consonants by 8-10 months. However, infants have not been shown to possess a large receptive vocabulary before the age of 12 months, and what words they do know may not include minimal pairs. Caselli et al. (1995) performed an extensive cross-

linguistic study on the content of the expressive and receptive lexicons of English-learning and Italian-learning infants between 8 and 16 months of age. According to their findings, the average 8-month-old can understand around 36 words. However, the 50 words most likely to be in the early receptive lexicon of English-learning babies do not include a single minimal pair. For Italian-learning babies, they found a single minimal triplet (*nonno* “grandpa”, *nonna* “grandma”, *nanna* “sleep/bedtime”), which differed minimally with respect to vowels, not consonants.

A second argument against the minimal pair-based hypothesis comes from recent experimental findings on word-learning and its interaction with speech perception. Stager & Werker (1997) conducted a study to test infants’ ability to discriminate minimal pairs of nonsense words (e.g. *bih* and *dih*). They found that it is not until 18 months of age that infants can discriminate minimal pairs, when those words have semantic referents. That is, when infants are beginning to learn word meanings, it interferes with the ability to discriminate fine phonetic detail. Only older infants, who are more adept at word-learning, are able to discriminate minimal pairs of words.

2.2 The Distribution-Based Hypothesis

In support of a distribution-based account of phoneme learning, previous research has demonstrated that infants utilize distributional information in other areas of language learning. One such study was conducted by Jusczyk, Luce, and Charles-Luce (1994), who showed that infants are aware of the relative frequency with which various phonotactic patterns occur in their language. They presented English-learning 9-month-old infants with nonsense words having either common English phonotactics (e.g. *mubb*), or legal but less frequent patterns (e.g. *jurth*). The infants preferred to listen to the words with frequently occurring phonotactics, indicating that they recognized the difference between the frequent and infrequent patterns.

Another study showing infants’ use of distributional information in language learning was conducted by Saffran, Aslin, and Newport (1996), who showed that 8-month-old infants can utilize statistical information about syllable co-occurrence, in order to segment a continuous stream of speech into words. This study presented infants with continuous synthetic speech composed of 3-syllable nonsense words (e.g. *tupiro*, *golabu*, *bidaku*, *padoti*), pronounced without stress or any information about word boundaries. The researchers hypothesized that infants might be able to learn which syllables went together to form words, on the basis of how often syllables occurred next to each other. Syllables within the same word would often occur next to each other, in a particular order (e.g. *golabu#padoti*), while syllables from different words would only occur next to each other if their words happened to be adjacent (e.g. *tupiro#bidaku*). After listening to this continuous speech for 2 minutes, infants showed a preference for the experimental nonsense words presented in isolation, over non-words (syllables presented in random order: e.g. *tidoku*) and part-

words (sequences of syllables that had occurred during training, but crossed a word boundary: e.g. padoti#bidaku).

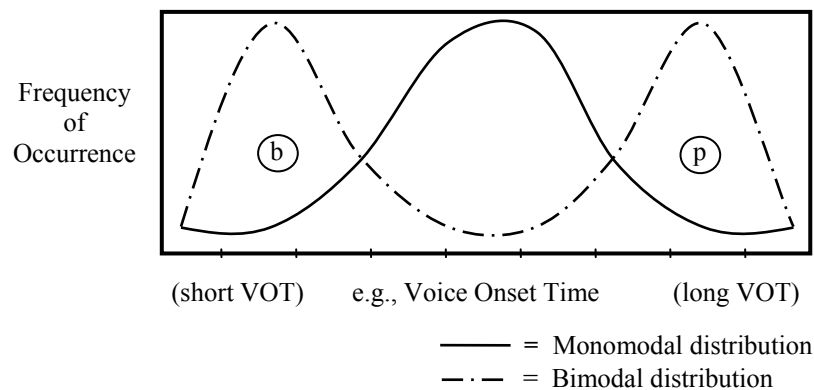
The findings of the above studies indicate that infants have access to detailed information about the relative frequency with which elements of their language occur. This type of information could potentially also be used for learning a language's phoneme categories. Guenther & Gjaja (1996) proposed a computational model of speech perception, which demonstrates how a neural correlate of phoneme categories might be acquired, on the basis of the statistical distribution of exemplars in a language. In their model, exposure to a particular language leads to nonuniformities in the distribution of the firing preferences of neural cells in the auditory system. Over time, proportionately more neurons become devoted to firing in response to the most frequently heard sounds.² This results in clusters of neurons that reflect the phoneme categories of the language. These clusters, in turn, give rise to the phoneme-based effects on speech perception.

3.0 Experiment

Guenther & Gjaja's model elegantly exemplifies the distribution-based hypothesis of phoneme acquisition, and argues that such a model is neurally plausible. However, what remains to be shown is that humans are actually capable of forming categories on the basis of the distribution of phonetic exemplars. Our experiment was designed to test this hypothesis by presenting adult participants with phonetic exemplars from an artificial language, and giving them no information about word meaning.

We designed the experiment according to the following reasoning. In real speech, speakers produce sounds with a large degree of phonetic variation. When a language has two, contrastive phoneme categories (like English /b/ and

Figure 1: Monomodal vs. Bimodal Distributions



/p/), although the categories are pronounced with much variation and even overlap, the most frequently heard tokens will fall into two clusters, forming a bimodal distribution, as shown by the broken line in Figure 1. When a language has a single phoneme category along some acoustic dimension (e.g. voice onset time), its tokens will fall into a single cluster, forming a monomodal distribution, as shown by the solid line in Figure 1.

In our experiment, we presented two groups of participants with the *same* stimuli, but varied the frequency with which they heard each token, such that one group was exposed to a monomodal distribution, and the other group a bimodal distribution. We then tested whether each group treated the stimuli as corresponding to one or two phoneme categories.

3.1 Stimuli

Because our intent was to test adult English speakers, we needed to find a contrast that English speakers are actually able to perceive, but one that they do not discriminate as a phonemic contrast. The reason for this is that we wanted to bias the two groups of participants in different directions: we wanted the bimodal group to attend to the acoustic differences, and treat them as phonemic, while the monomodal group should ignore the acoustic differences, treating them as non-phonemic phonetic variation. A study by Pegg & Werker (1997) provided us with an appropriate contrast: between English voiced /d/ (as in *day*) and voiceless unaspirated /t/ (as in *stay*). Although both of these sounds come from English, they do not constitute a phonemic contrast in English, since they never occur in the same environment (i.e. unaspirated /t/ only occurs after /s/ in English, while voiced /d/ never does). English speakers perceive both of these sounds as members of the /d/ category, when occurring in syllable initial position. For example, if the /s/ is removed from the word *stay*, English speakers will perceive it as the word *day*. However, Pegg & Werker showed that in a discrimination task, English speakers can hear the acoustic differences between voiced /d/ and unaspirated /t/.

It is also important to point out that, although the distinction between voiced /d/ and voiceless unaspirated /t/ is phonemic in many languages (e.g. Spanish, French, Japanese), the particular sounds used in this experiment (and in Pegg & Werker 1997) are taken from *English* syllables beginning with /d/ and /st/. The distinction between /d/ and /t/ is generally characterized as a “voicing” distinction; however, in English, actual tokens of /d/ and /st/ do not differ in terms of voice onset time. Instead, in our own measurements (cf. measurements reported by Pegg & Werker 1997), the differences between English /d/ and /t/ are in the formant onset frequencies for the following vowel, with /d/ having a more extreme transition from formant frequencies at the onset of the following vowel to the center of the vowel. Also, although prevoicing is not reliably produced in English, it was included in /d/ tokens, in order to ensure that the two sounds were different enough to be distinguished by participants. For our experiment, the cues that distinguish these two sounds are of little consequence,

so long as the stimuli satisfy two criteria: that they be not readily distinguished, yet still be discriminable to native English speakers.

To create our stimuli, we began with natural English productions of the syllables /da/, /sta/, /dæ/, /stæ/, /dr/, and /str/. We then removed the /s/ from the /s/-initial syllables, resulting in the syllables /da/, /ta/, /dæ/, /tæ/, /dr/, and /tr/. These six syllables were then re-synthesized into three continua, running from /d/ to /t/ (in each of the three vowel contexts) in eight equal steps.

Filler stimuli were syllables beginning with the consonants /m/ and /l/ (/ma/, /mæ/, /mr/, /la/, /læ/, and /lr/). There were four tokens (different utterances of each filler syllable), each presented twice per block of training, for a total of 24 filler stimuli per block.

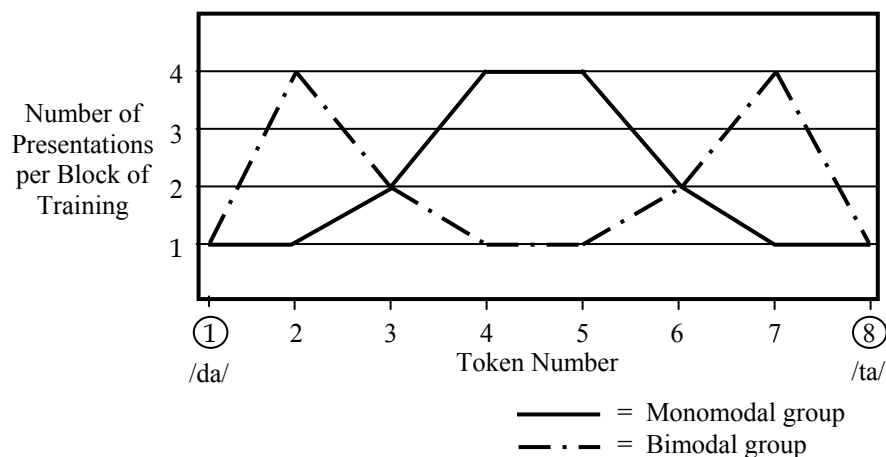
During the test phase, participants were presented with only the endpoint /d/ and /t/ stimuli. These stimuli were paired with themselves (e.g. /da/~da/) on “same” trials, paired with each other (e.g. /da/~ta/) on experimental “different” trials, or paired with filler items (e.g. /da/~ma/) on filler “different” trials. Filler pairs consisted of pairs of identical filler stimuli (e.g. the same utterance of /ma/ repeated twice), pairs of nonidentical filler stimuli (e.g. two different utterances of /ma/), and pairs of different filler stimuli (e.g. /ma/~la/).

3.2 Participants

Participants were 32 native English speakers (23 female, 9 male) enrolled in courses at the University of Arizona who received course credit for their participation, or who volunteered to participate. Their ages ranged from 18 to 41 years, with a mean age of 23. All had normal hearing and no language impairment.

Participants were randomly assigned to one of two groups for training. The two groups differed with respect to the distributional frequency of stimuli presented during training. One group was presented with a monomodal distribution of each continuum, in which the tokens from the center of each continuum were presented four times as often as the endpoint tokens, as shown by the solid line in Figure 2. The other group was presented with a bimodal distribution, in which the tokens from the near the endpoints of each continuum were presented four times as often as the center tokens, as shown by the broken line in Figure 2. In this way, both groups of participants were presented with all tokens along each continuum, the only difference being the frequency of stimulus presentation. Both groups of participants heard 16 experimental stimuli in each of the three vowel contexts, for a total of 48 experimental stimuli per block of training. Also, both groups heard tokens 1 and 8 from each continuum only once per block. This was because tokens 1 and 8 were used as the contrasting stimuli during the test phase (see *Procedure* section), and we wanted to ensure that both groups had heard the test stimuli the same number of times.

Figure 2: Stimuli Presentation Frequency during Acquisition Phase



3.3 Procedure

Participants were informed that in this experiment they would be listening to a language they had never heard before, with the purpose of learning about the sounds of the language. After listening to words from the language, they would be given a task in which they would hear pairs of similar-sounding words, and would have to decide whether each pair was the same word repeated twice, or two different words in this language.

Practice Task: Participants were first given a practice task, during which they heard 10 pairs of English words, half of which were “same” pairs (two utterances of the same word), and half of which were “different” pairs (two English words differing only in a single consonant). Participants were instructed to mark either “same” or “different” on a response sheet to indicate their answers. The inter-stimulus interval within each pair was 500 msec, and between trials there was a 2 second pause, during which participants recorded their responses.

Acquisition Phase: During the acquisition phase, participants were presented with words from the artificial language presented in list form, with an inter-stimulus interval of 1 second. No information was provided regarding the meaning of these words in the language. The words included all of the stimuli from the three experimental continua (*da*, *ta*, *dae*, *tae*, *dr*, and *tr*) as well as all of the filler stimuli (*ma*, *mae*, *mr*, *la*, *lae*, and *lr*), presented with the frequency distributions appropriate for each participant group, as discussed in the *Stimuli* section, above. The entire block of experimental and filler stimuli was repeated four times, for a total listening time of 9 minutes. In order to help the participants maintain their attention to the stimuli, they were given a check-sheet with 384 empty boxes on it (one for every stimulus presented during

acquisition), and instructed to check a box every time they heard a word. Participants were told that their task during this phase of the experiment was simply to listen carefully to the words of this language, and the way that the words sounded; and that the purpose of the check-sheet was simply to help them pay attention to the words.

Test Phase: After completing the acquisition phase, participants were given a decision task, in which they were presented with pairs of stimuli and asked to indicate on a response sheet whether the two stimuli in each pair were the same word repeated twice, or two different words in this language. Participants were reminded that this task was the same one they initially performed using English words. Participants were instructed to listen carefully to each pair since the items in each pair would sound very similar to each other; and that if they were unsure of their response, they should make their best guess and then be prepared to listen to the next pair.

During the test phase, the inter-stimulus interval was 500 msec, with 2 seconds between each pair. The items of interest were the pairs of words like /da/~ta/, in which one word began with /d/ (the /d/ endpoint of one of the continua), and the other word began with /t/ (the /t/ endpoint of the same continuum). We predicted that the two participant groups should differ on their responses to these pairs, and these pairs only. According to our hypothesis, the participant group that had been trained on the monomodal distribution should believe that in this language there is only one phoneme represented by the /d~/t/ stimuli; therefore, these participants were predicted to respond “same” to these items. The participant group trained on a bimodal distribution, however, should believe that these stimuli represent two phonemes in this language; therefore, these participants were predicted to respond “different” to these items.

3.4 Results

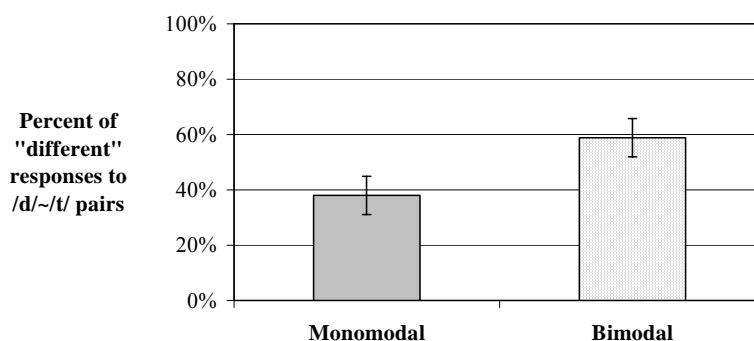
To calculate each participant’s performance on the test, responses were scored as either “correct” or “incorrect.” For the filler pairs, pairs of words that were either pairs of identical stimuli, or pairs that were two utterances of the same filler word (e.g. /ma~/ma/) were scored as correct if the subject responded “SAME”. Filler pairs that consisted of two different filler words (e.g. /ma~/la/) were scored as correct if the participant responded “DIFFERENT.” And for the experimental pairs (e.g. /da~/ta/), responses were scored as correct if the participant responded “DIFFERENT.” Because of this last scoring choice, participants from the bimodal training group, who were expected to distinguish the experimental contrast more often, were expected to receive higher scores on the test.

We performed a 2 Training Group x 5 Test Pair Type anova on the number-correct data. There was a significant effect of Group ($F = 5.35, p < .05$), with participants from the monomodal training group scoring lower than those from the bimodal group. There was also a significant effect of Test Type ($F = 97.58, p < .0001$), with scores on the /d~/t/ contrast pairs significantly lower than scores

for the four types of filler pairs, which were at ceiling. Importantly, there was also a significant interaction between Group and Test Type ($F = 7.65$, $p < .0001$). Follow-up comparison revealed the only significant effect of Group was for the /d~/t/ pairs, with participants from the monomodal training group scoring lower on the /d~/t/ contrast pairs than did participants from the bimodal training group ($t = 2.19$, $p < .05$).

The results for /d~/t/ pairs, the experimental contrast, are illustrated in Figure 3. Participants from the bimodal training group were more likely to respond “DIFFERENT” to the experimental /d~/t/ pairs, indicating that these sounds represent a phonemic contrast in this language. This finding confirms our hypothesis that humans can utilize distributional information in order to form phoneme categories. The group that was trained on a two-cluster distribution was more likely to indicate that the stimuli corresponded to two categories, than was the group trained on a one-cluster distribution.

Figure 3: Results, Experimental Contrast Pairs



4. Conclusion

The results of this experiment support a distribution-based model of phoneme learning. These results cannot be interpreted as arising from factors of the participants' native language, since participants from both groups were native English speakers. The only difference between the two groups of participants was the statistical distribution of sounds they were exposed to during training. What is especially interesting, is that participants learned to discriminate minimal pairs in this language, although they were not *trained* on minimal pairs, and were not given any information about the meanings of words in the language.

The findings of this experiment highlight the role of particular exemplars, and their frequency of occurrence, for learning phoneme categories. These findings suggest that humans maintain some sort of mental histogram for acoustic patterns they encounter in their language. We remain agnostic as to

how speech sounds are represented neurally, as this is an issue of much contention in the field of speech perception; however, Guenther & Gjaja (1996) and Guenther et al. (1999) present algorithms for how speech sounds could result in histogram-like organization of the auditory cortex.

There are several directions in which this research can be extended. First, we would like to know whether the categories formed are specific to training items, or whether they are generalizable to new tokens, for example, spoken by a new voice. In this experiment, participants were tested on their categorization of stimuli that they had heard during training (specifically, the endpoint /d/ and /t/ tokens). In reality, though, listeners rarely encounter the same exemplars more than once. To test generalization, we plan to test participants' categorization of *new* tokens not presented during training.

Another direction for extending this research is to test the language-specificity of category learning. Do learners have expectations about speech sound categories that are particular to language? Or is this type of category formation performed by a more general learning mechanism? To test this, we plan to test participants on their categorization of a new *contrast* that was not presented during training (e.g. /g/~k/). Since languages tend to have multiple analogous contrasts, such that a language with a /d~/t/ contrast might also have a /g~/k/ contrast, learners might expect to encounter multiple, analogous contrasts.

In addition, this type of generalization will enable us to test whether learning is constrained by linguistic markedness. Linguistic markedness refers to cross-linguistic regularities in linguistic inventories. For example, all languages have coronal sounds (pronounced with the tip of the tongue, like /d/), but not all languages have dorsal sounds (pronounced with the body of the tongue, like /g/). The commonly-occurring sound (/d/) is said to be “unmarked,” while the relatively less common sound (/g/) is said to be “marked.” It will always be the case that a language with the sound /g/ also has the sound /d/, but the reverse is not always true (having /d/ does not imply the presence of /g/). If the learning mechanism for phoneme categories is constrained by markedness implications, a learner who has heard the sound /g/ might assume the sound /d/ will also occur in the language; while a learner who has heard the sound /d/ will not have evidence for the presence of /g/. If markedness implications have this effect on phoneme learning, it would be evidence that the learning mechanism is constrained by factors that are specific to language.

And finally, because our hypothesis initially arose from findings regarding *infant* language development, an important extension of this research is to test whether infants are also capable of utilizing distributional information for the purposes of learning phoneme categories. The previous research showing infants' use of statistical information in other areas of language learning, suggests that infants will utilize this type of information for phoneme learning as well. Also, since young infants do not have well-established native language phoneme categories, they may learn new categories even better than adults do,

since pre-existing categories presumably inhibit the formation of new, competing categories.

In conclusion, we have shown that language learners can acquire phoneme categories on the basis of the distribution of sounds in the language that they hear. This provides an explanation for how infants are able to learn the phonemes of their language by the early age of eight months, before they know the meanings of many words. This finding contributes to the growing body of research showing that adults and infants make use of statistical information in many areas of language processing.

Endnotes

1. The categories we would like to account for are not “phonemes,” as the term is used by linguists. For linguists, the term applies to categories that include multiple allophones, which appear in different phonological positions. What we are interested in is perhaps more appropriately termed “phonetic categories” or “phonetic equivalence classes;” specifically, sounds which are categorized together in a *particular* phonological position. However, we point out that the psychological reality of the linguist’s “phoneme,” comprising multiple allophones, has not been experimentally demonstrated. The categories we account for here could plausibly be the *only* psychological correlate to phoneme categories.
2. Guenther et al. (1999) have since revised this model. In their new model, exposure to exemplars actually leads to a *reduced* number of neurons dedicated to the most prototypical sounds, if these sounds are encountered in a context that favors categorization.

References

- Caselli, Maria Cristina, Elizabeth Bates, Paola Casadio, Judi Fenson, Larry Fenson, Lisa Sanderl, & Judy Weir (1995) A cross-linguistic study of early lexical development. *Cognitive Development*, *10*, 159-199.
- Eimas, Peter D., Einar R. Siqueland, Peter W. Jusczyk, & James Vigorito (1971) Speech perception in infants. *Science*, *171*, 303-306.
- Guenther, Frank H., & Marin N. Gjaja (1996) The perceptual magnet effect as an emergent property of neural map formation. *Journal of the Acoustical Society of America*, *100* (2), 1111-1120.
- Guenther, Frank H., Fatima T. Husain, Michael A. Cohen, & Barbara G. Shinn-Cunningham (1999) Effects of categorization and discrimination training on auditory perceptual space. *Journal of the Acoustical Society of America*, *106*, 2900-2912.

- Jusczyk, Peter W., Paul Luce, & Jan Charles-Luce (1994) Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language*, 33, 630-645.
- Kuhl, Patricia 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.
- MacKain, Kristine S. (1982) Assessing the role of experience on infants' speech discrimination. *Journal of Child Language*, 9, 527-542.
- Miyawaki, Kuniko, Winifred Strange, Robert Verbrugge, Alvin M. Liberman, James J. Jenkins, & Osamu Fujimura (1975) An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception & Psychophysics*, 18, 331-40.
- Pegg, Judith E., & Janet F. Werker (1997) Adult and infant perception of two English phones. *Journal of the Acoustical Society of America*, 102 (6), 3742-3753.
- Polka, Linda, & Janet F. Werker (1994) Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, 20, 421-435.
- Saffran, Jenny R., Richard N. Aslin, & Elissa L. Newport (1996) Statistical learning by 8-month-old infants. *Science*, 274, 1926-1928.
- Stager, Christine L., & Janet F. Werker (1997) Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*, 388, 381-382.
- Tsushima, T., O. Takizawa, M. Sasaki, S. Siraki, K. Nishi, M. Kohno, P. Menyuk, & C. Best (1994) Discrimination of English /r-l/ and /w-y/ by Japanese infants at 6-12 months: Language specific developmental changes in speech perception abilities. Paper presented at International Conference on Spoken Language Processing, 4. Yokohama, Japan.
- Werker, Janet F., John. H. Gilbert, Keith Humphrey, Richard C. Tees (1981) Developmental aspects of cross-language speech perception. *Child Development*, 52, 349-355.
- Werker, Janet F., & Richard C. Tees (1984) Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, 7, 49-63.