The effect of variation on phonetic category learning

Madelaine Krehm[1], Adam Buchwald, and Athena Vouloumanos
New York University

Introduction

To understand speech, we must organize its natural variation into meaningful perceptual units that are appropriate to our language. Our perceptual systems divide the continuously varying speech signal into language-specific phonetic categories. For example, one linguistic parameter that varies continuously within and between languages is voicing, the vibration of the vocal cords. The timing of voicing onset relative to the release of the consonant is referred to as voice onset time (VOT). Different languages divide the voicing continuum differently, both in terms of the number of phonetic categories they specify within a continuum, and in terms of the specific VOTs that mark phoneme boundaries. For example, English has two alveolar stops /d/ and /t/ with mean voice onset times of 5 ms and 70 ms, respectively. In contrast, Thai divides alveolar stops into 3 categories /d/ (with mean VOT of -78 ms), /t/ (9 ms) and /t^h/ (65 ms; Lisker & Abramson, 1964). We perceive a continuous feature like VOT as specifying distinct, language-specific categories through categorical perception: Given the same magnitude of acoustic change, adult native speakers perceive a greater difference across the boundary between phonetic categories than within a phonetic category (e.g., Goldstone & Hendrickson, 2010; Liberman, Harris, Hoffman, & Griffith, 1957). Infants and adults learning English or Thai are faced with the task of learning how many alveolar stop categories their language has (two vs. three) and the specific VOT boundaries that distinguish these categories. In this paper we examine how adults learn the relevant phonetic categories for different languages.

Learners face yet another challenge when learning phonetic categories. Although the speech signal varies on dozens of different dimensions, only some, such as voicing, are crucial to understanding a given language, while others are irrelevant to that language. For example, change in VOT could change "ten" into "den", altering the meaning of the word, but a change in pitch does not affect word meaning in English. While pitch contours are irrelevant to word meaning in English, they are important for distinguishing tokens with different meanings in other languages. In Mandarin /ma/ with a flat pitch contour means *mother*, but /ma/ with a falling pitch contour means *scold*. Since languages differ in which dimensions of variation are linguistically relevant, listeners must learn which dimensions matter for the specific language they are learning.

One commonly proposed model for general category learning, the exemplar model, can be applied to the acquisition of phonetic categories in speech (Pierrehumbert,

[1] Reprint requests should be addressed to Madelaine Krehm, Department of Psychology, New York University, 6 Washington Place, New York, NY, 10003, USA, e-mail: madelaine.krehm@nyu.edu.

2001). In the exemplar model, a category is described by a cluster of instances of remembered tokens. The tokens are organized by similarity on any perceptually salient dimension such that similar tokens are closer together in parameter space. When the learner hears a new token, he or she categorizes it by measuring its similarity to the existing token clusters. This account of phonetic category learning builds up phonetic categories directly from the environmental input, creating detailed phonetic representations. Another feature of the exemplar model is that it does not require explicit feedback for phonetic category acquisition. In natural speech production, speakers are more likely to produce tokens near the center of the category boundary, and less likely to produce tokens near the edges of the continuum, creating frequency modes at the center of each category. Consistent with the exemplar model, learners use these frequencies of tokens in the input to infer likely categories (Pierrehumbert, 2003).

Frequency distributions also provide an explanation for how the exemplar model addresses the challenge that variability on irrelevant (as well as relevant) dimensions poses to phonetic category learning. The exemplar model includes the complete percept in the representation such that a token is represented as not only a /t/ but a /t/ with, for example, a specific pitch and in a female voice. Features like pitch and female voice are encoded at the same time as the VOT value, but English learners form indexical categories based on these features, and not linguistic categories (Pierrehumbert, 2001). The tracking of non-linguistic dimensions separately from the encoding of relevant linguistic properties in the exemplar model predicts that learners in an implicit task would be able to learn a new phonetic contrast using its distribution in the input despite non-linguistic variability that is irrelevant to the task.

The exemplar model's predictions have been tested empirically in adults (Gilkerson, 2003; Maye, 2000; Maye, Weiss, & Aslin, 2008). These studies evaluated whether adults could learn a new phonetic contrast in an implicit learning task from different distributions of the frequencies in the input. The participants were exposed to one of two training regimes. One group listened to a unimodal frequency distribution in which the central token was the most frequent, which approximates a single phonemic category. This distribution should reduce listeners' discrimination of the endpoints of the continuum because they should have induced a single category. The other group of participants listened to a bimodal distribution, in which frequency distributions were clustered around two tokens as if there were two distinct categories. Bimodal listeners should have an enhanced discrimination of the endpoints of the continuum because the bimodal distribution would have induced two different categories. In Maye (2000) the training phases consisted of syllables from along the VOT stop continuum with an alveolar place of articulation (/t/-/d/). There were three continua of eight tokens evenly spaced by VOT, ranging from /tɑ/ to /dɑ/, /tæ/ to /dæ/, and /tɚ/ to /dɚ/. Participants in the unimodal condition heard the most tokens in the middle of the VOT continuum, at the border between /t/ and /d/, while participants in the bimodal condition heard more tokens near the ends of the VOT continuum. The test phase evaluated participants' discrimination of endpoints of the continuum. Participants in the bimodal condition were more likely to say that the endpoints were different than adults in the unimodal condition in a same/different task (Maye, 2000). The frequency

distributions that listeners were exposed to in a controlled experimental setting thus affected their phonetic discrimination, consistent with the exemplar model.

Although adults' learning from different frequency distributions provides evidence for the exemplar model, phonetic category learning in the natural language environment requires learners to process many different dimensions. In the natural environment, language learners must ignore irrelevant non-linguistic information in order to consistently perceive phonemes produced by different speakers in different environments. But existing studies have tested learning of stimuli that varied only in the relevant dimension (Maye, Werker, & Gerken, 2002; Maye, 2000). In a natural environment, however, learners are simultaneously exposed to variation that is not relevant to the formation of phonetic categories including variation in pitch, speaker rate, and timbre. While previous studies show that listeners can learn where category boundaries are when presented with systematic variation on the relevant dimension, they do not address whether listeners can learn systematic variation in a relevant dimension in the presence of irrelevant variation. For the exemplar model to explain how we learn phonetic categories, people must be sensitive to the frequency distributions of linguistically relevant variation in the face of irrelevant acoustic variation.

Not only would the exemplar model predict that learners should accommodate irrelevant variation when learning a new category, but there is some evidence that added variation during a learning phase might improve category learning. Additional variation might encourage listeners to learn the distinction based on the relevant dimension rather than on characteristics specific to the training stimuli, improving discrimination after training. For example, in a task with explicit feedback, a more variable set of stimuli prevented listeners from learning to discriminate based on idiosyncrasies of individual tokens, such that adult learners could generalize discrimination to novel tokens varying on the same feature (Logan, Lively, & Pisoni, 1991). Fourteen-month-olds usually fail on an associative word-learning task when the words differ only on one phoneme (Stager & Werker, 1997) but succeed when the learning phase includes words that are recorded by different speakers (Rost & McMurray, 2009). This is consistent with the bottom-up aspect of the exemplar model. Since representations are created using only the tokens in the input, if the input is too narrow, learners might develop an overly specific representation of the category. Additional variation would improve the generalizability of the new category.

We test the exemplar model's prediction that learning to discriminate a new phonetic contrast will be robust in the face of non-linguistic variation. We use the same paradigm as Maye (2000) to test whether learners can discriminate a new phonetic contrast from a systematic frequency distribution of tokens in the input despite irrelevant variation. Learners heard either a bimodal or a unimodal frequency distribution of tokens along the t/d continuum, but with the frequency minimum in the bimodal distribution at a VOT that should not be a phonetic boundary for English speakers (but is a boundary for Spanish, French, and other languages). Unlike in the Maye (2000) experiment, the tokens varied not only on VOT, but on pitch as well. Listeners heard a high and a low pitch version of each token, introducing pitch variation irrelevant to the phonetic contrast. If learners in the bimodal condition were able to

distinguish the endpoints of the continuum better than learners in the unimodal condition, this would suggest that learners could use the structure of the input to learn a new phonetic category despite variation along a second dimension. The exemplar model could thus provide a possible explanation for how we learn phonetic categories in a naturalistic environment. If the additional non-linguistic variation reduced participants' ability to discriminate the endpoint tokens, this would call into question whether the exemplar model can adequately explain how we learn phonetic categories despite irrelevant variation in the speech signal.

Experiment 1

Method

Participants

Twenty native English speakers (12 female) participated in the study for course credit or $5. Eight participants were excluded because they rated themselves as 6 or higher on a 9-point fluency scale for a second language, 2 were excluded for response bias, and 1 was excluded for failing to achieve 80% accuracy on filler trials. The participants were an average of 20 years old (range: 18-27). Participants were randomly assigned to the bimodal or the unimodal familiarization group.

Stimuli

The stimuli were developed from the stimuli used in Maye (2000). The experimental stimuli were three continua of 8 steps each, spanning /dɑ/-/tɑ/, /dæ/-/tæ/, and /dəˠ/-/təˠ/. The stimuli were American English speech resynthesized using Kay Elemetrics ASL (Maye, 2000). The formant transitions were modified to gradually change from steeper formant transitions on the /d/ end of the continuum to less steep on the /t/ end. The rate of transition highly correlates with voice onset time and is perceived as a cue to voicing. Prevoicing was added to the /d/ half of the continuum (Maye, 2000). Stimuli from Maye (2000) were modified in Praat (Boersma & Weenik, 2010) to create two versions of each token, one 3 semitones higher and one 3 semitones lower in pitch than the original token. This created a high pitch and low pitch version of each of the tokens in the three continua. As in Maye (2000), the familiarization period included filler stimuli, 4 different utterances of /mɑ/, /lɑ/, /mæ/, /læ/, /məˠ/, and /ləˠ/, also modified to have high and low pitch versions. After the pitch of the stimuli was modified, they were leveled to have the same average perceived loudness using the Replay Gain algorithm in Praat.

Design

Familiarization phase. During the familiarization phase, participants heard an equal number of experimental and filler stimuli presented in a random order. The frequency of experimental stimuli within the continua was varied according to familiarization group. The familiarization phase included an equal number of stimuli from each of the three continua. Participants heard both the high and the low pitch versions of each stimulus, for a total of 96 experimental and 96 filler stimuli presented per block, double the number of the tokens presented in Maye (2000). There was a 750

ms interstimulus interval between each token. Participants listened to four blocks, with optional short breaks in between.

Half of the participants were assigned to a bimodal familiarization group, half to a unimodal familiarization group, varying on the frequency distribution of the three /d/-/t/ continua during the familiarization phase. The bimodal group heard an increased number of tokens at the endpoints of the continua, creating two modes. The unimodal group heard an increased number of tokens at the center of the continua, creating one mode. The two endpoint tokens were played the same number of times in each familiarization condition to control for participants' familiarity with these endpoint tokens in the test phase.

Test phase. There were a total of 48 test pairs, 24 filler and 24 experimental pairs. The 24 experimental pairs came in 4 different types, varying on whether or not the pairs were matched or mismatched on pitch and phoneme (see Table 1). Each of the six endpoints (/dɑ/, /tɑ/, /dæ/, /tæ/, /dɚ/, /tɚ/) from the three continua was tested four times for a total of 24 experimental pairs. Each participant also heard 24 filler pairs, also differing on pitch and phoneme. All the test pairs were matched for the vowels, and differed only on the initial phoneme. The two tokens in a test pair were separated by a 750 ms interstimulus interval.

Table 1. The Four Experimental Pair Types, Illustrated With /da/

	Same Pitch	**Different Pitch**
Same Phoneme	dɑHI-dɑHI *daLO-daLO*	dɑHI-dɑLO *daLO-daHI*
Different Phoneme	dɑHI-tɑHi *daLO-taLO*	dɑHi-tɑLO *daLO-taHI*

Procedure

Participants sat facing an iMac computer that presented visual instructions. Auditory stimuli were presented with a pair of Sennheiser HD 280 headphones at comfortable listening level. Participants were instructed that they would hear words from a different language, and that they should listen as carefully as they could because they would be tested on them later. Between one and nine tones of 200 Hz were interspersed between the speech tokens in each block. Participants were told to keep a tally of how many tones they heard to ensure that they attended to the task.

For the test trials, participants responded whether the test pairs were the same or different by pressing the 'p' and 'q' keys (counterbalanced) on a computer keyboard. Just before the test phase, participants were instructed that although different speakers may sound different when they say the same word, the point of the experiment was to identify when the words differed, not the speakers. They were also told to respond as quickly as they could while maintaining accuracy.

Results and Discussion

Discriminability was calculated for the test trials for each participant using signal detection theory (MacMillan & Creelman, 1991). "Different" responses to test trials that differed on phoneme were counted as hits, and "different" responses on test trials with the same phoneme were counted as false alarms. There was no effect of familiarization group on *d'* (bimodal, $M = 1.93$ $SD = 1.35$, unimodal $M = 0.95$, $SD = 1.28$, $t(18) = 0.98$, $p = .34$, Cohen's $d = 0.46$). The frequency distribution of tokens in the familiarization phase had no effect on discrimination in the test trials.

To compare the results to Maye (2000), an accuracy score was calculated based solely on the different phoneme test trials by dividing the number of "different" responses by the total number of different phoneme trials. Accuracy did not differ by familiarization group (bimodal $M = .35$, $SD = .24$, unimodal $M = .27$, $SD = .23$, $t(18) = 0.8$, $p = .43$, $d = 0.38$). Accuracy on filler trials was 93%, demonstrating that participants performed the task correctly.

To confirm that participants responded on the basis of VOT rather than pitch, accuracy was also scored as if participants had made same/difference judgments with respect to pitch, instead of VOT. Accuracy on same/different pitch judgments ($M = .50$, $SD = .08$) was reliably lower than accuracy on same/different VOT ($M = .62$, $SD = .10$; this accuracy measure includes both same and different trials unlike the accuracy scores reported above; $t(18) = 6.05$, $p < .001$, $d = 1.67$). This indicates that participants were more likely to have responded on the basis of VOT than pitch.

Experiment 1 thus showed that participants did not induce two categories when tokens included variable pitch. Experiment 2 tested whether participants exposed to a familiarization period that varied on a single dimension would be more likely to learn phonetic categories from structured input.

Experiment 2

Method

Participants

Twenty participants (13 female) participated in the study for course credit or $5. Eleven participants were excluded because they rated themselves as 6 or higher on a 9-point fluency scale for a second language, 5 were excluded for response bias, 3 were excluded for failing to achieve 80% accuracy on filler trials, and 3 were excluded for having reactions times more than 2 standard deviations below the mean. The participants were an average of 21 years old (range: 18-29). Participants were randomly assigned to the bimodal or the unimodal familiarization group.

Stimuli, Design, and Procedure

The stimuli, design, and procedure were identical to Experiment 1, except that there was no pitch variation present in the speech sounds. The sounds were identical to the stimuli used in Maye (2000) but we doubled the familiarization period to match the familiarization period from Experiment 1.

Results and Discussion

The bimodal group had marginally higher *d'* values than the unimodal group as measured by a two-tailed t-test ($t(18) = 1.54$, $p = .07$, $d = 0.73$; bimodal, $M = 2.93$, $SD = 1.12$, unimodal $M = 2.03$, $SD = 1.07$) and also reliably higher accuracy ($t(18) = 2.64$, $p = .01$, $d = 1.24$; bimodal, $M = .57$, $SD = .26$, unimodal $M = .29$, $SD = .19$).

Participants exposed to a familiarization period with no pitch variation were more likely to discriminate a new phonetic contrast based on the frequency distribution to which they were exposed. Notably, an attempt to directly replicate Maye (2000) using half the familiarization trials of the current experiment did not result in differences in *d'* or accuracy between the bimodal and unimodal groups.

General Discussion

Variability on a second non-linguistic feature affected implicit phonetic category learning. Pitch variability in the familiarization phase attenuated participants' ability to learn new phonetic categories from exposure to a frequency distribution. Although linguistic and non-linguistic features might be encoded separately (Pierrehumbert, 2001), variation on a second dimension interacted with the acquisition of the linguistic feature in this experimental task. However, in a natural language context, phonetic category learning occurs despite variability on more than one dimension. The failure to accommodate variability on an irrelevant dimension in this implicit task could have several explanations.

One possible explanation for the detrimental effect of variability on phonetic category learning is that different levels of encoding interact through their demands on shared cognitive resources. This is consistent with the PRIMIR model (Werker & Curtin, 2005) in which representations include multiple dimensions that encode different information, including phonetic and indexical information. Different types of information are available to the listener depending on the task and developmental time point, and indexical information can affect how linguistic information is processed. For example, adults are more likely to remember words when they are spoken in a familiar voice (Nygaard, Sommers, & Pisoni, 1995). In Experiment 1, learners encountered all the information available in Experiment 2, but had more difficulty using the VOT information than learners in Experiment 2 because the added indexical information placed demands on limited processing capacity rendering it more difficult to learn a new linguistic category. This hypothesis could be tested directly by extending the familiarization period, which could allow learners to accommodate irrelevant variation better.

Another possibility is that the distribution of the variation on the irrelevant dimension affected the learning of the relevant dimension. There may not have been enough variability in the secondary dimension for listeners to filter it out as noise, as the pitch in the current experiment is, in fact, bimodal. Higher variability in the irrelevant dimension makes it easier to filter out variability as noise in the system (Restle, 1955). This explanation is formalized by cue-weighting models (e.g., Love, Medin, & Gureckis), which have recently been applied to speech (Toscano & McMurray, 2010). There is also behavioral evidence supporting the beneficial role of heightened variability in a linguistic context. In an artificial grammar-learning task, adults can

learn non-adjacent dependencies more easily if the element interposed between the dependent elements is drawn from a larger pool of possibilities (Gomez, 2002). Presenting only two pitches may not have been enough variability–exposure to more variation in pitch might make pitch easier to filter out as an irrelevant dimension.

Alternatively, participants may have learned idiosyncratic acoustic differences between pairs of tokens rather than phonetic categories per se. Specifically, participants in Experiment 2 may have used acoustic or idiosyncratic cues that were specific to the tokens rather than creating phonetic categories that used the relevant phonetic feature, VOT. This type of learning, which relies on encoding idiosyncratic acoustic differences between pairs of tokens, would be disrupted by the addition of a new dimension of acoustic variation as we did in Experiment 1, because many more unique tokens were presented. Specifically, increasing the number of tokens may have created too many pairings of tokens to learn how to differentiate each one individually. Learning through acoustic cues could underlie the ability of adults to discriminate within phonetic category differences, both of native and non-native categories (Pisoni & Tash, 1974; Werker & Logan, 1985). The bimodal frequency distribution might have enhanced these acoustic differences in Experiment 2 without actually facilitating acquisition of the phonetic feature. This interpretation is consistent with adults' failure to generalize on a feature level in Maye (2000). Specifically, if adults were trained on the /d/-/t/ contrast, they did not generalize the newly learned voicing boundary to the /g/-/k/ continuum, unlike infants, who can generalize to a new /g/-/k/ contrast (Maye, Weiss, & Aslin, 2008). This suggests that adults might not have acquired VOT as a phonetic feature. To confirm that phonetic acquisition has occurred, it may be necessary to test listeners on a task that cannot be performed on the basis of idiosyncratic acoustic variability, such as phoneme identification (Pisoni & Tash, 1974).

Participants might be able to handle extraneous variation differently depending on their developmental stage. Adults may have greater difficulty learning than infants because adults are more likely to learn acoustic rather than phonetic differences (Werker & Logan, 1985). Humans generally form new phonetic categories by the end of the first year of life, and the decrement in ability to learn new phonetic structures with age is well documented (e.g., Miyawaki, Strange, Verbrugge, Liberman, Jenkins, & Fujimura, 1975). One reason that the adults may be more likely to encode at an acoustic rather than at a phonetic level may come from their existing linguistic experience. Specifically, adults have had far more experience with frequency distributions in the environment than infants simply as a function of time. Adults' phonetic categories are thus composed of many more exemplars, and, as a result, any new exemplar that is encountered is a much smaller percentage of the input, and will have a smaller effect on the adults' existing categories. Even by 10 months, infants require a longer familiarization period than 6- to 8-month-olds to show effects of frequency distribution on phonetic discrimination (Yoshida, Pons, Maye, & Werker, 2010). If the effects of variability are a function of age or experience, infants should be better able to learn phonetic categories despite variation on a second dimension.

Infants may show better performance on this task for yet another reason. Children are better than adults at L2 grammaticality judgments, and are more likely to report a preference for their second language (Jia & Aaronson, 2003; Johnson &

Newport, 1989). According to the critical period hypothesis, experience during childhood is crucial because of an increased sensitivity to language input (Newport, 2002). Children begin life with plastic neural structures that gradually become less plastic as they get older, making children better language learners than adults (Newport, 2002). Speech perception requires exposure to a phonetic contrast in infancy for discrimination to be maintained, which is possible evidence of an optimal period of phonetic category learning (see Werker & Tees, 2005, for a review). Adults in this task may only be able to learn at the acoustic level, lacking the plasticity required to learn new phonetic categories. The variability in Experiment 1 may have reduced their ability to discriminate based on acoustic differences.

These results emphasize the importance of studying speech perception in the wider context of general cognition and learning. The detrimental effect of irrelevant variability on a categorization task without feedback has been found using lines varying on length and orientation (Zeithamova, & Maddox, 2009). Corroborating previous findings in categorization studies without feedback (e.g., Clapper & Bower 1994), Zeithamova and Maddox (2009) show that seemingly benign factors such as presentation order can change the relative salience of the available dimensions and influence categorization. Applying principles from the general categorization literature to phonetic category learning could reveal the similarities and differences between how we learn categories for speech and other stimuli, as well as suggest effective strategies for L2 learning.

In conclusion, variation on a second dimension attenuates phonetic category learning from a frequency distribution. The availability of feedback, amount of variability, length of exposure or the type of learning taking place may explain why participants failed to learn when additional variability was introduced. Further studies with different stimuli and different age groups may be able to clarify why variability has an effect on phonetic category learning, and to explain how non-linguistic and linguistic information interact. These findings highlight the importance of non-linguistic cues in phonetic acquisition, and underscore the difficulty of learning phonetic categories from a frequency distribution.

References

Boersma, P., & Weenink, D. (2010). Praat: doing phonetics by computer [Computer program]. Version 5.1.07, retrieved 12 May 2009 from http://www.praat.org/.

Clapper, J. P., & Bower, G. H. (1994). Category invention in unsupervised learning. *Journal of Experimental Psychological Learning, 20*, 443-460.

Gilkerson, J. (2003). Categorical perception of natural and unnatural categories: Evidence for innate category boundaries. Poster presented at the Boston University Conference on Language Development, Boston University, Boston, MA.

Goldstone, R. L., & Hendrickson, A. T. (2010). Categorical perception. *Interdisciplinary Reviews: Cognitive Science, 1*, 65-78.

Gomez, R. L. (2002) Variability and detection of invariant structure. *Psychological Science. 13*, 431-436.

Jia, G. & Aaronson, D. (2003). A longitudinal study of Chinese children and adolescents learning English in the US. *Applied Psycholinguistics, 24*, 131-161.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology, 21*, 60-99.

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

Lisker, L., & Abramson, A.S (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word, 20*, 384-422

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 - 886.

Love, B. C., Medin, D. L., and Gureckis, T. M. (2004) SUSTAIN: A Network Model of Category Learning. *Psychological Review, 11*, 309-332.

MacMillan, N. A., & Creelman, C. D., (1991). Detection Theory: A User's Guide, New York: Cambridge University Press.

Maye, J. (2000). Learning speech sound categories on the basis of distributional information. Unpublished doctoral dissertation, University of Arizona, Tucson.

Maye, J., Weiss, D. J., & Aslin, R. N. (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental Science, 11*, 122-134.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82*, B101-B111.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. M., Jenkins, J. J., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception & Psychophysics, 18*, 331-340.

Newport, E. L. (2002). Critical periods in language development. In L. Nadel (Ed.), *Encyclopedia of Cognitive Science*. London, UK: Macmillan Publishers Ltd./Nature Publishing Group.

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1995). Effects of stimulus variability on perception and representation of spoken words in memory. *Perception & Psychophysics, 57*, 989-1001.

Restle, F. (1955). A theory of discrimination learning. *Psychological Review, 62*, 11-19.

Rost, G., & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental Science*, *12*, 339-349.

Pierrehumbert, J. (2001). Exemplar dynamics: Word frequency, lenition, and contrast. In J. Bybee and P. Hopper (eds.), *Frequency effects and the emergence of lexical structure* (pp.137-157). Amsterdam, Netherlands: John Benjamins.

Pierrehumbert, J. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech*, *46*, 115–154.

Pisoni, D. B., & Tash, J. (1974) Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics*, *15*, 285-290.

Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word learning tasks. *Nature*, *388*, 381-382.

Toscano, J. C., & McMurray, B. (2010). Cue integration with categories: Weighting acoustic cues in speech using unsupervised learning and distributional statistics. *Cognitive Science*, *34*, 434-464.

Werker, J. F., & Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development*, *1*, 197-234.

Werker, J. F., & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Perception and Psychophysics*, *37*, 35-44.

Werker, J. F. & Tees, R. C. (2005). Speech perception as a window for understanding plasticity and commitment in language systems of the brain. *Developmental Psychobiology*, *46*, 233-251.

Yoshida, K. A., Pons, F., Maye, J., & Werker, J. F. (2010). Distributional phonetic learning at 10 months of age. *Infancy*, *15*, 420-433.

Zeithamova, D., & Maddox, W. T. (2009). Learning mode and exemplar sequencing in unsupervised category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 731-741.