Statistical Cues Facilitate Infants' Discrimination of Difficult Phonetic Contrasts

Jessica Maye and Daniel Weiss
University of Rochester

Perhaps the best-known fact about developmental speech perception is that infants are remarkably adept at discriminating phonetic contrasts. In early infancy, this ability is unaffected by language environment, resulting in the surprising fact that infants can discriminate certain foreign contrasts that their parents cannot. For example, infants from English-speaking homes can hear the difference between [ž] and [ř], two fricatives used in Czech that English-speaking adults have difficulty discriminating (Trehub, 1976). Infants' advantage at foreign contrast discrimination wanes over the course of the first year, though, as they gain experience with their native language; and by the age of 12 months infants no longer discriminate those foreign contrasts (Werker & Tees, 1984). Developmental speech perception, then, can largely be described as a process of paring down previously discriminable contrasts, to just that set of contrasts that is utilized in the native language.

However, though infants' discrimination of many phonetic contrasts exceeds adults', there are in fact some phonetic contrasts that are difficult for infants. Among them are the contrasts between prevoiced and short-lag stops (Eimas, 1975; Aslin & Pisoni, 1980), between [d] and [ð] (Polka, Colantonio, & Sundara, 2001), and between fricative pairs such as [s]~[z] and [f]~[θ] (Eilers & Minifie, 1975; Eilers, 1977). Exposure to a language in which these contrasts are used phonemically apparently results in enhanced discrimination. For example, the [d]~[ð] contrast is discriminated well by adult English speakers, though not by speakers of French, a language in which it is not phonemic (Polka et al., 2001). The process of an infant's developing perception of speech must therefore involve not only paring down of initially discriminable contrasts, but also enhancement of initially difficult contrasts. The goal of the present study was to examine how the facilitation of difficult native language contrasts might occur.

In particular, the hypothesis we proposed was that changes in infants' perception of difficult contrasts might be driven by statistical cues, based on the distribution of phonetic tokens in the input. Previous research has shown that phonetic contrasts are instantiated in the distribution of speech sounds produced in a language. For example, in Thai there are three voicing categories, while in English there are only two. This fact is evident from the distribution of stop consonant VOT values produced in each language: for Thai there is a trimodal distribution (the most commonly produced sounds form three clusters, corresponding to the three Thai voicing categories: prevoiced, unvoiced, and aspirated), while in English the distribution is bimodal (Lisker & Abramson, 1964). The difference between the distribution of VOT values in the two languages is also reflected in the discrimination abilities of speakers of the languages. In the prevoiced~short-lag region of VOT, where Thai speakers have two categories and English speakers have only one, Thai speakers show good discrimination of sounds that cross the Thai voicing boundary, while English speakers' discrimination is poor throughout that region of VOT (Abramson & Lisker, 1970).

If infants are able to track this distributional information, then these cues might contribute to developmental changes in speech perception. If this is true, then exposure to a bimodal distribution of sounds should result in enhanced discrimination, while exposure to a unimodal distribution should result in reduced discrimination. In a previous study (Maye, Werker, & Gerken, 2002), we provided evidence supporting the latter half of this hypothesis. Namely, we found that two sounds that are discriminable in early infancy were no longer discriminated by infants who had been familiarized to the sounds within a unimodal distribution. In the current study, our goal was to test the first half of the hypothesis: does exposure to a bimodal distribution result in enhanced discrimination of a difficult phonetic contrast?

1. Experiment 1

We chose to test infants' discrimination of prevoiced vs. short-lag stop consonants, because this contrast has been shown to be a difficult one for infants (Eimas, 1975; Aslin & Pisoni, 1980). We predicted that infants would show greater discrimination following exposure to a bimodal distribution of these stimuli than they would with no prior exposure to these stimuli.

1.1. Method

1.1.1. Participants

64 8-month-old infants participated in the study. Infants were from English-speaking homes, and were recruited based on parental interest in research participation. Infants who received regular exposure to a language other than English were excluded from the study.

1.1.2. Stimuli

We recorded multiple natural tokens of the syllables [da] and [ga] (both prevoiced), and [ta] and [ka] (both unaspirated), as produced by a speaker of Hindi. Four tokens of each of the unaspirated syllables were chosen, from which to make four coronal and four velar continua. These syllables were then digitized and edited using SoundEdit 16.2. We removed portions of the unvoiced lag, to create tokens with 0, 7, 14, and 21 msec voicing lag. We then spliced naturally produced prevoicing (from [da] and [ga]) onto the 0 msec lag tokens, to create prevoiced stimuli with –100, –75, –50, and –25 msec voicing lead. The result was eight 8-step voicing continua: four continua from [da] to [ta], and four from [ga] to [ka].

1.1.3. Procedure

The experiment was conducted in two phases: familiarization and test. For the duration of the experiment, infants were seated on a parent's lap in front of a video monitor. Auditory stimuli were presented via a speaker located below the monitor. Parents listened to masking music through headphones.

Infants were randomly assigned to one of four conditions, which differed with respect to the auditory stimuli presented during familiarization. Infants in two of the conditions heard the experimental stimuli ([da]~[ta] or [ga]~[ka]) presented in a bimodal distribution (see Figure 1), such that stimuli near the endpoints of the continuum were presented most frequently, and center stimuli were presented infrequently. The two bimodal groups differed with respect to which place of articulation (coronal or velar) they were familiarized to and tested on. Infants listened to familiarization stimuli for 2.5 minutes, while they watched a short video clip.

In order to assess the effect of familiarization, as well as to ensure that this contrast is indeed a difficult one for infants, we compared each bimodal group's discrimination to that of a control group who received no exposure to the continuum stimuli prior to the test phase. Thus, there were two control groups that differed with respect to which place of articulation they were tested on. To make the test and control groups as similar as possible, the control groups were presented with irrelevant auditory stimuli (a random sequence of tones) while they watched the 2.5 minute video clip.

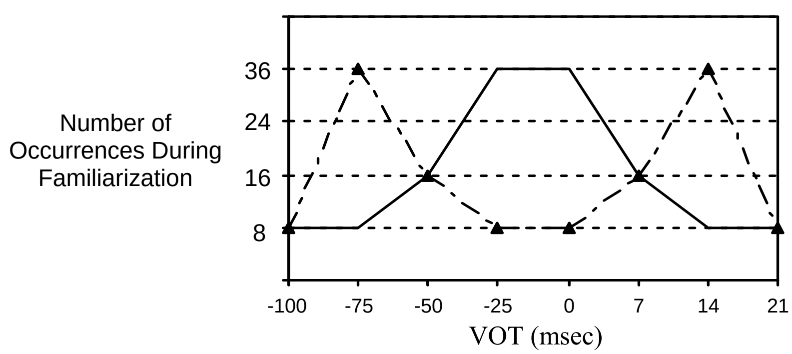

Figure 1. Presentation Frequency for Continuum Stimuli During Familiarization Phase. The Bimodal condition (Experiments 1 & 3) is shown by the broken line; the Unimodal condition (Experiment 2) is shown by the solid line.

Infants' discrimination was tested through a habituation procedure, in which the dependend measure was looking time. On each habituation trial, a colorful bullseye appeared on the video monitor, and the infant heard the four +7 msec lag tokens, from the appropriate place of articulation, presented in random order. Each trial continued until the infant looked away from the screen for 2 seconds, (upto a maximum of 60 seconds trial length). Habituation was assessed via a moving window that compared the total looking time for the first three trials to that of each subsequent set of three trials. The habituation criterion was satisfied when a window was reached in which looking time was at or below 50% of the initial window, at which point two change trials were presented. The change trials were identical to habituation trials, except that the auditory stimulus had –50 msec VOT. Discrimination is indicated by an increase in looking time between habituation and change trials.

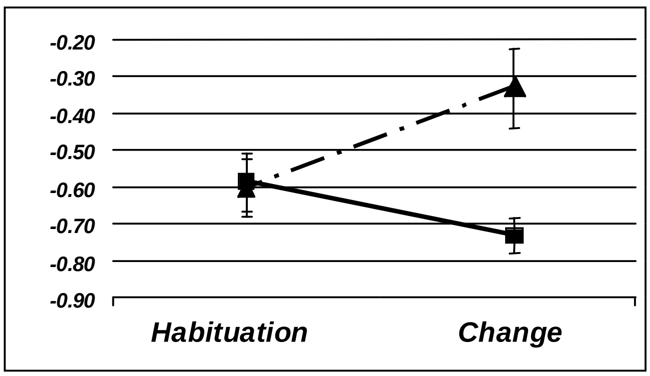

Figure 2. Normalized looking time scores for infants in the Bimodal (broken line) and Control (solid line) groups.

1.2. Results

To assess discrimination, we compared the average looking time for the last two habituation trials to the average looking time for the two change trials. To reduce variability, looking time scores were normalized by z-scoring and square root transformation. These results are shown in Figure 2, collapsed across place of articulation. A mixed-design ANOVA (2 familiarization types x 2 places of articulation x 2 trial types) revealed no main effect of place of articulation or trial type (both $F[1,60] < 1$, NS), but a significant effect of familiarization ($F[1,60] = 4.77$, $p < .05$) and a significant interaction between familiarization and trial type ($F[1,1] = 8.26$, $p < .01$). Planned pairwise comparisons revealed that infants in the bimodal conditions showed a significant dishabituation on change trials ($t[31]$, 2-tailed = 2.15, $p < .05$), while infants in the control conditions showed a significant *decrease* in looking times, indicating a transfer of habituation (t[31], 2-tailed = 2.15, $p < .05$).

1.3. Discussion

The results from Experiment 1 support our hypothesis that exposure to a bimodal distribution results in enhanced discrimination of a difficult contrast. The control groups' transfer of habituation confirms that this contrast is difficult for infants to discriminate in the absence of prior familiarization, while the bimodal groups' significant dishabituation shows that the contrast is discriminated following exposure to a bimodal distribution. However, there is an alternate explanation for these results. It may be that the bimodal groups' advantage came from simply being exposed to these sounds during the familiarization phase, not from the particular distribution in which the sounds occurred. To rule out this possibility, we conducted a second experiment in which we assessed infants' discrimination following exposure to the same sounds, but arranged in a distribution that is not expected to facilitate discrimination.

2. Experiment 2

Maye et al. (2002) found that exposure to a unimodal distribution of speech sounds hindered infants' ability to discriminate previously discriminable speech sounds. Since we wanted to rule out the possibility that the findings from Experiment 1 resulted from mere exposure to speech stimuli, in this experiment we familiarized infants to a unimodal distribution (see Figure 1) of the continuum stimuli. We compared their performance on the discrimination task to that of the Control groups from Experiment 1. We predict that if infants' discrimination is facilitated by mere exposure to the speech sounds, then infants in the unimodal condition should perform similarly to the bimodal condition of Experiment 1, and dishabituate to the change stimulus. However, if infants are attending to the shape of the distribution, then infants in the unimodal condition should perform similarly to the control condition of Experiment 1, and fail to dishabituate.

2.1. Methods

We familiarized 27 8-month-olds to a unimodal distribution of the continuum stimuli (16 to the velar stimuli, 11 to the coronal stimuli). In all other respects, the methods were identical to Experiment 1.

2.2. Results

As in Experiment 1, we compared the average looking time for the last two habituation trials to the average looking time for the two change trials, and normalized looking time scores by z-scoring and square root transformation. These results are shown in Figure 3, collapsed across place of articulation. The two familiarization types compared were Unimodal (Experiment 2) and Control (Experiment 1). A mixed-design ANOVA (2 familiarization types x 2 places of articulation x 2 trial types) revealed no main effects of place of articulation, familiarization, or trial type (all $F[1,55] < 1$, NS), and no significant interaction between familiarization and trial type ($F[1,1] = 1.26$, $p = .27$). Infants in the Unimodal conditions performed similarly to infants in the control conditions, also transferring habituation.

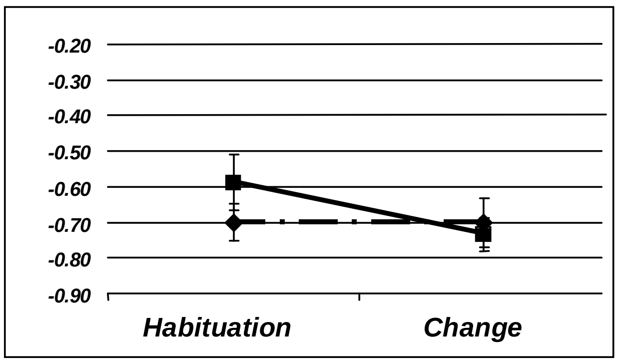

Figure 3. Normalized looking time scores for infants in the Experiment 2 Unimodal (broken line) and Experiment 1 Control (solid line) groups.

2.3. Discussion

The results from Experiment 2 confirm that the effect found in Experiment 1 does not result from mere exposure to the speech stimuli, but rather from exposure to the particular type of distribution indicative of a contrast, namely a bimodal distribution. As predicted, exposure to the same sounds arranged in a unimodal distribution, which is not indicative of a contrast, does not result in enhanced discrimination.

Together, the results from Experiments 1 and 2 show that the contrasts presented in these experiments are difficult for infants to discriminate in the absence of familiarization. The results of Experiment 1 indicate that familiarization to a bimodal distribution of sounds facilitates discrimination. The results of Experiment 2 rule out the possibility that this effect is merely a product of exposure to the speech stimuli; rather, enhancement occurs on the basis of exposure to the particular distribution of sounds (i.e. bimodal) that is indicative of a phonetic contrast.

3. Experiment 3

The results of Experiments 1 and 2, combined with the findings of Maye et al. (2002), support the notion that distributional information affects early perceptual development, resulting in both the loss of non-native contrast discrimination as well as the facilitation of difficult contrasts. In Experiment 3 we investigate the level at which infants encode these contrasts during the course of distributional learning. One possibility is that infants initially learn at the level of the segment. That is, an infant exposed to a bimodal distribution of [g]~[k] learns that the categories [g] and [k] are contrastive in the language. A second possibility is that infants encode these speech sounds at the more abstract level of the phonetic feature. If this is the case, then an infant exposed to a bimodal distribution of [g]~[k] learns that in this language there is a *voicing* contrast. If infants are learning at the level of the feature, then their learning should not be specific to the speech sounds encountered during familiarization; rather, bimodally familiarized infants should also discriminate other contrasts exemplifying the same featural relationship (e.g. [d]~[t]). However, if infants are learning at the level of the segment, exposure to a bimodal distribution of [g]~[k] should have no effect on discrimination of other voicing contrasts, such as [d]~[t]. To test the level at which infants encode speech sounds during the course of distributional learning, in Experiment 3 we familiarized infants to a bimodal distribution of sounds at one place of articulation and tested discrimination of the same featural contrast at an untrained place of articulation.

3.1 Methods

In this experiment, 22 8-month-olds were familiarized to a bimodal distribution (see Figure 1) of the prevoiced/short-lag contrast at one place of articulation (11 to the velar stimuli, 11 to the coronal stimuli), and subsequently tested on their discrimination of the prevoiced/short-lag contrast at the other place of articulation. For example, one group of infants were familiarized to a bimodal distribution of the coronal stimuli, and then tested on discrimination of the velar contrast. We then compared their discrimination performance to that of the control groups from Experiment 1. In all other respects, the methods were identical to Experiment 1.

3.2. Results

The results from Experiment 3 are shown in Figure 4, collapsed across place of articulation. As in Experiments 1 and 2, we compared the average looking time for the last two habituation trials to the average looking time for the two change trials, and normalized looking time scores by z-scoring and square root transformation. The two familiarization types compared were Generalization (Experiment 3) and Control (Experiment 1). A mixed-design ANOVA (2 familiarization types x 2 places of articulation x 2 trial types) revealed that the only significant effect was an interaction between familiarization and trial type ($F[1,1] = 6.52$, $p < .02$), indicating that the Generalization infants discriminated the untrained contrast.

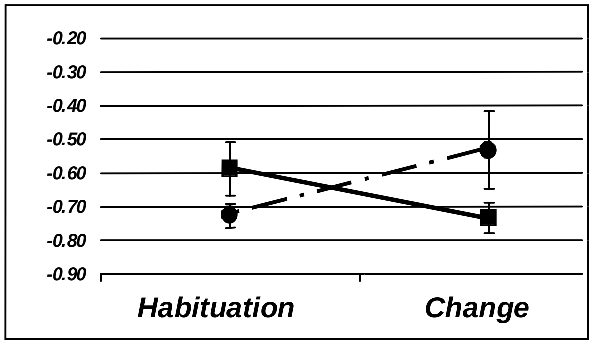

Figure 4: Normalized looking time scores for infants in the Experiment 3 Generalization (broken line) and Experiment 1 Control (solid line) groups.

3.3. Discussion

The results from Experiment 3 indicate that exposure to a bimodal distribution at one place of articulation facilitates discrimination at a second, untrained place of articulation. In other words, the infants in Experiment 3 appear to have extracted the featural properties of the input speech, after only 2.5 minutes of exposure. This finding is in line with research showing that 9-month-old infants demonstrate an awareness that sets of sounds forming a natural class are more similar than sets with no unifying phonetic features (Jusczyk, Goodman, & Baumann, 1999). At least by the age of 8 months, infants appear to encode speech sounds on an abstract level, on the basis of the featural relationship between the sound categories of a language.

This finding is particularly interesting in light of previous research showing that adult participants who learn to discriminate a contrast via exposure to a bimodal distribution do *not* generalize to an untrained place of articulation (Maye & Gerken, 2001). However, the methods used in the adult study were not completely analogous to those used in this experiment (the previous study may have been confounded by metalinguistic factors). Thus, one future direction will entail a replication of the adult study, employing a methodology more closely resembling that of the current study, in order to determine whether the discrepant findings are due to methodological differences. Another possibility is that infants, who are in the process of honing in on a first language, may extract information at a different level than adults. Infants appear to extract the featural properties of the input speech, while adult learning may be restricted to the segmental level.

Another direction we plan to explore is whether nonhuman primates will show similar facilitation effects when familiarized to difficult speech contrasts. A number of recent studies have shown that nonhuman primates are capable of processing many features of human speech in a manner similar to humans. For example, cotton-top tamarin monkeys are sensitive to the rhythmic properties of human languages (Ramus et al., 2000), are capable of segmenting words from a speech stream using transitional probabilities (Hauser, Newport, & Aslin, 2001), and may extract rudimentary rules from linguistic patterns (Hauser, Weiss, & Marcus, 2002). Recently, we replicated the Maye et al. (2002) phonetic learning study with tamarin monkeys, and found that, like human infants and adults, tamarins are sensitive to the distributional properties of a speech stream (Weiss, Maye, & Tincoff, in prep.). This places us in the ideal position to replicate the current series of experiments with tamarins, and investigate the level at which these non-human primates analyze human speech input.

Replicating this study with nonhuman primates is particularly important in light of the findings of many of the aforementioned nonhuman primate studies. Those studies found similar abilities in speech perception for both human infants and nonhuman primates. However, since only humans go on to ultimately acquire human language, there must at some level be differences in the way that the speech stream is processed by the different species. These differences may take the form of unique language-specific mechanisms, or an elaboration of mechanisms that are homologous in both species. The latter possibility would entail that although a distributional learning mechanism is common to multiple species, this mechanism is shaped by natural selection to take advantage of each species' unique environmental needs, resulting in differences in how the mechanism operates in different species.

Our current line of research will provide a testing ground for this comparison. We predict that distributional information *will* facilitate tamarins' discrimination of difficult contrasts (in line with the findings of Weiss, Maye and Tincoff, in prep.). However, we can make no prediction regarding tamarins' ability to generalize to analogous contrasts. It seems likely that only humans have access to phonetic features, and if this is the case then tamarins are not expected to encode speech input at a featural level. However, tamarins have been shown to form categories over linguistic input and generalize to acoustically different stimuli (Hauser, Weiss, & Marcus, 2002), and thus they may in fact encode speech input at a relatively abstract level. If tamarins do learn at the level of the feature, it would suggest that adult humans' failure to do so arises from their having already acquired a language.

4. Conclusion

The overall results from this study have shown that exposure to a bimodal distribution of speech sounds results in facilitated discrimination. We have shown that, for a phonetic contrast that is previously not discriminable, infants familiarized to a unimodal distribution remain unable to discriminate the contrast, whereas familiarization to a bimodal distribution results in enhanced discrimination. In addition, infants appear to be performing this computation at the level of the phonetic feature, resulting in the bimodal infants' ability to also discriminate a previously non-discriminable, untrained contrast that exhibits the same feature. Combined with the results from the Maye et al. (2002) study, these data suggest that infants' sensitivity to the distributional properties of speech can account for both the pruning and enhancement patterns seen in infants' development of native language speech perception.

References

Abramson, A. S., & L. Lisker (1970). Discriminability along the voicing continuum: Cross-language tests. In *Proceedings of the Sixth International Congress of Phonetic Sciences.* Prague: Academia.

Aslin, R. N., & D. B. Pisoni (1980). Some developmental processes in speech perception. In G. H. Yeni-Komshian, J. H. Kavanagh, & C. A. Ferguson (eds.), *Child Phonology, 2: Perception.* New York: Academic Press.

Eilers, R. E. & F. D. Minifie (1975). Fricative discrimination in early infancy. *Journal of Speech & Hearing Research, 18*, 158-167.

Eilers, R. E. (1977). Context-sensitive perception of naturally produced stop and fricative consonants by infants. *Journal of the Acoustical Society of America, 61*, 1321-1336.

Eimas, P. D. (1975). Speech perception in early infancy. In L. B. Cohen & P. Salapatek (eds.), *Infant Perception, 2: From Sensation to Cognition.* New York: Academic Press.

Hauser, M. D., E. L. Newport, & R. N. Aslin (2001). Segmentation of the speech stream in a non-human primate: Statistical learning in cotton-top tamarins. *Cognition, 78*, B53-B64.

Hauser, M. D., D. J. Weiss, & G. Marcus (2002). Rule learning by cotton-top tamarins. *Cognition, 86*, B15-B22.

Jusczyk, P. W., M. B. Goodman, & A. Baumann (1999). Nine-month-olds' attention to sound similarities in syllables. *Journal of Memory and Language, 40*, 62-82.

Lisker, L., & A. S. Abramson (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word, 20*, 384-482.

Maye, J., & L. Gerken (2001). Learning phonemes: How far can the input take us? In A. H-J. Do, L. Domínguez, & A. Johansen (eds.), *Proceedings of the 25th Annual Boston University Conference on Language Development* (p. 480-490). Somerville, MA: Cascadilla Press.

Maye, J., J. F. Werker, & L. Gerken (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82 (3)*, B101-B111.

Polka, L., C. Colantonio, & M. Sundara (2001). A cross-language comparison of /d/~/D/ discrimination: Evidence for a new developmental pattern. *Journal of the Acoustical Society of America, 109*, 2190-2201.

Ramus, F., M. D. Hauser, C. T. Miller, D. Morris, & J. Mehler (2000). Language discrimination by human newborns and cotton-top tamarin monkeys. *Science, 288*, 349-351.

Trehub, S. E. (1976). The discrimination of foreign speech contrasts by infants and adults. *Child Development, 47,* 466-472.

Weiss, D. J., J. Maye, & R. Tincoff (in preparation). Sensitivity to speech sound distributions in a non-human primate.

Werker, J. F., & R. C. Tees (1984). Developmental changes across childhood in the perception of nonnative speech sounds. *Canadian Journal of Psychology, 37,* 278-286.