Available online at www.sciencedirect.com
SCIENCE DIRECT
ELSEVIER
COGNITION
Cognition 95 (2005) B15–B26
www.elsevier.com/locate/COGNIT

Brief articles

Infants are sensitive to within-category variation in speech perception

Bob McMurray[*,1], Richard N. Aslin

Department of Brain and Cognitive Sciences, University of Rochester, Rochester, NY 14627, USA

Received 12 May 2004; accepted 19 July 2004

Abstract

Previous research on speech perception in both adults and infants has supported the view that consonants are perceived categorically; that is, listeners are relatively insensitive to variation below the level of the phoneme. More recent work, on the other hand, has shown adults to be systematically sensitive to within category variation [McMurray, B., Tanenhaus, M., & Aslin, R. (2002). Gradient effects of within-category phonetic variation on lexical access, *Cognition*, *86* (2), B33–B42.]. Additionally, recent evidence suggests that infants are capable of *using* within-category variation to segment speech and to learn phonetic categories. Here we report two studies of 8-month-old infants, using the head-turn preference procedure, that examine more directly infants' sensitivity to within-category variation. Infants were exposed to 80 repetitions of words beginning with either /b/ or /p/. After exposure, listening times to tokens of the same category with small variations in VOT were significantly different than to both the originally exposed tokens and to the cross-category-boundary competitors. Thus infants, like adults, show systematic sensitivity to fine-grained, within-category detail in speech perception.

Keywords: Speech perception; Infancy; Categorical perception

For over half a century, a major focus of research on speech perception has been the mapping of continuous perceptual detail onto discrete categories. Early work on this

* Corresponding author. Address: Department of Psychology, E11 Seashore Hall, University of Iowa, IA 52242, USA. Tel.: +1 319 335 2408; fax: +1 319 335 0191.
E-mail address: bob-mcmurray@uiowa.edu (B. McMurray).
[1] Author is now in the Department of Psychology, University of Iowa.

0022-2860/$ - see front matter
doi:10.1016/j.cognition.2004.07.005

mapping in adults suggested that consonants are perceived *only* in terms of language-relevant phoneme categories—variation below the level of the phoneme (i.e. subphonemic variation) is discarded and inaccessible to higher level processing. In particular, findings of categorical perception (Liberman, Harris, Hoffman, & Griffith, 1957) motivated the view that the analysis of speech is best described as an autonomous pattern recognition process that simply transforms acoustic input into symbolic output. This view received significant support when Eimas, Siqueland, Jusczyk, and Vigorito (1971) discovered that infants as young as 2 months displayed a pattern of discrimination consistent with categorical perception.

A sizeable body of work with adults, however, has not supported this strong version of categorical perception (Carney, Widin, & Viemeister, 1977; Miller, 1997; Pisoni & Lazarus, 1974; Pisoni & Tash, 1974; see McMurray, Aslin, Tanenhaus, Spivey, & Subik, in preparation, for a review). More importantly, a number of studies have recently demonstrated that not only are within-category variants discriminable, but that these fine-grained distinctions are retained and affect lexical processes (Dahan, Magnuson, Tanenhaus, & Hogan, 2001; Gow, McMurray, & Tanenhaus, 2003; McMurray et al., in preparation; McMurray, Tanenhaus, & Aslin, 2002). These studies suggest that subphonemic variation, rather than being treated as noise due to coarticulatory, talker and speaking rate variability, can actually be used during word recognition to take advantage of regularities in the speech signal.

Despite this emerging consensus with respect to adult speech perception, there has been no significant parallel challenge to findings of infant categorical perception—in fact, virtually all relevant studies have reported categorical perception (Aslin, Pisoni, Hennessy, & Perey, 1981; Eimas, 1974; Eimas et al., 1971; see Jusczyk, 1997, for an overview; Miller & Eimas, 1996, for an exception discussed below). Yet there are important reasons why analogous sensitivities to within-category detail could augment word recognition and word learning in infants. First, sensitivity to within-category detail may play an important role in learning the phonological categories of the infant's native language. Maye, Werker, and Gerken (2002), for example, have shown that infants' speech categories can be modified by the distribution of exemplars during a brief period of passive exposure. Infants exposed to a bimodal distribution of tokens on a da/ta continuum showed discrimination performance consistent with two categories, while infants exposed to a unimodal distribution appeared to have only one category. This evidence suggests that infants may extract categories, in part, from the distributional statistics in the input by implicitly building a frequency histogram along one or more acoustic/phonetic dimensions. For this model to work, infants must detect precisely where individual tokens lie along each dimension. While this approach has since become an important account of speech-category learning (see also Guenther & Gjaja, 1996; Pierrehumbert, 2004), no research to date has explicitly demonstrated the within-category sensitivity that it requires.

Second, sensitivity to within-category detail may be involved in learning phonological regularities of the language. Such regularities, like English coronal place assimilation (Gow, 2001), effects of prosodic strength on consonant articulation (Fougeron & Keating, 1997), coarticulatory coloration of vowels (Clark & Hillenbrand, 2003) and the relationship between voice onset time (VOT) and vowel length (e.g. Allen & Miller, 1999; Kessinger & Blumstein, 1998), take the form of *subphonemic* signal modifications.

While it is clear that *adults* use these regularities during on-line word recognition to integrate material over time and segment fluent speech (e.g. Gow, 2001; Gow & Gordon, 1995), their abilities may be augmented by the lexicon—lexical feedback may implicitly store between-unit regularities (e.g. Magnuson, McMurray, Tanenhaus, & Aslin, 2003; Samuel & Pitt, 2003), and lexical competition may segment the signal without explicit cues (McClelland & Elman, 1986). However, given infants' sparse lexicons, this route may not be available, thereby increasing the importance of bottom-up mechanisms. Indeed, it has been shown that infants are capable of using allophonic differences between phonemes at word boundaries to segment words from fluent speech (Hohne & Jusczyk, 1994; Jusczyk, Hohne, & Bauman, 1999).

Thus, previous work suggests that subphonemic sensitivity is present in infants, but does not speak to the organization and representational form of early phonetic categories. A study by Miller and Eimas (1996) on voicing discrimination in 3-month-old infants, however, provides some clues. Infants who were habituated to non-prototypical voiceless sounds (VOTs longer or shorter than the prototypical value) dishabituated to same-category, prototypical variants, implying within-category discrimination. Infants habituated to prototypical sounds, however, did not show evidence of discrimination to any tokens. While this study is important in revealing an overall prototypicality to infant speech categories, it assessed VOT differences that were too large to be of use in statistical learning (45 and 100 ms). Moreover, the asymmetry of the effects (no discrimination when habituated to prototypical tokens) makes it difficult to interpret how such VOT information might be used.

The present study contrasts with Miller and Eimas in three ways. First, we assessed discrimination after exposure to more prototypical exemplars. Second, much smaller VOT differences were used. And third, discrimination on both the voiced and voiceless sides of the category boundary was measured. Thus, the present study examines within-category discrimination at a much finer grain—at a resolution necessary to support statistical learning of the phonemic category structure and extraction of the phonological regularities of the language.

The present study differs from previous studies of categorical perception in a number of important ways. First, we used the head-turn-preference procedure (HTPP: Jusczyk & Aslin, 1995), a technique that has become a standard method for assessing word recognition during infancy. Although similar to habituation, the HTPP does not utilize as much initial exposure to the stimuli (e.g. 2 min rather than the 10 min or more used in High Amplitude Sucking or Conditioned Headturning). As a result, it is less susceptible to selective adaptation effects that can shift category boundaries (Samuel & Newport, 1979; Sawusch & Jusczyk, 1981).

Second, while many previous studies have employed synthetic stimuli, the present study used stimuli constructed from natural speech, which is more engaging to infants. Finally, rather than presenting infants with a single token and speech continuum, we used four pairs of words (as in other studies using multiple exemplars of natural speech: see Werker & Tees, 1984). This not only better maintains the infant's interest, but also introduces slight within-category acoustic variability, encouraging a category-based representation.

1. Experiment 1

Experiment 1 used the HTPP in a word memory paradigm. After exposure to items from a single category, listening time to *novel* items within and across category boundaries was assessed. If infants perceive and remember words categorically, their response to the novel, within-category tokens should be equivalent to their response to the originally exposed items. If infants show a graded response to within-category variation, their listening time should be midway between the original familiar items and the novel, cross-category tokens.

1.1. Methods

Participants. Ninety-five infants from monolingual English environments between the ages of 6.5 and 9 months ($M=8.0$ mon; $SD=0.60$) were tested in this experiment. Infants were randomly assigned to one of two exposure conditions: /b/ or /p/, with 35 infants in the /p/ condition and 60 in the /b/ condition.[2] Ten infants were excluded for failing to reach our minimum criterion of 2 s of listening time to 11 out of the 12 test tokens, leaving 85 infants in the final sample (52 for /b/, 33 for /p/).

Stimuli. Stimuli were taken from four 9-step b/p continua: beach/peach, bear/pear, bale/pail, and bomb/palm (originally created as part of another study: McMurray et al., in preparation). Each token was constructed from natural recordings of both endpoints made by a single male talked sampled at 11,025 kHz in a quiet room on a Kay Elemetrics Computerized Speech Lab (model 4300B). For each continuum, 7–10 exemplars of each of the two endpoints were recorded. From these tokens pairs of endpoints were selected that best matched each other on voice quality as well as first and second formant frequencies during the initial portion of the syllable.

Once b/p endpoints were selected for each of the four continua, they were normalized to have equivalent amplitudes (root-mean-square energy). Continua were then constructed using digital-editing software by progressively removing material from the onset of the voiced endpoint and replacing it with approximately the same amount of material from onset of the voiceless endpoint. Care was taken to only cut and splice at zero-crossings of the waveform (to avoid artifactual clicks). For each continuum, eight splice points were used at approximately 5 ms increments to create VOT continua ranging from approximately 3 ms (good exemplars of /b/) to 40 ms of VOT (good exemplars of /p/). In addition, 1–3 ms of the release burst of the voiceless tokens was spliced onto the endpoint voiced tokens, so that each continuum step would have a spliced portion.[3]

The stimuli selected from these continua for the present infant study were based on adult identification data (see McMurray et al., in preparation). Category boundaries for each continuum were estimated by fitting logistic functions to these identification data

[2] Because preliminary results suggested that within-category sensitivity on the voiced side of the continuum would have a smaller effect size, more infants were assigned to that condition. However, this proved to be unnecessary.

[3] For more information and examples of these continua, see http://www.psychology.uiowa.edu/faculty/mcmurray/publications/mcmurray_aslin_supplement

Table 1
Adult category boundaries and voice onset time (VOT) values (in ms) for stimuli used in Experiment 1

Item	Boundary	B	B*	P*	P
Bale	17.6	5.7	15.1	30.8	42.9
Beach	23.8	4.8	12.2	29.9	41.7
Bear	23.3	4.6	10.4	30.2	39.4
Bomb	18.15	4.0	9.7	29.7	38.7

(within-subjects). Means are reported in Table 1 along with VOTs of the tokens selected for the present infant study.

From these 9-step continua, endpoint tokens (B and P) were chosen to be as close to 5 ms and 40 ms (respectively) as possible. Tokens for the intermediate B* and P* conditions were selected to have at least a 5 ms difference in VOT from the endpoints, while maintaining at least 90% "correct" identification by the adult subjects (mean performance in Table 2).

Procedure. For each infant, either the voiced or voiceless endpoints of the four voicing continua were designated as targets, with the opposite endpoint designated the competitors. At the beginning of the experiment, infants were seated on their mother's lap in a dimly illuminated, sound-attenuated booth and exposed to 20 repetitions of each of the four target words in random order (400 ms inter-stimulus interval) while watching a children's movie on a small video screen. This 80-word familiarization phase lasted 75 s.

Immediately after familiarization, the video monitor was extinguished and testing began. On each trial, the infant was presented with repetitions of one of the tokens and listening time was assessed. Test tokens consisted of the same four target items, four novel competitor items, and four novel, within-category tokens (either B* or P*), yielding a total of 12 testing trials.

Each test trial started with a blinking red light immediately in front of the infant that served to center the infant's attention. When the infant looked at this light (determined by an observer watching the infant on a video monitor), it was extinguished and one of two lights to the left or right began blinking. When the observer indicated that the infant was looking at this side-light, a single auditory item was presented repeatedly until the infant looked away from the light for more than two consecutive seconds.

The total time spent listening to each of the 12 stimuli was automatically computed, based on the observer's responses. Trials in which total listening time was less than 2 s were automatically excluded and rerun at the conclusion of the 12 testing trials. If, after these additional trials, the infant did not listen for 2 s to at least 11 of the 12 tokens,

Table 2
Mean proportion correct identification from adults for B, B*, P* and P stimuli used in Experiment 1

Item	B	B*	P*	P
Bale	0.95	0.91	0.98	0.99
Beach	1.00	1.00	0.90	0.98
Bear	0.98	0.96	0.95	1.00
Bomb	1.00	1.00	1.00	1.00

Table 3
Number of infants in each exposure condition who showed a preference for the novel or familiar test tokens

Exposed to:	Novel	Familiar
B	36	16
P	21	12
Pre-B (exp 2)	25	34

the infant was excluded from further analysis. Throughout the procedure, the parent listened to music over headphones to prevent them from hearing the stimuli and covertly biasing their infant's responses.

1.2. Results and discussion

The logic of the HTPP is that the familiarization phase biases stimulus preferences. Unfortunately, preferences are free to vary in two directions, with listening times either longer (or shorter) to test items from the familiarization phase than to novel test items. Although there are a number of hypotheses as to what factors affect the direction of preferences (i.e. novelty or familiarity), no consensus has emerged, and few studies make a priori predictions (Aslin, 2000; Hunter & Ames, 1988; Wagner & Sakovits, 1986; but see Thiessen & Saffran, 2003).

Fortunately, conclusions from the present study rest on an analysis of *three* test conditions (e.g. B, B*, and P for infants familiarized to B). That is, the question of gradiency hinges on the *relative* performance on the B* and P* test items compared to the *two* endpoints. Thus, we sorted infants into two groups on the basis of their listening times to these endpoint stimuli. Infants who listened longer to the target stimuli presented during familiarization than to the novel competitors were classified as familiarity-preferring (FAM), and those who listened longer to the competitor stimuli were novelty-preferring (NOV). Note that this classification relies only on the target and competitor stimuli (B and P), not on the within-category stimuli (B* and P*). Table 3 reports the number of infants in each exposure condition classified as NOV or FAM.

After each infant was classified, data were analyzed, *within each group*, with respect to the target and competitor tokens. Recall that our hypothesis is that if infants show a gradient effect, listening times to within-category tokens should differ from *both* endpoints. A mixed design ANOVA with two factors was conducted separately for each of the two preference groups (NOV and FAM), with the target presented during familiarization as the between-subjects factor (B or P). The within-subjects factor was the three types of test trials: for infants exposed to /b/, B was designated Target, B* was Target* and P was the Competitor; for infants exposed to /p/, P was designated Target, P* was Target* and B was the Competitor.[4]

[4] Note that our sorting procedure guaranteed a main effect and linear trend, so these analyses were not conducted. It did not guarantee our predicted results, that Target* differed from *both* Target and Competitor. Therefore, these analyses were conducted as a series of planned comparisons within an ANOVA framework.

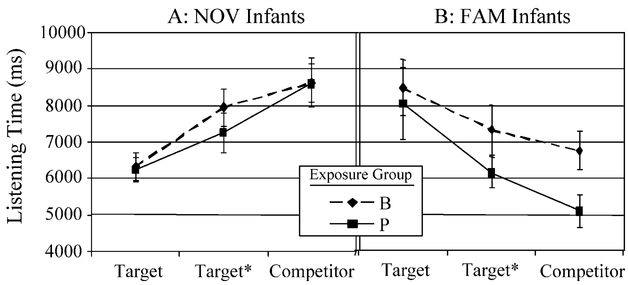

Fig. 1. Mean listening time to Target, Target* and Competitor as a function of exposure group. Panel A shows a clear gradient effect for infants classified as NOV. Panel B shows the same for infants classified as FAM. Error bars represent standard error of the mean.

NOV infants (Fig. 1, panel A) showed significant differences between Target and Target* ($F(1,55)=23.7$, $P<0.001$) and between Target* and Competitor ($F(1,55)=6.1$, $P=0.017$). Neither comparison interacted with the two exposure conditions (T vs. T*: $F(1,55)=2.0$, $P>0.1$; T* vs. C: $F(1,55)=2.7$, $P>0.1$). FAM infants (Fig. 1, panel B) showed similar results (despite the smaller sample size). Target* differed significantly from both Target ($F(1,26)=11.0$; $P=0.003$) and Competitor ($F(1,26)=7.3$; $P=0.012$). Neither interaction with exposure condition was significant (F's<1).

A series of planned comparisons was performed to verify within-category sensitivity in each exposure condition. Infants familiarized to /b/ and classified as NOV showed a significant difference between B and B* ($t(35)=4.9$, $P<0.0001$) but not between B* and P ($t(35)=0.6$; $P>0.1$). Infants who were classified as FAM showed a similar pattern, with a marginally significant difference between B and B* ($t(15)=2.0$, $P=0.062$) but not between B* and P ($t(15)=1.5$; $P>0.1$). Infants familiarized to /p/ exhibited much stronger effects. The NOV infants showed differences between P and P* ($t(20)=2.5$, $P=0.02$) and also between P* and B ($t(20)=2.9$, $P=0.009$). The FAM infants showed similar effects (P vs P*: $t(14)=2.42$, $P=0.03$; P* vs. B: $t(14)=2.62$; $P=0.02$).

To verify that sorting infants did not inflate the likelihood of obtaining gradient effects, a Monte Carlo simulation was conducted. We created a dataset that contained no effect of VOT but which generally matched the mean and variability associated with the experimental data. This simulated dataset consisted of each infant's mean listening times to the three types of test stimuli (B, B* and P or P, P* and B), but randomly shuffled (i.e. reassigned) to these three conditions. This shuffling disrupted any consistent relationship between VOT and listening time across subjects. As a result, when these shuffled datasets were analyzed using the same sorting procedure and statistical tests as the original data, evidence of gradiency should be extremely weak. If our NOV/FAM sorting procedure itself, rather than true differences in looking time, was the source of our gradient results, then the dominant pattern in this new (random) dataset should be the same—significant differences between B* or P* and the two endpoints. Five hundred random reshufflings (and analyses) revealed that the pattern of gradiency observed in our actual data was very rare in the "null"-datasets, occurring with a probability of .051 for NOV and .047 for FAM. Thus, our sorting procedure did not yield a greater likelihood of finding reliable effects than our predetermined alpha of 0.05.

2. Experiment 2

Overall, the results of Experiment 1 provided clear evidence for graded sensitivity to within-category differences. The significant differences between Target and Target*, and between Target* and Competitor, coupled with the lack of an interaction with VOT, suggest that gradiency extends to both sides of the category boundary. However, results from planned comparisons demonstrated that while infants exposed to /p/ showed clear evidence for graded sensitivity, this sensitivity was attenuated for the infants exposed to /b/. Contrary to the predicted null-hypothesis, however, it appeared that for these infants, responses to B* were more similar to P than to B. In a two-alternative-forced choice framework, this pattern would suggest that the infants in this condition may have a category boundary between B and B* rather than between B* and P. Although this contrasts with the adult data, we cannot be sure that infants' category boundaries are at the same location along the VOT continuum. Thus, in Experiment 2 we altered the VOT values of the B and B* tokens to shift them away from the adult category boundary.

2.1. Method

Participants. Sixty-nine infants raised in a monolingual English environment between the ages of 6.5 and 8.5 months ($M=8.15$ mos. $SD=0.49$) were tested. None of them participated in Experiment 1. Ten infants were excluded for failing to reach our criterion of 2 s of listening time to 11 of the 12 test tokens, leaving 59 infants in the final sample.

Stimuli. The target B tokens were shifted from a mean VOT of +4.8 ms in Experiment 1 to −9.8 ms in Experiment 2 (although in English, such prevoicing is still robustly judged as exemplars of /b/). The prevoiced tokens were constructed from recordings of the same speaker in Experiment 1 by taking a single voicing-pulse from the onset of a prevoiced exemplar and splicing it to the onset of the original voiced tokens. The B* tokens in Experiment 2 were identical to the B tokens in Experiment 1, with a mean VOT of +4.8 ms. Competitor (P) tokens were the same as in Experiment 1 (VOT = 40 ms).

Procedure. An identical procedure to that used in Experiment 1 was used to assess listening times after familiarization. Infants were exposed to 20 repetitions each of the four prevoiced B words, followed by 12 testing trials with B, B* and P words.

2.2. Results and discussion

As in Experiment 1, infants were classified as NOV or FAM with 25 NOV infants and 34 FAM infants. Infants in the NOV (Fig. 2A) group showed significant differences in listening times between B and B* ($t(24)=3.73$, $P=0.001$) and between B* and

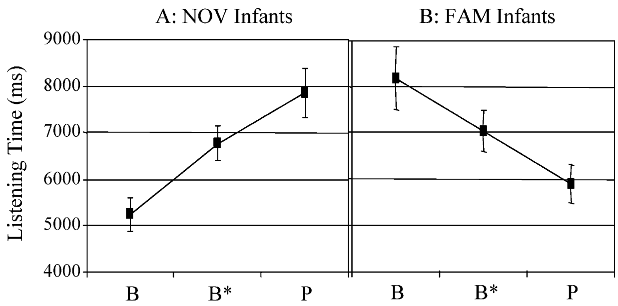

Fig. 2. Mean listening time to B, B* and P for infants exposed to prevoiced stimuli. Panel A shows a clear gradient effect for infants classified as NOV. Panel B shows the same for infants classified as FAM. Error bars represent standard error of the mean.

P ($t(24)=2.5$, $P=0.02$). Infants in the FAM group (Fig. 2B) showed a similar pattern with significant differences between B and B* ($t(33)=2.5$, $P=0.016$) and a marginally significant difference between B* and P ($t(33)=2.0$, $P=0.057$).

These results are consistent with gradient categories for 8-month-old infants, and suggest that the attenuation of the gradiency seen in the voiced stimuli from Experiment 1 may have been due to the fact that the contrasts we chose were located suboptimally with respect to the location of the voiced category. An alternative view, however, is that by measuring responses to *prevoiced* stimuli, we tapped a non-native category that had not yet been "lost" (e.g. Aslin et al., 1981). Under this view, our B stimuli activated this prevoiced category, while the B* activated the "native" voiced category. While a strictly categorical version of this explanation would not yield the graded results seen here (since B* and P are equally non-B), category representations with significant overlap (and gradiency *outside* the prototypical range) would yield such results since B would be more confusable with B* than P. This explanation is broadly consistent with gradient categories, but does not require gradations *within-categories*. However, it does not seem likely, given the small VOT difference tested here (14.5 ms). Moreover, it cannot explain the graded effect found on the *voiceless* side of the continuum in Experiment 1, since there is no reason to think that VOTs of 30 and 40 ms are in different categories.

3. General discussion

The experiments presented here support the hypothesis that infants are sensitive to within-category variation in stop consonants. Our results provide clear evidence for graded representations for both voiced and voiceless speech categories. The attenuation of this effect for voiced consonants in Experiment 1 can be explained in at least two ways. As mentioned earlier, it is possible that the infants categorized the B* stimuli as /p/ which would posit a category boundary somewhere between 4.8 and 11.8 ms. This boundary is quite small and inconsistent with prior work on VOT in infants (e.g. Aslin et al., 1981; Eimas et al., 1971).

The alternative explanation rests on the fact that the HTPP is a one-alternative task. Infants in this procedure are not "deciding" whether a stimulus is B or P, but rather whether the stimulus is B or *not-B*. Under this view, B* may be sufficiently far from B that it is not a part of that category, while still not being a part of the P category. Infants' early phonetic categories may reside in dimensional "islands", with gaps near category boundaries (and possibly at extreme values along a dimension as well) where no category is strongly activated.[5] Such a representation may in fact be adaptive—by withholding judgment at certain regions of the acoustic-phonetic space, the infant may in fact be preserving plasticity until more data are obtained from native-language input. Work in progress is using a two-alternative task, the Anticipatory Eye Movement (AEM) procedure (McMurray & Aslin, 2004), to further examine this hypothesis.

We acknowledge that our claim of graded representations, like all other data collected on infant speech perception, is based on between-subject averages. It is possible that the graded sensitivity we see across infants is actually the result of categorical representations within-subjects, (with highly variable category boundaries between subjects). We do not believe this alternative is likely because the medial tokens (B* and P*) were quite far from expected boundaries; this extreme boundary-variability would contrast with prior reports of consistent boundaries during infancy. Moreover, this alternative cannot be ruled out by any current infant methodology (e.g. HTPP, High Amplitude Sucking, Habituation, or Conditioned Head-turning) because none of these methods yields sufficient repeated measurements within individual infants.

The present study bridges several areas of research in speech perception. We have demonstrated that infant voicing categories have graded sensitivity to subphonemic variation. This within-category sensitivity bears on theories of underlying phonological representations in early language acquisition and provides a necessary mechanism for statistical accounts of phonological learning. These results also link infant abilities with a growing body of work with adults demonstrating sensitivity to and use of within-category detail during word recognition.

Acknowledgements

The authors would like to thank Julie Markant for recruiting and running the participants in this study, Katherine Crosswhite who provided the tools for amplitude normalization, and Richard Pastore for helpful discussions leading to experiment 2. This study was supported by a NIH predoctoral NRSA (F31DC006537-01) to BM and a NIH research grant (HD-37082) to RNA.

[5] Such gaps may also map to non-native categories, or more likely, their remnants as they are lost through language exposure.

References

Allen, J. S., & Miller, J. L. (1999). Effects of syllable-initial voicing and speaking rate on the temporal characteristics of monosyllabic words. *Journal of the Acoustical Society of America*, *106*(4), 2031–2039.

Aslin, R. N. (2000, July). Interpretation of infant listening times using the headturn preference technique. Paper presented at the *International Conference on Infant Studies*, Brighton, UK.

Aslin, R. N., Pisoni, D. B., Hennessy, B. I., & Perey, A. J. (1981). Discrimination of voice onset time by human infants: new findings and implications for the effects of early experience. *Child Development*, *52*, 1135–1145.

Carney, A. E., Widin, G. P., & Viemeister, N. F. (1977). Non categorical perception of stop consonants differing in VOT. *Journal of the Acoustical Society of America*, *62*, 961–970.

Clark, M. J., & Hillenbrand, J. M. (2003). Quality Of American English front vowels before /R/. *Journal of the International Phonetic Association*, *33*(1), 1–16.

Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. (2001). Subcategorical mismatches and the time course of lexical access: evidence for lexical competition. *Language and Cognitive Processes*, *16*(5/6), 507–534.

Eimas, P. (1974). Auditory and linguistic processing of cues for place of articulation by infants. *Perception and Psychophysics*, *16*(3), 513–521.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, *171*, 303–306.

Fougeron, C., & Keating, P. (1997). Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America*, *101*, 3628–3740.

Gow, D. (2001). Assimilation and anticipation in continuous spoken word recognition. *Journal of Memory and Language*, *45*, 133–139.

Gow, D., & Gordon, P. C. (1995). Lexical and prelexical influences on word segmentation: evidence from priming. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(2), 344–359.

Gow, D., McMurray, B., & Tanenhaus, M. K. (2003, November). Eye movements reveal the time course of multiple context effects in the perception of assimilated speech. Poster presented at *The 44th Annual Meeting of the Psychonomics Society*, Vancouver, Canada.

Guenther, F., & Gjaja, M. (1996). The perceptual magnet effect as an emergent property of neural map formation. *Journal of the Acoustical Society of America*, *100*, 1111–1112.

Hohne, E., & Jusczyk, P. (1994). Two-month-old infants' sensitivity to allophonic differences. *Perception and Psychophysics*, *56*(6), 613–623.

Hunter, M. A., & Ames, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. *Advances in Infancy Research*, *5*, 69–95.

Jusczyk, P. (1997). *The discovery of spoken language*. Cambridge, MA: MIT Press.

Jusczyk, P., & Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology*, *29*, 1–23.

Juscyk, P., Hohne, E., & Bauman, A. (1999). Infants' sensitivity to allophonic cues for word segmentation. *Perception and Psychophysics*, *61*(8), 1465–14767.

Kessinger, R. H., & Blumstein, S. E. (1998). Effects of speaking rate on voice onset time and vowel production: some implications for perception studies. *Journal of Phonetics*, *26*, 117–128.

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*(5), 358–368.

Magnuson, J. S., McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2003). Lexical effects on compensation for coarticulation: the ghost of Christmash past. *Cognitive Science*, *27*(2), 285–298.

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

McClelland, J., & Elman, J. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1–86.

McMurray, B., & Aslin, R. N. (2004). Anticipatory eye-movements reveal infants' auditory and visual categories. *Infancy*, *6*(2), 203–229.

McMurray, B., Aslin, R. N., Tanenhaus, M. K., Spivey, M. J, & Subik, D. (in preparation). *Two B or not Two B: categorical perception in lexical and nonlexical tasks*.

McMurray, B., Tanenhaus, M., & Aslin, R. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition*, *86*(2), B33–B42.

Miller, J. L. (1997). Internal structure of phonetic categories. *Language and Cognitive Processes*, *12*, 865–869.

Miller, J., & Eimas, P. (1996). Internal structure of voicing categories in early infancy. *Perception and Psychophysics*, *58*(8), 1157–1167.

Pierrehumbert, J. (2004). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech*, *46*(2/3), 115–154.

Pisoni, D. B., & Lazarus, J. H. (1974). Categorical and noncategorical modes of speech perception along the voicing continuum. *Journal of the Acoustical Society of America*, *55*(2), 328–333.

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

Samuel, A., & Newport, E. (1979). Adaptation of speech by nonspeech: evidence for complex acoustic cue detectors. *Journal of Experimental Psychology: Human Perception and Performance*, *5*(3), 546–578.

Samuel, A., & Pitt, M. (2003). Lexical activation (and other factors) can mediate compensation for coarticulation. *Journal of Memory and Language*, *48*, 416–434.

Sawusch, J., & Jusczyk, P. (1981). Adaptation and contrast in the perception of voicing. *Journal of Experimental Psychology: Human Perception and Performance*, *7*(2), 408–421.

Thiessen, E. D., & Saffran, J. R. (2003). When cues collide: use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology*, *39*(4), 706–716.

Wagner, S., & Sakovits, J. (1986). A process analyses of infant visual and cross-modal recognition memory: Implications for an amodal code. In L. Lipsitt, & C. Rovee-Collier, *Advances in infancy research* (Vol. 4), 196–217.

Werker, J. F., & Tees, R. (1984). Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*, 49–63.