Perceptual magnet and phoneme boundary effects in speech perception: Do they arise from a common mechanism?

PAUL IVERSON and PATRICIA K. KUHL
University of Washington, Seattle, Washington

The question of whether sensitivity peaks at vowel boundaries (i.e., phoneme boundary effects) and sensitivity minima near excellent category exemplars (i.e., perceptual magnet effects) stem from the same stage of perceptual processing was examined in two experiments. In Experiment 1, participants gave phoneme identification and goodness ratings for 13 synthesized English /i/ and /e/ vowels. In Experiment 2, participants discriminated pairs of these vowels. Either the listeners discriminated the entire range of stimuli within each block of trials, or the range within each block was restricted to a single stimulus pair. In addition, listeners discriminated either one-step or two-step intervals along the stimulus series. The results demonstrated that sensitivity peaks at vowel boundaries were more influenced by stimulus range than were perceptual magnet effects; peaks in sensitivity near the /i/–/e/ boundary were reduced with restricted stimulus ranges and one-step intervals, but minima in discrimination near the best exemplars of /i/ were present in all conditions.

Speech perception is marked by distortions in sensitivity that are more easily linked to phonemic categorization than to auditory processing. When listeners hear items from a stimulus continuum that spans two phoneme categories and has acoustically uniform steps, their sensitivity is rarely uniform. There are peaks in sensitivity at category boundaries and poor sensitivity within phoneme categories (Fujisaki & Kawashima, 1970; Liberman, Harris, Hoffman, & Griffith, 1957; Repp, 1984; Studdert-Kennedy, Liberman, Harris, & Cooper, 1970), reaching minima near the best exemplars of phoneme categories (Aaltonen, Eerola, Hellström, Uusipaikka, & Lang, 1997; Iverson & Kuhl, 1995, 1996; Kuhl, 1991; Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Sussman & Lauckner-Morano, 1995; cf. Lively & Pisoni, 1997, and Lotto, Kluender, & Holt, 1998). It is as if the perceptual space underlying speech sounds is distorted, being stretched at category boundaries and shrunk near best exemplars.

Although the link between phonemic status and sensitivity could be seen to imply that perceptual distortions are caused by phonemic categorization, it is important to distinguish between the perceptual phenomena of the distortions and the mechanisms that have been hypothesized as their cause. For example, the term *categorical*

perception is mostly used as an empirical description of the correlations between discrimination sensitivity and phoneme labeling in perceptual experiments, but it is strongly linked with a class of hypothesized mechanisms in which phonemes are perceived in terms of their phonemic labels rather than by their acoustic properties (see Repp, 1984, for a review). Researchers have disputed whether the perceptual phenomena associated with categorical perception (particularly the sensitivity peaks at phonemic boundaries) are actually due to a phonemic labeling mechanism, attributing them instead to such factors as nonlinearities in auditory processing (e.g., Diehl, 1987; Kuhl, 1987, 1988) or cognitive encoding strategies (e.g., Macmillan, Goldberg, & Braida, 1988). However, sensitivity peaks at phoneme identification boundaries remain ubiquitous in speech perception research, despite disputes about their origin. The present paper will use the term *phoneme boundary effect* (Wood, 1976) for this perceptual phenomenon, to distinguish it from hypothesized mechanisms.

The perceptual phenomenon of the occurrence of poor sensitivity near best exemplars has been called the *perceptual magnet effect* (Kuhl, 1991); the perceptual space is shrunk near best exemplars, effectively drawing neighboring stimuli toward best-exemplar locations. As with the phoneme boundary effect, the perceptual magnet effect has been strongly linked to a hypothesized mechanism; the perceptual magnet effect has been seen to imply that phoneme categories are represented in terms of prototypes (i.e., a single abstract exemplar that represents all the members of a category) and that phonemes are at least partially perceived in terms of their distance from these prototypes. However, the perceptual magnet effect may, in fact, be due to other mechanisms. For instance, it could

arise because of experience-related distortions in auditory processing (Guenther & Gjaja, 1996) or because of categorization processes based on multiple stored exemplars (Lacerda, 1995). The term *perceptual magnet effect* refers to the perceptual phenomenon of sensitivity minima near best exemplars, and the exact cause of the perceptual magnet effect is unknown.

Empirical Distinctions

Research on the perceptual magnet effect has primarily been concerned with establishing the correlation between locations of best exemplars and poor sensitivity. For example, Kuhl (1991) synthesized an excellent exemplar of the American English vowel /i/ (the prototype, P) and a poor exemplar that was still thought to be consistently identified as a member of the /i/ category (the nonprototype, NP). Variants in the region of these two vowels were synthesized by varying $F1$ and $F2$ frequencies. Adults rated the goodness of all the stimuli on a 7-point integer scale, and the results confirmed that P and its variants were perceived to be better instances of the category than were NP and its variants. Sensitivity was measured for adults, 6-month-old infants, and rhesus monkeys, using a go/no-go discrimination procedure in which one of the target stimuli (P or NP) was played repeatedly in the background and was occasionally changed to one of its variants. The results revealed that adults and infants were better at detecting changes in NP than in P, demonstrating that they were less sensitive to acoustic changes in the region of the best exemplar. Monkeys had equivalent detection performances for P and NP, suggesting that the sensitivity exhibited by humans was due to linguistic experience (see Kuhl et al., 1992, for supporting evidence).

One of the earliest criticisms of the Kuhl (1991) study was that it did not effectively eliminate the possibility that the perceptual magnet effect was the same as the phoneme boundary effect. That is, if the NP stimulus was located near a phoneme boundary—and was not constantly labeled as /i/—the increased sensitivity for NP could have been due to the increased sensitivity at the category boundary previously described by the phoneme boundary effect. Kuhl (1991) did not explicitly ask the participants to make phoneme identifications, so it was not certain whether NP was identified consistently as /i/. Moreover, additional studies that obtained identification judgments revealed that stimuli near NP are perceived by some listeners to be outside of the /i/ category (Iverson & Kuhl, 1995; Lively & Pisoni, 1997; Sussman & Lauckner-Morano, 1995). To address this issue, Iverson and Kuhl (1995) examined the link between goodness, identification, and sensitivity within the /i/ category, using multidimensional scaling and detection-theoretic measures. Although the participants in this experiment thought that the Kuhl (1991) NP was outside of the /i/ category, the perceptual magnet effect was found even when analysis was restricted to stimuli that were clearly within the /i/ cate-

gory. Therefore, the perceptual magnet effect seems distinct from distortion at the category boundary.

Recent work by Lotto et al. (1998) adds a caveat to this conclusion. Phonemic-labeling tasks typically involve the identification of individual stimuli, but Lotto et al. examined the identification of stimuli presented in pairs, identical to the way stimuli are presented in AX discrimination experiments (see also Repp, Healy, & Crowder, 1979). In tests using stimuli analogous to those used by Iverson and Kuhl (1995), they found that stimuli were identified less often as /i/ when they were paired with P than when they were paired with NP (i.e., there was a contrast effect, so that presenting stimuli paired with P shifted the identification boundary toward P). In discrimination experiments, the patterns of sensitivity replicated those of Iverson and Kuhl (1995), but Lotto et al. attributed these sensitivity patterns to the context-related shifts in the phonemic boundary. Specifically, poorest sensitivity was found for the best exemplars of /i/ because these tokens were most consistently identified as /i/ when presented in the P context. Higher sensitivity was found for exemplars that received low goodness ratings and consistent /i/ identifications in Iverson and Kuhl's (1995) study, because these tokens became ambiguous in terms of identification when presented in the P context (identifications in the NP context obtained by Lotto et al. matched the single-interval identifications obtained by Iverson & Kuhl, 1995). Given this result, it seems possible that the effect of goodness on discrimination could be mediated by phonemic labeling. That is, the goodness of tokens in pairs could influence phonemic identifications (i.e., the best exemplars shift identification boundaries), but phonemic boundary locations could directly be responsible for the sensitivity minima observed near P.

However, the perceptual magnet effect has also been found for consonants (Davis & Kuhl, 1994; Iverson & Kuhl, 1996), whose identification functions are much less prone to context-related boundary shifts than are those of vowels. Furthermore, in Iverson and Kuhl's (1996) two-dimensional mapping of the American English /r/ and /l/ categories, distortions occurred that were orthogonal to those predicted by phoneme identification. Specifically, the perceptual space was shrunk along both stimulus dimensions ($F2$ and $F3$) near the best exemplars of /r/ and /l/, in accord with the perceptual magnet effect. But the identification of these stimuli was a function only of $F3$. The distortion along the $F2$ dimension cannot be described by the phoneme boundary effect, because the $F2$ values of stimuli did not affect their proximity to the phoneme boundary. It seems doubtful that the effects of goodness on discrimination were mediated by phonemic labeling (see Lotto et al., 1998) in this case.

Causes of Phoneme Boundary and Perceptual Magnet Effects

Although the phenomena of phoneme boundary and perceptual magnet effects may be distinct, the possibility

remains that these effects arise from the same underlying cause. Many causes currently seem plausible, including processes related to phonemic categorization, nonlinearities in auditory processing, and cognitive encoding strategies. A few possible mechanisms are described below.

Phonemic encoding. Categorical perception provided the earliest evidence that speech perception involves a phonemic encoding stage in which the perceptual input is represented in terms of discrete phonemic category labels (Fujisaki & Kawashima, 1970; Liberman et al., 1957; Studdert-Kennedy et al., 1970). Perception is in terms of the input's phonemic identity, although it may be possible to retain some residual perceptual information. This hypothesis has been discredited by diverse evidence from sources such as animal studies (e.g., Kuhl & Miller, 1975, 1978) and nonspeech studies with humans (e.g., Kluender, Diehl, & Wright, 1988). Most significant for the present discussion, the perceptual magnet effect itself provides arguments against a discrete phonemic encoding stage. Categorical perception requires stimuli that receive the same phonemic label to be perceptually identical, but listeners are able to judge which stimuli are the best exemplars of a phonemic category even when they are labeled identically (Miller & Volaitis, 1989; Samuel, 1982; Volaitis & Miller, 1992). Furthermore, sensitivity to stimuli that receive the same phonemic label is not uniformly poor; sensitivity is highest for poor exemplars of a category and lowest for excellent exemplars (Aaltonen et al., 1997; Iverson & Kuhl, 1995, 1996; Kuhl, 1991; Kuhl et al., 1992; Sussman & Lauckner-Morano, 1995). Listeners apparently encode more information from speech sounds than discrete phonemic labels.

Although this makes it seem unlikely that phonetic processing involves a categorical phonemic encoding stage, it is still plausible that phoneme boundary and perceptual magnet effects arise from phonemic encoding of a different form. For example, it is possible that stimuli are encoded in terms of their distance from prototypes that represent phoneme categories. One key component of such a model is that this distance is graded. In other words, rather than being perceived in terms of discrete category membership, phonemes could be perceived along continuous dimensions of how closely they represent each phoneme in an individual's language. It is possible that this encoding process could somehow distort perception, stretching distances in regions in which stimuli are equidistant between prototypes (i.e., at phonemic boundaries) and shrinking distances near prototypes.

It can be argued that the notion of prototype representations is implausible, because it is hard to conceive of how individual prototypes would be able to represent the varied acoustics of speech produced by different talkers under different conditions. It is probably more plausible that categories could be represented in terms of multiple exemplars produced by many talkers. Such a model has been detailed by Lacerda (1995). Rather than each stimulus being encoded in terms of its distance from an individual prototype, it could be encoded in terms of its aggregate distance from all of the exemplars of a category. Goodness would be maximized when the distance from the exemplars is minimized, and category boundaries would occur at locations equally distant from the exemplars of two or more categories. A more general point is that exemplar representations, and even abstract representations, can have central tendencies that mimic prototypes (i.e., certain stimuli can best match category representations, even if the categories are not represented in terms of prototypes). Thus, it is difficult to distinguish between the empirical predictions of prototype- and exemplar-based models (Estes, 1993). The perceptual magnet effect could arise from any type of phonemic encoding stage that can account for variations in goodness.

Auditory processing. Probably the most convincing evidence against phonemic categorization explanations of the phoneme boundary effect is that phoneme boundary effects occur under conditions in which phonemic categorization cannot plausibly play a role. For example, chinchillas (Kuhl & Miller, 1975, 1978), monkeys (Kuhl & Padden, 1982, 1983), and quail (Kluender, 1991) have all been shown to exhibit phoneme boundary effects for consonant stimuli. In addition, humans can show boundary effects for nonspeech stimuli (e.g., Kluender et al., 1988). Phoneme boundary effects may not arise because of phonemic categorization. Rather, these regions of increased sensitivity may arise because of general, not speech-specific, characteristics of auditory processing. Phonemic boundaries and sensitivity peaks may coincide because language systems evolve to enhance intelligibility (see Diehl, Kluender, Walsh, & Parker, 1991). Thus, phonemic boundaries may tend to be located in positions that take advantage of existing regions of high perceptual sensitivity.

One of the problems with such auditory theories is that they have difficulty accounting for perceptual changes related to experience. Kuhl's native language magnet theory of speech perception (Kuhl, 1992, 1993a, 1993b) solves this problem by suggesting that speech perception is initially determined by nonlinearities in auditory perception but that perception is altered by the acquisition of phoneme categories. Perception becomes fine-tuned to the characteristics of one's native language. The perceptual magnet effect is hypothesized to be central to this tuning process, because monkeys do not exhibit this effect (Kuhl, 1991) and its realization in humans has been linked to linguistic experience (Kuhl et al., 1992).

Although it is easiest to think of the perceptual magnet effect as being related to cognitive aspects of phonemic categorization, Guenther and Gjaja (1996) have recently proposed a way in which it could be explained by the tuning of auditory neural maps that are sensitive to the acoustic dimensions of speech. For example, it is proposed that, given a set of neurons that are sensitive to $F1$ and $F2$

formant frequencies of vowels, the initial (i.e., prior to sensory experience) firing preferences of the individual neurons are randomly distributed within the $F1$–$F2$ space. Sensory experience alters the firing preferences of these neurons, so that more neurons become tuned to the formants of frequently experienced vowels and fewer neurons become tuned to the formants of infrequently experienced vowels. The distribution of firing preferences begins to reflect the nonuniform distribution of experienced vowels. Under this theory, the perceptual magnet effect is caused by these nonuniform firing preferences. Any input vowel will cause a population of neurons within this map to fire, on the basis of the similarity of the vowel's formants to the firing preferences of the neurons. However, considering that the distribution of firing preferences is not uniform, more of the neurons that fire will be in the direction of the prototype (i.e., the center of the distribution of experienced stimuli) than in directions away from the prototype (i.e., less frequent stimuli). The center of the population of neurons that fire for a given stimulus will thus be shifted toward prototypic locations, because more neurons are tuned to prototypic stimuli. The projection of vowels into these neural maps will cause vowels to be perceived as being closer to the prototype than they really are. In computer simulations of this model, Guenther and Gjaja were able to closely reproduce Kuhl's (1991) and Iverson and Kuhl's (1996) experimental data.

The most important implication of this model is that the distribution of experienced speech can plausibly alter processing at an auditory, noncognitive level. Introspective notions of phonemic goodness are probably also related to the distribution of experienced speech. That is, listeners begin to perceive frequently presented stimuli and stimuli far from phonemic boundaries as being good exemplars. However, it is possible that the locations of best exemplars and the center of the perceptual magnet effect are only coincidentally related. That is, they tend to occur at the same locations because both are related to the distribution of sounds in the ambient language, but best-exemplar locations and perceptual distortions could be determined by entirely different stages of neural processing.

Perceptual anchors. Although the perception of consonants has often met the criteria for categorical perception, the perception of vowels has often been described as being more continuous (e.g., Fry, Abramson, Eimas, & Liberman, 1962). Vowels do not meet the strict criteria for categorical perception, because sensitivity to vowel stimuli within phoneme categories is higher than is predicted by phonemic labeling. However, vowels have been found in most studies to have phonemic boundary effects. The presence of phonemic boundary effects has been found to vary with experimental task. Specifically, experimental tasks that reduce demands on attention and memory diminish sensitivity peaks for vowel stimuli (Macmillan et al., 1988; Pisoni, 1973, 1975; Repp et al., 1979).

Macmillan et al. (1988) have adapted Durlach and Braida's (1969) model of intensity discrimination to pro-

vide a detection-theoretic account of why vowel boundary peaks are influenced by experimental task. According to their theory, the accuracy of listeners in discrimination tasks is limited by three main types of variance: (1) sensory variance owing to resolution limits of the auditory system, (2) context variance owing to the range of stimuli presented in each block of trials (limiting the ability to label individual stimuli), and (3) trace variance owing to the length of time that each stimulus is stored in memory. In tasks in which there is low context and trace variance, sensitivity peaks at vowel boundaries are eliminated, because performance reflects the basic resolution of the auditory system. In tasks in which there is a high degree of context variance, sensitivity peaks at vowel boundaries emerge, owing to strategies that listeners adopt to compensate for their labeling difficulty. Listeners in these tasks tend to label stimuli in terms of their distance from a particular reference point along the stimulus series (a *perceptual anchor*). These distance estimates become less accurate as distances increase, so this causes relatively high discrimination performance near perceptual anchors. Boundary effects for vowels occur when there is a high degree of context variance, because listeners tend to choose perceptual anchors that are near phonemic boundaries (Macmillan et al., 1988). It should be noted that increases in trace variance also may promote vowel boundary effects (Pisoni, 1973), but this receives less emphasis in Macmillan et al.'s (1988) model.

Thus, Macmillan et al.'s (1988) perceptual anchor model provides an alternative account of vowel boundary effects: They arise from the strategy of encoding stimuli in terms of perceptual anchors. Perceptual anchors do not account for consonant boundary effects, because these effects are found even when context and trace variance is minimized (Macmillan et al., 1988). Within Macmillan et al.'s (1988) model, consonant boundary effects are, therefore, attributed to the resolution of the auditory system.

It is plausible that the perceptual magnet effect can be explained by encoding in terms of perceptual anchors. That is, regions of low sensitivity within vowel categories may occur because they are far from perceptual anchors, rather than because they are near best exemplars of phonemic categories.

Perceptual magnet effects have been found both for vowels (Aaltonen et al., 1997; Iverson & Kuhl, 1995; Kuhl, 1991; Kuhl et al., 1992; Sussman & Lauckner-Morano, 1995) and for consonants (Davis & Kuhl, 1994; Iverson & Kuhl, 1996), but it is unknown whether the strength of the perceptual magnet effect is altered by experimental procedures. Previous tests of the perceptual magnet effect for vowels have employed procedures with a large degree of context variance. The effect was originally identified (Grieser & Kuhl, 1989; Kuhl, 1991; Kuhl et al., 1992) with a go/no-go discrimination design that employed fixed standards (P and NP) in separate blocks of trials, but these standards were paired with 64 comparison

stimuli that varied on two independent dimensions (F1 and F2 frequency). Iverson and Kuhl (1995) examined the perceptual magnet effect with a roving discrimination design in which there was no fixed standard and tokens from the entire stimulus series were presented within each block of trials. Examinations of the perceptual magnet effects for consonants (Davis & Kuhl, 1994; Iverson & Kuhl, 1996) have employed similarity scaling designs in which every possible stimulus pair within the stimulus set is presented to participants in a randomized order. It is, thus, unknown whether the perceptual magnet effect is reduced by experimental procedures that reduce context variance.

An Investigation of Mechanisms

In the present investigation, context variance in discrimination tasks for vowels was manipulated in order to examine whether perceptual magnet and phoneme boundary effects arise from common mechanisms. Phoneme boundary effects are eliminated under conditions with reduced context variance, but it is unknown whether the perceptual magnet effect is eliminated under the same conditions. If reducing context variance diminishes phoneme boundary effects more than it does perceptual magnet effects, this would provide strong evidence that these phenomena arise from different perceptual processes. Within Macmillan et al.'s (1988) framework, this result would suggest that the perceptual magnet effect arises from auditory processing, because performance under reduced context variance is thought to reflect sensory resolution. It is also conceivable that this result could be attributed to phonemic processing, but it would suggest that this particular phonemic processing stage must be relatively immune to effects of focused attention (see Schouten & van Hessen, 1992). If reducing context variance has similar effects on phoneme boundary and perceptual magnet effects, this would support the conclusion that phoneme boundary and perceptual magnet effects for vowels both arise from cognitive encoding strategies. It would then be left for future research to examine whether perceptual magnet effects for consonants are similarly diminished under conditions with reduced context variance.

The experiments employed the stimulus set used by Iverson and Kuhl (1995; Kuhl, 1991) and included exemplars of the vowel categories /i/ and /e/. The stimuli varied in F1 and F2 frequency in 30-mel steps along a single stimulus vector. In Experiment 1, the listeners identified and rated the goodness of each stimulus in terms of the /i/ and /e/ phoneme categories. In Experiment 2, the listeners discriminated pairs of these stimuli (AX discrimination) in four conditions: 30-mel roving, 30-mel fixed, 60-mel roving, and 60-mel fixed. The fixed and roving conditions differed in context variance. Roving conditions presented pairs of tokens from the entire stimulus set within each block of trials and, therefore, had high context variance. Fixed conditions presented only a single pair of

tokens within each block of trials and, therefore, had low context variance. The 30- and 60-mel conditions differed in the acoustic interval between each pair of stimuli; the participants in the 60-mel conditions discriminated acoustic differences that were twice as large as those in the 30-mel conditions. This manipulation tested whether the perceptual magnet effect was present for stimulus intervals that were near threshold (30 mels), as well as for differences that were twice as large (60 mels). Although this was not motivated by Macmillan et al.'s (1988) model, manipulations of this sort have been used by others (Healy & Repp, 1982; Pisoni, 1975) to assess vowel discrimination. Small and large step sizes both exhibit phoneme boundary effects. It is unknown whether the perceptual magnet effect is apparent for both threshold and suprathreshold intervals.

EXPERIMENT 1
Identification and Goodness

In this experiment, the listeners identified each stimulus as /i/ or /e/ and rated its goodness on a scale from 1 (bad) to 7 (good). The aim of this study was to establish the location of the best /i/ exemplar and the location of the /i/–/e/ identification boundary. These locations were determined for comparison with the discrimination results of Experiment 2.

Method

Participants. Forty-nine members of the University of Washington community participated in this experiment. Twenty-seven participants completed the 60-mel condition, and 22 participants completed the 30-mel condition. All of the participants were native English speakers, reported having no known hearing impairments, had no training in phonetics, and were between 18 and 40 years of age. They received course credit for their participation.

Apparatus. The stimuli were reproduced at 10,000 twelve-bit samples per second by a Data Translation DT2821 digital I/O board and were low-pass filtered with a 4.6-kHz cutoff frequency. The participants heard these sounds through the right ear of Telephonics TDH-39P headphones while each participant sat in a sound-treated booth. An NEC 386 microcomputer controlled the stimulus presentation and recorded responses.

Stimuli. Thirteen /i/ and /e/ vowel stimuli were synthesized, using the Klatt (1980) cascade-parallel speech synthesizer. These stimuli were identical to those used by Iverson and Kuhl (1995) and were a subset of the 64 stimuli used by Kuhl (1991). The tokens varied in F1 and F2 frequency along a one-dimensional stimulus series, with a combined 30-mel difference[1] between neighboring tokens (formant frequencies are listed in Table 1). Equal stimulus spacing along the mel scale (Stevens, Volkmann, & Newman, 1937) was designed to equate the stimulus differences with regard to peripheral auditory processing, as was recommended by Fant (1973). The mel scale tends to equate the thresholds necessary to detect differences in vowel formant frequencies (Fant, 1973) and the width of critical bands at most frequencies (Zwicker, Flottorp, & Stevens, 1957).

The stimuli were identical in all other respects. The formant frequencies were set to 3010 Hz for F3, 3300 Hz for F4, and 3850 Hz for F5. The bandwidths were set to 53 Hz for F1, 77 Hz for F2,

Table 1
Stimulus Formant Frequencies (in Hertz)

Stimulus No.	F1	F2
1*	197	2489
2	215	2438
3*	233	2388
4	251	2339
5*	270	2290
6	289	2242
7*	308	2195
8	327	2148
9*	347	2102
10	367	2057
11*	387	2012
12	408	1968
13*	429	1925

*Stimuli heard by the participants in the 60-mel condition.

111 Hz for F3, 175 Hz for F4, and 281 Hz for F5. The F0 contour of each token rose from 112 to 130 Hz over the first 100 msec and dropped to 92 Hz over the rest of the stimulus. Each token was 435 msec long. The stimuli were equalized in root-mean-square (RMS) intensity and were played to participants at a comfortable level.

Procedure. The participants gave identification and goodness ratings for the stimuli that they would discriminate in Experiment 2. The participants in the 30-mel condition gave ratings on all 13 stimuli. The participants in the 60-mel condition gave ratings only on the 7 odd-numbered stimuli (see Table 1).

The task was identical for both conditions. A single token was played on each trial, and the participants made identification and goodness judgments. For the identification task, the participants judged whether the token sounded like the vowel in *he* (/i/), like that in *hay* (/e/), or like neither of those vowels. After making this judgment, the participants rated the degree to which the token was a good example of the identified category by giving an integer rating from 1 (*bad*) to 7 (*good*). Goodness ratings were not collected if the participant judged that the token was not /i/ or /e/. They were instructed to distribute their responses so that the worst category exemplars in the set would receive the lowest ratings, the best category exemplars in the set would receive the highest ratings, and the other tokens would receive intermediate ratings. The participants were allowed to hear each token as often as they needed to make their judgments.

The participants in the 30-mel condition completed a practice block of 13 trials, with each token presented once in a random order. After the practice, they completed an experimental session of four blocks of the 13 tokens (52 trials), with the order randomized within each block.

The participants in the 60-mel condition completed a practice block of 7 trials, with each token presented once in a random order. After the practice, they completed an experimental session of four blocks of seven tokens (28 trials), with the order randomized within each block.

Results and Discussion

Average identification percentages for /i/ and /e/ are displayed in Figure 1. In accord with the results of Iverson and Kuhl (1995), Stimuli 1–9 were most often identified as /i/, and Stimuli 10–13 were most often identified as /e/. Stimuli 1–7 were identified as /i/ on at least 93% of all the trials, demonstrating that the participants were consistent in their categorization of these tokens.

The participants were less consistent at making /e/ identifications. Of the four tokens that were more often identified as /e/, none received this identification on more than 68% of the trials. The participants infrequently judged that the stimuli sounded like neither /i/ nor /e/. This response was given on 1% of the trials for Stimuli 1–9 (those that were most often identified as /i/) and was given on 11% of the trials for Stimuli 10–13 (those that were most often identified as /e/).

The location of the /i/–/e/ identification boundary (i.e., the point at which tokens were identified as /i/ on 50% of trials) appeared to differ somewhat for the 60- and the 30-mel conditions. That is, the boundary, in the averaged data, was located near Stimulus 11 in the 60-mel condition and near Stimulus 10 in the 30-mel condition. From inspection of Figure 1, this seemed to be solely a result of more frequent /i/ identifications for Stimulus 9 by listeners in the 60-mel condition. A post hoc independent samples t test verified that the /i/ identification percentages for Stimulus 9 in the 60-mel ($M = 75.0\%$) and 30-mel ($M = 47.7\%$) conditions were significantly different [$t(47) = 2.26, p < .05$].

Average goodness judgments are displayed in Figure 1. As can be observed from the confidence intervals plotted in the figure, goodness judgments were consistent. Stimulus 2 received the highest rating among the /i/ tokens, although the ratings of Stimuli 1–4 all had overlapping confidence intervals (i.e., equivalent goodness). None of the tokens was an excellent exemplar of the /e/ vowel category. These results are in accord with the results of Iverson and Kuhl (1995).

To further examine the location of the best /i/ exemplar, the location of the token with the highest goodness rating was identified for each participant (locations were averaged when more than one token received the same highest rating). The average best location was 2.2 mels ($SE = 7.0$ mels) to the right of Stimulus 2 in the 60-mel condition and was 1.6 mels ($SE = 8.0$ mels) to the right of Stimulus 2 in the 30-mel condition. In other words, the average best location was near the center of the four tokens with equivalently high goodness ratings (Stimuli 1–4).

From inspection of Figure 1, the goodness judgments in the 60- and 30-mel conditions did not substantially differ within the /i/ category. However, goodness judgments within the /e/ category seemed higher in the 30-mel condition. Post hoc independent samples t tests[2] on the /e/ tokens demonstrated that the goodness judgments for the two conditions were not significantly different [$t(33) = -1.20, p > .05$] for Stimulus 11 ($M_{60\text{-mel}} = 3.1$, $M_{30\text{-mel}} = 3.6$) but were significantly different [$t(40) = -2.80, p < .01$] for Stimulus 13 ($M_{60\text{-mel}} = 2.6$, $M_{30\text{-mel}} = 3.6$). The latter result is consistent with the differences observed in the /i/ identification percentages. That is, listeners in the 30-mel condition seemed marginally more disposed to judge that some tokens sounded like the vowel /e/.
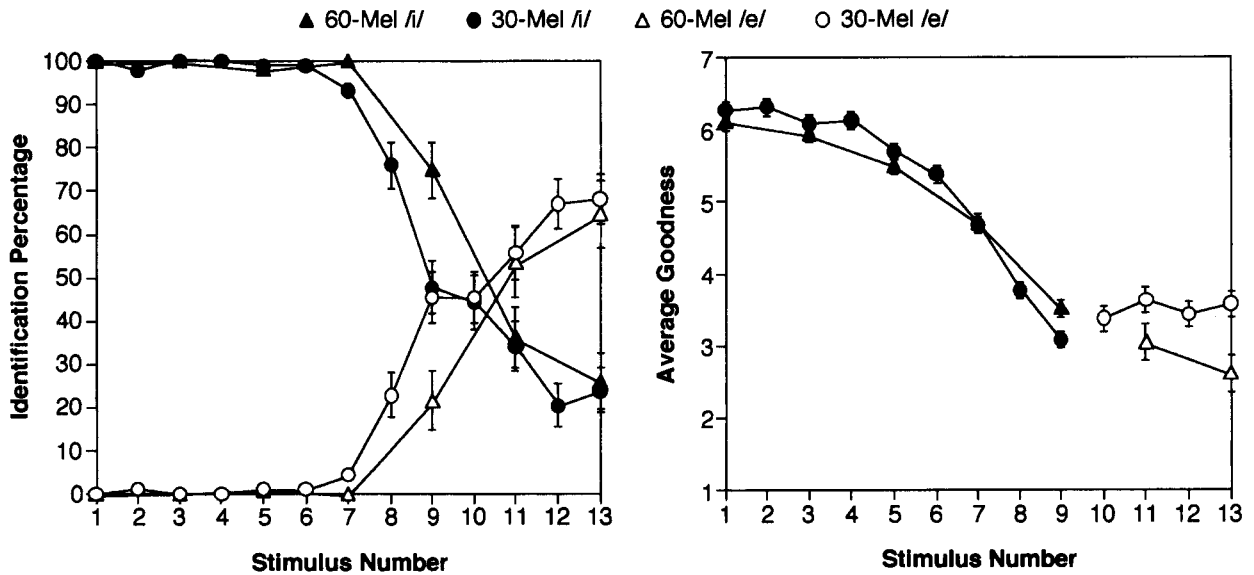
Figure 1. Phoneme identification percentages and average goodness ratings from Experiment 1. Error bars are 95% within-subjects confidence intervals (Loftus & Masson, 1994). The results demonstrate that the /i/−/e/ identification boundary was located near Stimulus 10 and that the best exemplar of the /i/ category was located near Stimulus 2. None was a good exemplar of the /e/ category.

EXPERIMENT 2
Fixed and Roving Discrimination With 30- and 60-mel Intervals

This experiment measured AX discrimination sensitivity for the stimuli of Experiment 1, with the aim of testing whether the contributions of phoneme identification and goodness are influenced by context variance. Specifically, the range of stimuli within each block of trials and the acoustic interval between each pair of stimuli were varied in separate conditions, to manipulate the role of attention and memory in AX discrimination tasks. Previous studies (Macmillan et al., 1988; Pisoni, 1973, 1975; Repp et al., 1979) have demonstrated that phoneme boundary effects are diminished in experimental conditions that reduce demands on attention and memory, but it is unknown whether perceptual magnet effects are also diminished.

Method

Participants. The participants were the same 49 individuals who completed Experiment 1. Each participated in a single condition. Ten participated in the 30-mel roving condition, 12 participated in the 30-mel fixed condition, 15 participated in the 60-mel roving condition, and 12 participated in the 60-mel fixed condition.

Apparatus and Stimuli. The apparatus and stimuli were the same as those in Experiment 1.

Procedure. The participants heard two stimuli on each trial and indicated whether the stimuli were the same or different by pressing marked keys on a computer keyboard. After each response, they received feedback that indicated whether their answers were correct. On half of the trials, the participants heard the same stimulus presented twice (*same* trials). On the other trials, the participants heard two different stimuli (*different* trials). For each pair of stimuli (A and B), each stimulus was presented on half of the *same* trials (i.e., half of the *same* trials were AA, and half were BB). The order of stimuli in the *different* trials was counterbalanced (i.e., half of the *different* trials were in the order AB, half were in the order BA). The offset-to-onset interval was 250 msec for each stimulus pair.

There were four conditions: 30-mel roving, 30-mel fixed, 60-mel roving, and 60-mel fixed. The 30- and 60-mel conditions were distinguished by the acoustic interval size between pairs of stimuli. The participants in the 30-mel conditions discriminated adjacent pairs of tokens along the full 13-token stimulus series. The participants in the 60-mel conditions discriminated adjacent pairs of tokens along a series composed of the 7 odd-numbered tokens in the set (see Table 1). The roving and the fixed conditions were distinguished by the range of stimulus pairs presented in each block of trials. The participants in the roving conditions heard all the stimulus pairs within each block. The participants in the fixed conditions heard only a single pair of stimuli within each block.

The participants in the 30-mel conditions first completed a practice session composed of four repetitions of the 12 stimulus pairs. The practice was a single block of 48 trials in the roving condition and 12 blocks of 4 trials in the fixed condition. The participants then completed an experimental session consisting of 12 blocks of 40 trials (20 *same* and 20 *different* trials for each pair), a total of 480 trials.

The participants in the 60-mel conditions first completed a practice session composed of four repetitions of the six stimulus pairs. The practice was a single block of 24 trials in the roving condition and six blocks of 4 trials in the fixed condition. The participants then completed an experimental session consisting of six blocks of 80 trials (40 *same* and 40 *different* trials for each pair), a total of 480 trials.

For all the conditions, the order of trials was randomized between and within blocks for each participant.

■ AX Discrimination   O Goodness Prediction   △ Identification Prediction

Fixed Discrimination

30-Mel Intervals

60-Mel Intervals

Roving Discrimination

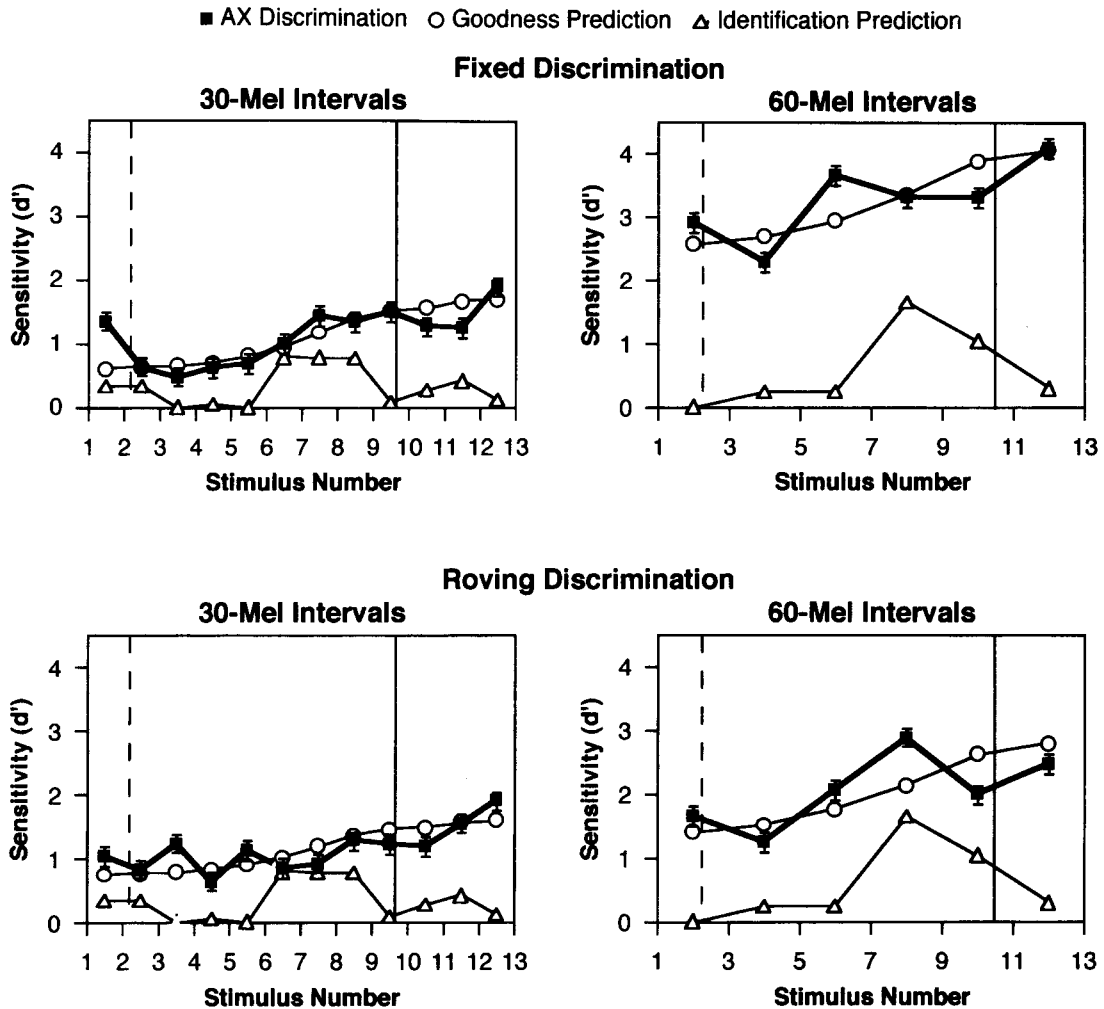### 30-Mel Intervals

60-Mel Intervals

Figure 2. Discrimination sensitivity in all conditions of Experiment 2 and sensitivity predictions based on the identification and goodness ratings of Experiment 1. The average best-exemplar locations are marked by vertical dashed lines, near Stimulus 2. The /i/–/e/ identification boundaries are marked by vertical solid lines, between Stimuli 9 and 11. Error bars are 95% within-subjects confidence intervals (Loftus & Masson, 1994). Goodness predicted sensitivity in all the conditions, but identification predicted sensitivity only in the 60-mel roving condition.

Results and Discussion

The mean results for each stimulus pair are displayed in Figure 2. Discrimination sensitivity (d') was measured for each participant through the application of signal detection theory (Macmillan & Creelman, 1991; Macmillan, Kaplan, & Creelman, 1977). As was suggested by Macmillan and Creelman, a differencing model was adopted to calculate d' in the roving conditions and an independent-observation model was adopted to calculate d' in the fixed conditions.

As a first attempt to detect whether perceptual magnet and phoneme boundary effects were present, the patterns of sensitivity (means and confidence intervals) in Figure 2 were inspected. The main criterion for the presence of a perceptual magnet effect is a minimum in sensitivity near the best-exemplar location. For each condition, there was a minimum in sensitivity near the best-exemplar lo-

cations (marked in Figure 2 by vertical dashes). However, the locations of sensitivity minima and best exemplars did not coincide exactly (i.e., the best-exemplar locations were near Stimulus 2, but the sensitivity minima, for all the conditions except 30-mel roving, appeared closer to Stimuli 3 and 4). A second criterion for the presence of a perceptual magnet effect is an inverse correlation between goodness and sensitivity. For this comparison, the average /i/-category goodness (from Experiment 1) was calculated for each pair of stimuli. The inverse of these goodness averages are plotted in Figure 2, with the means and variances scaled to match the sensitivity measure.[3] In all the conditions, sensitivity appears to be inversely correlated with goodness. This suggests that the perceptual magnet effect was present regardless of whether the context was fixed or roving or whether the stimulus interval was 30 or 60 mels.

The main criterion for the presence of a phoneme boundary effect is a sensitivity peak near the phoneme boundary. There were no peaks in sensitivity at the identification boundary between /i/ and /e/ (i.e., the vertical solid lines in Figure 2). However, there was a peak in sensitivity in the 60-mel roving condition approximately 60 mels to the left of the identification boundary, suggesting that a phoneme boundary effect may have been present in this condition.[4] A second criterion for the presence of a phoneme boundary effect is a correlation between sensitivity and the degree to which the stimuli within each pair were identified differently (i.e., sensitivity will be highest for stimuli that consistently receive different category labels, if a phoneme boundary effect is present). For this comparison, the mean percentages of /i/ identifications from Experiment 1 were used to estimate d' values, by calculating the difference between the z-transformed /i/-identification percentages for each pair of stimuli (Macmillan & Creelman, 1991). In the 60-mel roving condition, the peak in sensitivity predicted by identification percentages coincided with the location of the sensitivity peak obtained in the discrimination experiment, further suggesting that a phoneme boundary effect occurred in this condition. Note that the sensitivity peak predicted by identification coincides with the location at which the phoneme identification percentages had the steepest slope and that this location was to the left of the phoneme boundary. For the other conditions, no obtained sensitivity peaks coincided with those predicted by identification, suggesting that the phoneme boundary effect did not occur in those conditions.

Analyses of covariance were conducted for each condition, to test statistically whether perceptual magnet and phoneme boundary effects were present, on the basis of the criteria discussed above. In each analysis, discrimination sensitivity was the dependent measure, average goodness and the d' estimates based on identification were each continuous independent variables, and participant was coded as a categorical independent variable. The main effect of goodness was significant in each condition [30-mel fixed, $F(1,130) = 17.06, p < .001$; 30-mel roving, $F(1,130) = 17.06, p < .001$; 60-mel fixed, $F(1,58) = 21.02, p < .001$; 60-mel roving, $F(1,73) = 7.85, p < .01$]. This supports the conclusion that a perceptual magnet effect was present in each condition. The main effect of identification was significant in the 60-mel roving condition [$F(1,73) = 12.99, p < .001$] but was not significant (i.e., $p > .05$) in the other conditions [30-mel fixed, $F(1,130) = 0.91$; 30-mel roving, $F(1,108) = 1.826$; 60-mel fixed, $F(1,58) = 1.08$]. This supports the conclusion that a phoneme boundary effect occurred only in the 60-mel roving condition. In addition, the participant variable was significant ($p < .01$) in each condition [30-mel roving, $F(9,108) = 6.58$; 30-mel fixed, $F(11,130) = 3.90$; 60-mel roving, $F(14,73) = 2.13$; 60-mel fixed, $F(11,58) = 1.87$]. This demonstrates that there were differences in accuracy among participants.

In light of Macmillan et al.'s (1988) findings, it is surprising that there was no apparent phoneme boundary effect in the 30-mel roving condition. In their experiments, phoneme boundary effects were shown to occur under conditions in which context variance was high. One explanation for this lack of a phoneme boundary effect is that the listeners in the 30-mel roving condition were able to perform close to the limits of their sensory resolution. Specifically, performance in fixed conditions was predicted by Macmillan et al. (1988) to reflect sensitivity only on the basis of sensory resolution, and the total d' across the stimulus series for the 30-mel roving condition ($M = 13.6$) was not significantly different from that for the 30-mel fixed condition [$M = 13.4; t(20) = -0.08, p = .940$]. Perceptual anchor effects may not have emerged because such encoding strategies are only needed when context variance interferes with performance. Furthermore, in the 60-mel conditions, the total d' across the stimulus series was significantly poorer for the roving condition ($M = 12.25$) than for the fixed condition [$M = 19.5; t(25) = 5.38, p < .001$]. The phoneme boundary effect in the 60-mel roving condition is thus in accord with Macmillan et al.'s (1988) model, because the context variance interfered with performance and, thus, required encoding in terms of perceptual anchors.

The ratio of roving to fixed sensitivity for each stimulus pair in the 30- and 60-mel conditions are plotted in Figure 3. Under Macmillan et al.'s (1988) model, sensitivity in fixed conditions is predicted to exceed that in roving conditions (i.e., the fixed/roving ratio should be less than 1), and sensitivity in the roving condition is predicted to come closest to sensitivity in the fixed conditions at the location of a perceptual anchor. That is, a peak in the fixed/roving ratio indicates the presence of the perceptual anchor. For 60-mel conditions, there is a clear peak at Stimulus 8, suggesting that the listeners may have encoded stimuli in terms of a perceptual anchor at that location. In the 30-mel conditions, there is a peak between Stimuli 3 and 4, possibly indicating the presence of a perceptual anchor. However, this interpretation is suspect, because roving performance exceeds fixed performance at this location, contrary to the model's predictions. Furthermore, a post hoc t test demonstrated that the difference in fixed ($M = 0.5$) and roving ($M = 1.2$) d' values for this pair was not significant [$t(20) = -1.90, p = .07$]. There is no strong evidence that the listeners encoded stimuli in terms of a perceptual anchor in the 30-mel roving condition.

GENERAL DISCUSSION

The results from these experiments demonstrate that the perceptual magnet effect for vowels is not a result of increased sensitivity at phonemic boundaries. In accord with previous findings (Macmillan et al., 1988; Pisoni, 1973, 1975; Repp et al., 1979), the experiments showed that the peaks in sensitivity near vowel boundaries di-
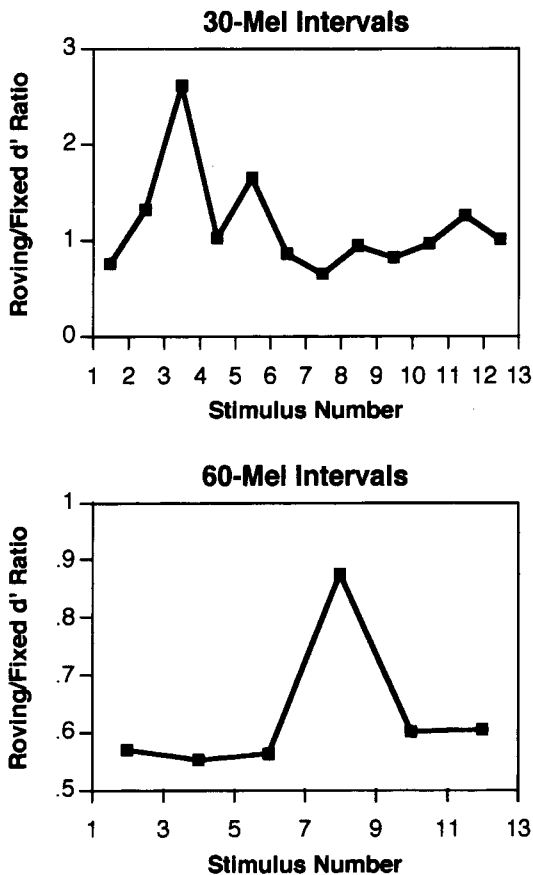
30-Mel Intervals

60-Mel Intervals

Figure 3. Ratio of roving to fixed discrimination d' for 30- and 60-mel intervals in Experiment 2. It is likely that the ratio peak between Stimuli 3 and 4 for the 30-mel intervals is unreliable; accuracy is not significantly different at that point for the fixed and the roving conditions. The ratio peak at Stimulus 8 for the 60-mel intervals is reliable (i.e., fixed and roving discriminations are significantly different) and indicates the existence of a perceptual anchor effect at that location.

minish in fixed discrimination tasks (at least for 60-mel intervals). In contrast, the low sensitivity near the best exemplars of /i/ was present in both the fixed and the roving conditions. It thus seems that perceptual magnet and phoneme boundary effects are caused by different mental processes, because the effects are differentially influenced by context variance.

It is somewhat surprising that interval size influenced boundary sensitivity peaks in the present study. Previous research (Healy & Repp, 1982; Pisoni, 1975) found sensitivity peaks when participants discriminated both one- and two-step intervals along a vowel stimulus series,[5] and Macmillan et al.'s (1988) perceptual anchor model does not directly predict that interval size should influence sensitivity peaks. However, some findings (see Macmillan, 1987, for a review) have suggested that listeners' sensitivity for one-step intervals tends to be about 8% better than would be expected from their performance on two-step intervals, suggesting that listeners may pay more at-

tention to small acoustic details of the stimuli when the task is difficult. Although this increased attention to small acoustic details may not usually be sufficient, in itself, to eliminate the phoneme boundary effect, other aspects of the experimental design (e.g., the short interstimulus interval, the use of isolated vowel stimuli, and the lack of tokens that were consistently labeled as /e/) may have contributed to the weakness of the phoneme boundary effect (i.e., the phoneme boundary effect may have been relatively weak, although not eliminated, even in the 60-mel roving condition). It is unclear exactly why a boundary sensitivity peak was not present in the 30-mel roving condition, but the fact that the perceptual magnet effect was present in this condition still demonstrates that the two effects rely on different processes.

The presence of the perceptual magnet effect in fixed conditions may suggest that it arises from general auditory—not cognitive—processes; sensitivity measured within fixed discrimination designs has been thought to reflect the basic sensory resolution of the auditory system (Macmillan et al., 1988). Furthermore, Sussman and Lauckner-Morano (1995) have hypothesized that auditory processing may account for the poor performance of listeners for the best /i/ stimuli used in the present experiments, independent of any effects of experienced speech (cf. Diesch, Iverson, Kettermann, & Siebert, 1999). It is, therefore, necessary to reconsider whether the mel-scale spacing of this stimulus set effectively equated auditory resolution. To further evaluate the effectiveness of mel-scale spacing, the acoustics of our stimuli were analyzed with a computational model of the auditory periphery. Following Moore and Glasberg's (1987) method, the Fourier spectrum was calculated for each sound, and this was used to estimate how each stimulus would excite an auditory filter bank that was designed to match basilar membrane excitation patterns. To examine how these excitation patterns differed, the RMS difference was calculated for every pair of the 13 stimuli, and this was used as a distance metric for multidimensional scaling (Kruskal, 1964a, 1964b). The multidimensional scaling solution revealed that the stimulus set was one-dimensional and that the stimuli were equally spaced. This suggests that it is likely that the mel-scale spacing of the stimulus set yielded uniform differences in peripheral excitation. However, it remains possible that the unequal sensitivity resulted from some higher level of auditory processing, especially at levels that may be altered by the distribution of experienced speech sounds.

There is some doubt that the fixed conditions of the present study were actually sufficient to allow listeners to attend to pure sensory differences. The listeners were untrained, the entire experiment took less than 1 h to complete, and the stimuli varied between blocks. It is possible that highly trained listeners in tasks that better reduce stimulus uncertainty (e.g., Kewley-Port & Watson, 1994) would have had equal sensitivity to these stimulus pairs. In addition, it is questionable whether any type of fixed discrimination truly measures sensory processing,

because sensitivity peaks at consonant boundaries do not diminish in these conditions. Macmillan et al. (1988) consider this to be evidence that consonant sensitivity peaks are due to general auditory processing, but it is also reasonable to conclude that phonemic labeling for consonants is an automatic process that is not altered by such experimental designs (Schouten & van Hessen, 1992). In other words, phonemic categorization is extremely well learned and ecologically important, and fixed discrimination procedures may not be sufficient to alter its influence. The fact that the perceptual magnet effect does not diminish in fixed discrimination tasks does not necessarily demonstrate that it is a sensory phenomenon. Rather, it indicates that perceptual magnet effects are less influenced by manipulations of context variance than are phoneme boundary effects.

The fixed discrimination results are consistent with Aaltonen et al.'s (1997) demonstration of the perceptual magnet effect with a physiological measure, the mismatch negativity component (MMN) of event-related potentials. MMN is elicited in response to a change in an auditory stimulus, and it is thought to reflect the resolution of preattentive auditory processing (Näätänen, 1992). It seems to be present regardless of whether listeners attend to or ignore the stimuli in an experiment. For example, Aaltonen et al. had participants watch a subtitled film with the sound off, while the experimental stimuli were played continuously in the background. The results demonstrated that MMN responses were smallest (lowest sensitivity) near stimuli that were judged by most participants[6] to be the best exemplars of the Finnish /i/ category, suggesting that the perceptual magnet effect alters sensitivity at a preattentive level.

Listeners' awareness of phonemic category identification or goodness may be independent of these preattentive changes in sensitivity (Aaltonen et al., 1997; see also Morais & Kolinsky, 1994). The correlation between goodness and sensitivity may arise because both are dependent on the acoustic distribution of native-language phonemes. For example, Kuhl et al. (1992) found that the perceptual magnet effect distorts sensitivity to vowels in the first 6 months of life, prior to the time at which infants learn word meanings. The acoustics of infants' ambient language could initially alter perceptual processing (see Guenther & Gjaja, 1996), and conscious notions of which sounds are best exemplars may develop at a later stage.

Although there is converging evidence that the perceptual magnet effect is present at low levels of perceptual processing, it is important to note that the magnitude of the perceptual magnet effect has been linked to individual differences in phonemic categorization. Aaltonen et al. (1997) were only able to find a relationship between MMN and category goodness among participants who were clearly able to identify their stimuli as /i/ or /y/; listeners who had inconsistently labeled these synthetic

tokens did not exhibit a perceptual magnet effect on the MMN measure. Iverson and Kuhl (1996) similarly found a perceptual magnet effect for participants who were clearly able to label stimuli as /r/ or /l/, but the clustering within the /r/ category was reduced for participants who were less willing to identify any stimuli as good exemplars of /r/. In addition, Lively and Pisoni (1997) were unable to find any perceptual magnet effect for exemplars of the vowel /i/, but their listeners exhibited substantially more variability in goodness judgments than has been found in other studies of that category (Iverson & Kuhl, 1995; Kuhl, 1991; Experiment 1 of the present study). Listeners thus seem to have less of a perceptual magnet effect when they are unable to reliably label synthetic stimuli as good exemplars of phoneme categories. It is unknown what causes these individual differences in categorization. However, it seems more plausible that they are linked to differences in cognitive categorization than to differences in auditory processing. The possibility of a causal relationship between cognitive categorization and the perceptual magnet effect, therefore, cannot be entirely dismissed.

To summarize, the present results suggest that the perceptual magnet effect for vowels is independent of the phoneme boundary effect. Perceptual magnet effects seem less influenced by experimental manipulations of attention and memory than are phoneme boundary effects, suggesting that these effects arise from different processes.

REFERENCES

AALTONEN, O., EEROLA, O., HELLSTRÖM, Å., UUSIPAIKKA, E., & LANG, A. H. (1997). Perceptual magnet effect in the light of behavioral and psychophysiological data. *Journal of the Acoustical Society of America*, **101**, 1090-1105.

DAVIS, K. D., & KUHL, P. K. (1994). Tests of the perceptual magnet effect for American English /k/ and /g/. *Journal of the Acoustical Society of America*, **95**, 2976.

DIEHL, R. L. (1987). Auditory constraints on speech perception. In M. E. H. Schouten (Ed.), *The psychophysics of speech perception* (NATO ASI Series D: Behavioural and Social Sciences, No. 39, pp. 210-219). Dordrecht: Nijhoff.

DIEHL, R. L., KLUENDER, K. R., WALSH, M. A., & PARKER, E. M. (1991). Auditory enhancement in speech perception and phonology. In R. R. Hoffman & D. S. Palermo (Eds.), *Cognition and the symbolic processes: Vol. 3. Applied and ecological perspectives* (pp. 59-76). Hillsdale, NJ: Erlbaum.

DIESCH, E., IVERSON, P., KETTERMANN, A., & SIEBERT, C. (1999). Measuring the perceptual magnet effect in the perception of German /i/. *Psychological Research*, **62**, 1-19.

DURLACH, N. I., & BRAIDA, L. D. (1969). Intensity perception: I. Preliminary theory of intensity resolution. *Journal of the Acoustical Society of America*, **46**, 372-383.

ESTES, W. K. (1993). Concepts, categories, and psychological science. *Psychological Science*, **4**, 143-153.

FANT, G. (1973). *Speech sounds and features*. Cambridge, MA: MIT Press.

FRY, D. B., ABRAMSON, A. S., EIMAS, P. D., & LIBERMAN, A. M. (1962). The identification and discrimination of synthetic vowels. *Language & Speech*, **5**, 171-189.

FUJISAKI, H., & KAWASHIMA, T. (1970). Some experiments on speech perception and a model for the perceptual mechanism. In *Annual Re-*

port of the Engineering Research Institute (No. 29, pp. 207-214). Tokyo: University of Tokyo Faculty of Engineering.

GRIESER, D., & KUHL, P. K. (1989). Categorization of speech by infants: Support for speech-sound prototypes. Developmental Psychology, 25, 577-588.

GUENTHER, F. H., & GJAJA, M. N. (1996). The perceptual magnet effect as an emergent property of neural map formation. Journal of the Acoustical Society of America, 100, 1111-1121.

HEALY, A. F., & REPP, B. H. (1982). Context independence and phonetic mediation in categorical perception. Journal of Experimental Psychology: Human Perception & Performance, 8, 68-80.

IVERSON, P., & KUHL, P. K. (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. Journal of the Acoustical Society of America, 97, 553-562.

IVERSON, P., & KUHL, P. K. (1996). Influences of phonetic identification and category goodness on American listeners' perception of /r/ and /l/. Journal of the Acoustical Society of America, 99, 1130-1140.

KEWLEY-PORT, D., & WATSON, C. S. (1994). Formant-frequency discrimination for isolated English vowels. Journal of the Acoustical Society of America, 95, 485-496.

KLATT, D. H. (1980). Software for a cascade/parallel formant synthesizer. Journal of the Acoustical Society of America, 67, 971-995.

KLUENDER, K. R. (1991). Effects of first formant onset properties on voicing judgements result from processes not specific to humans. Journal of the Acoustical Society of America, 90, 83-96.

KLUENDER, K. R., DIEHL, R. L., & WRIGHT, B. A. (1988). Vowel-length differences before voiced and voiceless consonants: An auditory explanation. Journal of Phonetics, 16, 153-169.

KRUSKAL, J. B. (1964a). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika, 29, 1-27.

KRUSKAL, J. B. (1964b). Nonmetric multidimensional scaling: A numerical method. Psychometrika, 29, 115-129.

KUHL, P. K. (1987). The special-mechanisms debate in speech research: Categorization tests on animals and infants. In S. Harnad (Ed.), Categorical perception: The groundwork of cognition (pp. 355-386).

KUHL, P. K. (1988). Auditory perception and the evolution of speech. Human Evolution, 3, 19-43.

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

KUHL, P. K. (1992). Infants' perception and representation of speech: Development of a new theory. In J. J. Ohala, T. M. Nearey, B. L. Derwing, M. M. Hodge, & G. E. Wiebe (Eds.), Proceedings of the International Conference on Spoken Language Processing (pp. 449-456).

KUHL, P. K. (1993a). Infant speech perception: A window on psycholinguistic development. International Journal of Psycholinguistics, 9, 33-56.

KUHL, P. K. (1993b). Innate predispositions and the effects of experience in speech perception: The native language magnet theory. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. MacNeilage, & J. Morton (Eds.), Developmental neurocognition: Speech and face processing in the first year of life (pp. 259-274). Dordrecht: Kluwer.

KUHL, P. K., & MILLER, J. D. (1975). Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. Science, 190, 69-72.

KUHL, P. K., & MILLER, J. D. (1978). Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. Journal of the Acoustical Society of America, 63, 905-917.

KUHL, P. K., & PADDEN, D. M. (1982). Enhanced discriminability at the phonetic boundaries for the voicing feature in macaques. Perception & Psychophysics, 32, 542-550.

KUHL, P. K., & PADDEN, D. M. (1983). Enhanced discriminability at the phonetic boundaries for the place feature in macaques. Journal of the Acoustical Society of America, 73, 1003-1010.

KUHL, P. K., WILLIAMS, K. A., LACERDA, F., STEVENS, K. N., & LINDBLOM, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. Science, 255, 606-608.

LACERDA, F. (1995). The perceptual-magnet effect: An emergent consequence of exemplar-based phonetic memory. In Proceedings of the

XIIIth International Congress on Phonetic Sciences (Vol. 2, pp. 140-147).

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.

LIVELY, S. E., & PISONI, D. B. (1997). On prototypes and phonetic categories: A critical assessment of the perceptual magnet effect in speech perception. Journal of Experimental Psychology: Human Perception & Performance, 23, 1665-1679.

LOFTUS, G. R., & MASSON, M. E. J. (1994). Using confidence intervals in within-subject designs. Psychonomic Bulletin & Review, 1, 476-490.

LOTTO, A. J., KLUENDER, K. R., & HOLT, L. L. (1998). Depolarizing the perceptual magnet effect. Journal of the Acoustical Society of America, 103, 3648-3655.

MACMILLAN, N. A. (1987). Beyond the categorical/continuous distinction: A psychophysical approach to processing modes. In S. Harnad (Ed.), Categorical perception: The groundwork of cognition (pp. 53-85). New York: Cambridge University Press.

MACMILLAN, N. A., & CREELMAN, D. (1991). Detection theory: A user's guide. New York: Cambridge University Press.

MACMILLAN, N. A., GOLDBERG, R. F., & BRAIDA, L. D. (1988). Resolution for speech sounds: Basic sensitivity and context memory on vowel and consonant continua. Journal of the Acoustical Society of America, 84, 1262-1280.

MACMILLAN, N. A., KAPLAN, H. L., & CREELMAN, C. D. (1977). The psychophysics of categorical perception. Psychological Review, 84, 452-471.

MILLER, J. L., & VOLAITIS, L. E. (1989). Effect of speaking rate on the perceptual structure of a phonetic category. Perception & Psychophysics, 46, 505-512.

MOORE, B. C. J., & GLASBERG, B. R. (1987). Formulae describing frequency selectivity as a function of frequency and level, and their use in calculating excitation patterns. Hearing Research, 28, 209-225.

MORAIS, J., & KOLINSKY, R. (1994). Perception and awareness in phonological processing: The case of the phoneme. Cognition, 50, 287-297.

NÄÄTÄNEN, R. (1992). Attention and brain function. Hillsdale, NJ: Erlbaum.

PISONI, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. Perception & Psychophysics, 13, 253-260.

PISONI, D. B. (1975). Auditory short-term memory and vowel perception. Memory & Cognition, 3, 7-18.

REPP, B. (1984). Categorical perception: Issues, methods, findings. In N. J. Lass (Ed.), Speech and language: Vol. 10. Advances in basic research and practice (pp. 243-335). New York: Academic Press.

REPP, B. H., HEALY, A. F., & CROWDER, R. G. (1979). Categories and context in the perception of isolated steady-state vowels. Journal of Experimental Psychology: Human Perception & Performance, 5, 129-145.

SAMUEL, A. G. (1982). Phonetic prototypes. Perception & Psychophysics, 31, 307-314.

SCHOUTEN, M. E. H., & VAN HESSEN, A. J. (1992). Modeling phoneme perception: I. Categorical perception. Journal of the Acoustical Society of America, 92, 1841-1855.

STEVENS, S. S., VOLKMANN, J., & NEWMAN, E. B. (1937). A scale for the measurement of the psychological magnitude pitch. Journal of the Acoustical Society of America, 8, 185-190.

STUDDERT-KENNEDY, M., LIBERMAN, A. M., HARRIS, K. S., & COOPER, F. S. (1970). Motor theory of speech perception: A reply to Lane's critical review. Psychological Review, 77, 234-249.

SUSSMAN, J. E., & LAUCKNER-MORANO, V. J. (1995). Further tests of the "perceptual magnet effect" in the perception of [i]: Identification and change/no-change discrimination. Journal of the Acoustical Society of America, 97, 539-552.

VOLAITIS, L. E., & MILLER, J. L. (1992). Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories. Journal of the Acoustical Society of America, 92, 723-735.

WOOD, C. C. (1976). Discriminability, response bias, and phoneme categories in discrimination of voice onset time. *Journal of the Acoustical Society of America*, **60**, 1381-1389.

ZWICKER, E., FLOTTORP, G., & STEVENS, S. S. (1957). Critical band width in loudness summation. *Journal of the Acoustical Society of America*, **29**, 548-557.

NOTES

1. For each step along the stimulus series, there was a 21.2-mel change in $F1$ frequency and a 21.2-mel change in $F2$ frequency. The combined Euclidean formant frequency difference between neighboring tokens along the series was 30 mels [i.e., $30 = (21.2^2 + 21.2^2)^{0.5}$].

2. The degrees of freedom are different for the two tests because some participants did not identify these tokens as /e/ and, thus, did not give any goodness ratings with regard to the /e/ category.

3. Goodness was scaled in this way to make goodness and sensitivity easier to visually compare in Figure 2. This scaling is valid, because the criterion is that goodness and sensitivity must be correlated (i.e., correlation measures are insensitive to differences in means or variances).

4. One could conclude that there is a small sensitivity peak for Stimuli 5 and 7 in the 60-mel fixed condition. It is considered here to be less likely to be indicative of a phoneme boundary effect, because it occurred further from the phoneme boundary. In addition, this "peak" did not seem to influence the roving/fixed ratio analysis (see below), suggesting that it was not caused by a perceptual anchor.

5. Sensitivity peaks can actually diminish for some larger intervals because of ceiling effects (Healy & Repp, 1982; Macmillan, 1987).

6. MMN indicated only a perceptual magnet effect for participants who were good categorizers of these stimuli; listeners who were poor at judging whether these stimuli were /i/ or /y/ did not demonstrate a perceptual magnet effect for MMN. This is described in more detail later in the General Discussion section.