

# The perceptual magnet effect in Australian English vowels

NICK THYER and LOUISE HICKSON  
*University of Queensland, Brisbane, Australia*

and

BARBARA DODD  
*University of Newcastle, Newcastle upon Tyne, England*

Recent research (Kuhl, 1991) has suggested that the internal structure of vowel categories is graded in terms of stimulus goodness. It has been proposed that a *best instance* stimulus reflects a central point or prototype, which effectively renders within-category members perceptually more similar. Discrimination experiments suggest a nonlinear relationship between acoustic and perceptual space near category centers (Iverson & Kuhl, 1995b). This phenomenon has been described as the *perceptual magnet effect*. The present study investigated the presence of the perceptual magnet effect in five Australian vowel categories. Australian English speakers identified, rated, and discriminated between a pool of 32 vowel stimuli that varied in  $F1$  and  $F2$  values. The results from Experiments 1 and 2 showed that subjects were able to judge the quality and identity of each stimulus and that a general grading of stimulus quality was reported. This was not symmetrical, and the subjects' responses varied considerably. In Experiment 3, closer control of the methodology in the discrimination task and of contextual factors influencing the test materials was exercised. Despite this, evidence of the warping of perceptual space in discrimination data was not found. In general, these results do not provide support for the existence of the perceptual magnet effect, and explanations for this finding are discussed.

For three decades or more, a large body of research on speech perception has concentrated on the nature of between-category perception. Recent work has focused on the internal structure of phonetic categories. This change of focus has been largely fueled by studies suggesting that listeners may organize phonetic categories around a central prototypical<sup>1</sup> member of the category (Eimas & Corbit, 1973; Massaro & Cohen, 1991; Samuel, 1982). Various studies have used different techniques to reveal fine, within-category structure. For example, adaptation experiments have shown that prototypical stimuli are more effective adapters than are stimuli taken from around the category boundary area (Miller, 1977; Miller, Conine, Schermer, & Kluender, 1983; Samuel, 1982). Other direct measures have shown a systematic increase in category goodness ratings by listeners for various voice onset time (VOT) values, as stimuli were moved from a boundary toward a more central position (Miller & Volaitis, 1989). A subsequent decrease in goodness rating was observed as the stimuli were moved toward the opposite boundary. A similar structure has also been dem-

onstrated to exist in vowels (Grieser & Kuhl, 1989; Kuhl, 1991; Miller & Eimas, 1996).

Clearly, there is strong evidence indicating a fine structure within phonetic categories that is somehow influenced by or related to the perceived quality or goodness of exemplars. Research into categorization processes (Nosofsky, 1986, 1987, 1991) has suggested that category members may cluster to form groups of exemplars that have similar attributes. This clustering has the effect of distorting the perceptual space, so that there is no longer a one-to-one mapping of physical dimensions onto psychological dimensions. Kuhl (1991) also suggested that category goodness influences the way in which humans categorize phonemes and that, within a category, there exists a prototypical example that is perceived as being a best instance of that category. The function of this prototype is to act as a perceptual magnet that draws members of a group together, rendering them perceptually similar to the category prototype. This *perceptual magnet effect* has been investigated for one vowel category with infants and adults (Kuhl, 1991). It has been suggested to be specific to human speech, since it is not present in primates (Kuhl, 1991), and is claimed to be native language dependent (Kuhl, 1993, 1994; Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992).

Two broad suppositions underlie Kuhl's (1991) perceptual magnet effect. The first is that members of a phonetic category are organized within that category in terms of their phonetic quality. Furthermore, phonetic quality degrades symmetrically as the distance from a

---

The authors thank Gerry Docherty, Peter Bailey, and two anonymous reviewers for extremely helpful comments on earlier drafts of this paper. We also thank Dianne Bates for invaluable help in manuscript preparation and Tony Baglioni for statistical advice. Correspondence concerning this article should be addressed to N. Thyer, Department of Speech Pathology and Audiology, University of Queensland, Brisbane, Queensland 4072, Australia (e-mail: n.thyer@mailbox.uq.edu.au).

—Accepted by previous editor, Myron L. Braunstein

central, prototypical stimulus increases. Evidence for the first supposition is provided by Kuhl in her original experiment (Kuhl, 1991). Two arrays of 32 variants of the vowel /i/ were synthesized. The array coordinates of each variant was defined by its *F1* and *F2* values, measured on the mel scale, which placed it on one of four orbits centered on a stimulus representing either a prototype (good quality) or a nonprototype (poor quality) category center. Thus, there were 8 equally spaced variants on each orbit. The orbits were 30, 60, 90, and 120 mels from the center. Sixteen adult listeners were required to make quality judgments about each /i/ in terms of its goodness or representativeness of the category as a whole. The results showed that listeners gave good quality judgments to stimuli at the center of the array of stimuli and progressively poorer quality judgments to stimuli at the extremes of the array. Kuhl (1991) noted remarkable symmetry in quality judgments, in that judgments became poorer as a function of distance from the center of the prototype array. This was not seen in the nonprototype array. These findings were taken as evidence of a symmetrical graded structure within the vowel space.

Attempts to replicate these data have delivered mixed results. The variation in goodness ratings observed by Kuhl in the prototype and nonprototype arrays was evident in recent data (Lively & Pisoni, 1997), but the symmetrical gradation was not obvious in either the prototype or the nonprototype stimulus arrays.

For the second supposition, much of the evidence for the perceptual magnet effect comes from discrimination experiments (Iverson & Kuhl, 1995a, 1995b; Kuhl, 1991; Miller, 1994; Sussman & Lauckner-Morano, 1995). Specifically, discrimination between members of a category placed near a prototype is described as being poor in relation to discrimination of members situated further away from a prototype. Poor discrimination at the center reflects the perceptual magnet effect in operation, since stimuli are rendered more similar. This was demonstrated, for example, in Kuhl's (1991) second experiment. She asked adults to discriminate between either the prototype /i/ and its variants or the nonprototype /i/ and its variants. Results revealed significantly poorer discrimination performance when the prototype /i/ was the referent than when the nonprototype /i/ was the referent. The evidence of discrimination changes within categories found in this experiment led Kuhl (1991) to suppose that the prototype exerted an effect much like a perceptual magnet, shrinking the perceptual space in the center of the category (Nosofsky, 1987). This conclusion is certainly consistent with Kuhl's (1991) results, but the subsequent investigations (Iverson & Kuhl, 1995b; Sussman & Lauckner-Morano, 1995) revealed some problematic aspects of Kuhl's work. For example, Kuhl's (1991) nonprototype stimulus was identified in an /i/-not-/i/ paradigm as belonging to the /i/ category only 26% of the time (Sussman & Lauckner-Morano, 1995). In fact, it was probably more readily perceived as either /e/ or /ɪ/. This would mean that better discrimination involving the nonprototype /i/ may have been a result of across-boundary

identification, and not of any clustering of perceptual space within the category that was due to the perceptual magnet effect.

Consequently, Sussman and Lauckner-Morano (1995) synthesized a more appropriate nonprototype from Kuhl's (1991) stimulus values that was likely to be identified as /i/ 50% of the time. This item was situated at the category boundary. Five groups of adults were tested: one with a prototype and variants, and four with a range of nonprototype stimuli and variants. The subjects' discrimination, response strategy, and reaction times (RTs) were measured. Overall, the prototype group had the poorest sensitivity, the highest number of misses, and the slowest RTs for hits. However, poorer discrimination performance of the prototype group, in comparison with other groups, could have been a result of improved discrimination owing to either across-boundary judgments or membership of a different category. For example, differences revealed by comparison of the prototype group with the 75-up and 120-mel groups may have involved separate category members, since subjects identified the category boundary to be 75 mels from the prototype. Stimuli heard by these groups were between 75 and 135 mels distant from the prototype. The important exception was the comparison between the prototype and the 75-down groups. Here, the variant stimuli were common to both continua, simply changing mel steps in opposite directions. Listeners in the prototype group showed significantly poorer discrimination than those in the 75-down groups, which is consistent with a perceptual magnet effect. Sussman and Lauckner-Morano suggest that this difference might be caused by closer to optimal performances when *F1* values of consecutive tokens are decreasing, as in the 75-down group stimuli.

In a third experiment, they hypothesized that a nonprototype that was identified as /i/ 50% of the time may not have been an appropriate nonprototype, since it may have allowed across-boundary discriminations and artificially contributed to a *magnet effect*. They considered the possibility that a better nonprototype might be a token identified as /i/ 65% of the time (67-mel stimulus). Results indicated that the hypothesized, more optimal nonprototype stimulus was not always more easily discriminated than the prototype and was not always discriminated differently to the stimuli 75-mel groups. It was, therefore, not shown convincingly to have been a better nonprototype stimulus than one situated at the category boundary. They concluded that the results for the prototype group in their Experiments 2 and 3 partly supported those found by Kuhl (1991, 1992) but did not provide conclusive evidence of the existence of phonological prototypes operating as perceptual magnets.

Category membership of Kuhl's nonprototype stimulus was again investigated by Lively and Pisoni (1997), who tested phonetically naive subjects in an /i/-not-/i/ identification task. They found that their subjects identified Kuhl's nonprototype stimulus as /i/ only 28% of the time. In addition, they reported that their subjects did not agree on the same prototype stimulus as Kuhl's nonnaive

subjects and that individual subjects appeared to have their own internal representation of prototypicality. This finding suggests that, if phonetic prototypes do exist, they may be different for each subject. Lively and Pisoni proceeded to test discrimination with arrays of stimuli synthesized around individual subjects' own measured prototypes and again failed to demonstrate evidence of the perceptual magnet effect.

Finally, Iverson and Kuhl (1995b) reviewed Kuhl's (1991) earlier work and addressed the issue of identification, by asking subjects to identify a subset of Kuhl's (1991) stimuli, to ensure they were perceived as belonging to the category /i/, before asking them to rate the stimuli on a scale from 1 to 7 (1, poor /i/; 7, good /i/). Kuhl (1991) had proceeded directly to a discrimination task and had not studied identification. Iverson and Kuhl (1995b) found that subjects identified the good-quality stimuli more quickly than the poor-quality stimuli. Their data broadly corresponded to the data collected by Kuhl (1991). Kuhl's (1991) nonprototype was identified as /i/ 55% of the time and gained a mean rating of 3.2. Kuhl's previous work had been restricted to investigating overall percentage correct and proportion of misses as a measure of generalization. In this second study, they used signal detection theory (Macmillan & Creelman, 1991) to analyze discrimination, and they used RTs and multidimensional scaling analysis to describe mathematically how the perceptual space within a category might be distorted. Results showed that, although the stimuli's formant values were equally spaced on the mel scale,<sup>2</sup> in perceptual space the good instances of stimuli were clustered together, implying poor discrimination around them. Iverson and Kuhl (1995b) point out, importantly, that this pattern is different to that expected from classical categorical perception. The perceptual magnet effect predicts *differences* in discrimination for a range of identically labeled stimuli. Categorical perception predicts the *same* level of discrimination for a range of identically labeled stimuli. They both predict discrimination peaks at identification boundaries.

In summary, two fundamental assumptions underpin the logic of Kuhl's (1991) perceptual magnet effect. The first is that the internal structure of phonetic categories is invariant and symmetrically graded. This can be demonstrated by varying levels of quality judgments given to stimuli lying within the category. The second is the ability to show differences in discrimination and identification scores within a category that cannot be explained by categorical perception and that suggest a warping of perceptual space. Research to date has shown clear evidence for graded category membership (see, e.g., Flege, Schmidt, & Wharton, 1996; Iverson & Kuhl, 1995b; Kuhl, 1991; Miller & Volaitis, 1989; Sussman & Lauckner-Morano, 1995). However, it does not consistently demonstrate a symmetrical gradation, invariant quality ratings, or distortion of the perceptual space suggestive of magnet effects (Lively & Pisoni, 1997; Sussman & Lauckner-Marano, 1995).

Robustness of the effect warrants further investigation, particularly since the perceptual magnet effect has

been argued to play a central part in Kuhl's native language model of speech perception (Kuhl, 1994; Kuhl & Meltzoff, 1996; Kuhl et al. 1992).

## THE PRESENT RESEARCH

In the present study, the intention was to provide a fairer test for the presence of the perceptual magnet effect in five synthetic Australian vowel categories by addressing some factors not previously accounted for in a single study. Previous investigations have been confined to the American English vowel /i/ and the Swedish /Y/. Evidence of the magnet effect in five vowel categories of a different English vowel system would add to our understanding of speech categorization processes.

Three experiments were conducted. In the first experiment, the claim (Kuhl, 1991) that vowel categories have a graded structure that is invariant and declines symmetrically in quality as the distance from a prototypical center increases was tested in five Australian vowel categories. The aim of the second experiment was to investigate category membership of the stimuli rated in Experiment 1.

By scrutinizing the rating and identification data from Experiments 1 and 2, suitable prototype and nonprototype stimuli from each category were selected, which had opposing levels of category goodness, yet retained category membership. The first two experiments were expected to add to the literature in regard to graded category membership and were necessary to define the test materials for Experiment 3. The third experiment was designed to address some methodological concerns overlooked by previous research.

First, the internal structure of speech categories has been shown to be sensitive to changes of context. For example, speaking rate can radically change the location of best category exemplars, and their quality ratings change as the acoustic context changes (Miller & Volaitis, 1989; Volaitis & Miller, 1992). Lively and Pisoni (1997) tested for the perceptual magnet effect in a vowel discrimination task, using both a *same/different* and a modified four-alternative forced choice task, but they compared across context. Their previous experiment had indicated that the nonprototype stimulus was rated quite differently when presented in the context of the nonprototype array, as compared with the prototype array. This may indicate that, according to context models of categorization, prototypes or category best instances may change with stimulus set ranges (cf. Iverson & Kuhl, 1995b). This may have influenced their listeners' discrimination results, since the prototype, the nonprototype, or, indeed, any stimulus in the array would have differing levels of goodness or category membership. The tasks are not, therefore, equivalent. Comparing discrimination performance without equivalence could obscure any observable magnet effect. In Experiment 3 of the present study, the prototype and the nonprototype stimuli were taken from the same array, in order to avoid influences of context, making the prototype and the nonprototype test conditions equivalent.

lent. If the rating scores in this array are related to prototypicality (and since the stimulus context is unchanged), it should be possible simply to select any good-quality example as a prototype and any poor-quality example as a nonprototype and test for a perceptual magnet effect. In the third experiment, it was hypothesized that performance would be poorer when subjects discriminated within-category stimuli from the selected prototype stimulus than when they discriminated within-category stimuli from the selected nonprototype stimulus. In addition, prior phonetic experience would influence their discrimination performance.

Since the perceptual magnet effect has been argued to be apparent in discrimination tasks (e.g., Kuhl, 1991), Experiment 3 was designed to maximize discrimination performance between consecutive stimuli. Kuhl (1991) used a 1,000-msec interstimulus interval (ISI) in her change/no-change discrimination task. Long ISIs have been shown to promote the use of phonetic labeling when discriminating pairs of stimuli, rather than comparing them on the basis of their acoustic traces (Fujisaki & Kawashima, 1969). This may have made between-category distinctions within Kuhl's nonprototype stimuli more obvious and, therefore, facilitated discrimination, falsely indicating a magnet effect. The same effect may also influence discrimination of category boundary stimuli, since for these ambiguous stimuli, subjects may be prepared to label them in favor of the across-boundary category in order to optimize performance (Nosofsky, 1986). Sussman and Lauckner-Morano (1995) used a 500-msec ISI, which would have improved the salience of acoustic cues in their discrimination task. But since Pisoni (1973) found that vowel formant discrimination improves when the ISI is shortened from 500 to 250 msec, to further increase the salience of spectral cues available in our *same/different* task, a short ISI of 250 msec was used in Experiment 3. Iverson and Kuhl (1995b), in fact, used an ISI of 250 msec in their second experiment. However, even though small spectral differences in their adjacent stimuli may have been easier to discriminate as a result, their demonstration of a magnet effect requires some consideration. Their stimuli were taken from Kuhl's (1991) array and were made up of 13 tokens from the diagonal where  $F1$  and  $F2$  change in an opposing manner in 30-mel steps. More specifically, they divided these stimuli into two groups. The first was Kuhl's (1991) prototype stimulus and tokens 30, 60, and 90 mels to the left and right of it (prototype condition), and the second was Kuhl's nonprototype stimulus and tokens 30, 60, and 90 mels to the right and left of it.

Iverson and Kuhl's (1995b) main result was that their subjects were significantly worse at discriminating stimuli from the prototype than from the nonprototypes. However, the same criticism regarding across-boundary comparisons in the nonprototype condition that has been leveled at Kuhl (1991) can be applied here.

Better evidence of a magnet effect came from Iverson and Kuhl's (1995b) second experiment. First, they found a significant difference between discrimination of stimuli positioned to the right and to the left of the prototype.

However, stimuli to the left of the prototype had previously been given high quality ratings and identification scores, which indicated that they were all more or less highly representative of the category. Claims of a magnet effect from these stimuli should be considered in light of the fact that all of these stimuli were discriminated from a prototype (one of many). The prototype(s) may have assimilated tokens, but since there was no comparison with discrimination from a nonprototype for these stimuli, it is difficult to interpret this finding as evidence for a perceptual magnet effect. Nevertheless, the data does indicate, as Iverson and Kuhl (1995b) point out, that category goodness can influence discrimination within a category (see also their third experiment). Stimuli to the right of the prototype, however, had been given a quality rating of between 3.7 and 5.5 and identified as /i/ between 83% and 98% of the time. In this case, the same stimulus set was discriminated from the prototype and the nonprototype referents. These results showed a difference in performance that only approached significance ( $p = .11$ ). Although it is not explicitly stated, it seems likely that this significance results from the comparison for the 30-mel distance. These results are interesting, since it is this condition that is free from the effects of the context. In the other conditions of this experiment and in previous research (Kuhl, 1991; Lively & Pisoni, 1997; Sussman & Lauckner-Morano, 1995), the stimulus set was different when discrimination was compared between tokens and a prototype or a nonprototype condition. In Experiment 3 of the present study, discrimination of stimuli from the same stimulus set (with a 15-mel step size, smaller than that used by Iverson & Kuhl, 1995b), from a prototype and a nonprototype is compared.

Finally, Kuhl's (1991) stimulus arrays (and those tested by Lively & Pisoni, 1997) were described in terms of radial distance from a central stimulus. It is not explicitly stated how the coordinates were calculated, but if each token was placed on a circular orbit, the change in  $F1$  and  $F2$  values for the diagonals was different from the change in  $F1$  and  $F2$  values for the vertical and horizontal axes. In studies in which all 32 stimuli in the prototype and nonprototype arrays are used, this fact introduces another nonlinearity along the physical dimension that may influence the plausibility in interpreting results as a nonlinearity in perception. In this study,  $F1$  and  $F2$  changes were made in equal steps. The result was a series of rectangular orbits, rather than circular orbits. Circular orbits mean that the increments that effect simultaneous  $F1$  and  $F2$  changes are larger than those where  $F1$  and  $F2$  change separately, by a factor of 1.4.

The following series of three experiments were designed to be a fairer test of the perceptual magnet effect, since methodological factors were more tightly controlled than in any single investigation to date.

## EXPERIMENT 1

The five vowels chosen for this experiment were /i/, /ɜ/, /ɔ/, /a/, and /æ/. The vowels were chosen because

**Table 1**  
**Stimulus Coordinates of the Five Central Tokens for the**  
**Australian English Vowels Used in Experiments 1 and 2**

Formant	/ɔ/	/ɜ/	/a/	/æ/	/i/
F1	440	485	730	756	273
F2	750	1,430	1,100	1,850	2,328
F3	2,495	2,525	2,440	2,575	3,010
F4	3,300	3,300	3,300	3,300	3,300
F5	3,750	3,750	3,750	3,750	3,750

/i/, /ɔ/, /a/, and /æ/ have F1 and F2 values that define the four corners of the Australian vowel quadrilateral and the vowel /ɜ/ is situated roughly in the middle of this formant space. An initial trial synthesis of prototype and nonprototype stimulus arrays for these vowels suggested that stimuli positioned further than 120 mels from the center were, more often than not, situated in another category or were so extreme in their F1 and F2 values that they sounded unnatural. Because the focus of this study was on within-category structure, all the stimuli following Kuhl's nonprototype array were not synthesized, the exception being those four stimuli that were on the vector that was common to both her prototype and her nonprototype arrays. Our aim was to describe a large enough proportion of each vowel space to be able to examine structure and to be able to define a prototype and a nonprototype within that space.

## Method

**Subjects.** Sixteen adults (10 females and 6 males), 22–41 years of age (mean = 28.6), participated in this experiment. All the subjects had normal hearing (i.e., pure tone air conduction responses at 20 dBHL for octave interval frequencies from 250 to 8000 Hz). They had no reported history of hearing disorder or of speech and language disorder and had no training in phonetics. The subjects were native Australian English speakers, were right-handed, and were staff and students at the University of Queensland.

**Stimuli.** The stimuli were 33 synthetic vowels in each of the categories /i/, /ɔ/, /ɜ/, /a/, and /æ/. For each vowel category, an initial prototype was selected that had formant values corresponding to single utterances of vowels found in Australian English (Bernard, 1989). Bernard's data were originally collected in the late 1960s. On the basis of phonetic transcription, Bernard grouped Australian English speakers into the three categories of cultivated, general, and broad speakers. He documented the acoustic specifications for each group in terms of formant values for the major vowels. No large-scale national investigation into the acoustic specifications of Australian English has been undertaken since. However, studies of adult pronunciation in Sydney (Horvath, 1985) and of Australian adolescents nationally (Mitchell & Delbridge, 1965) have suggested that, overall, individual speakers appear to use a mixture of broad, general, and cultivated vowels. Furthermore, for specific vowels, they appear to consistently use a pronunciation from one category.

In a pilot test, a range of stimuli centered on the vowels /æ/, /i/, /ɔ/, /a/, and /ɜ/ were synthesized. The central vowels were synthesized using the first three formant values from Bernard's (1989) general category. Fifteen subjects were asked to rate each stimulus in terms of category goodness. Results showed that, for /æ/ and /i/, the subjects thought that the general prototype was atypical and chose a different

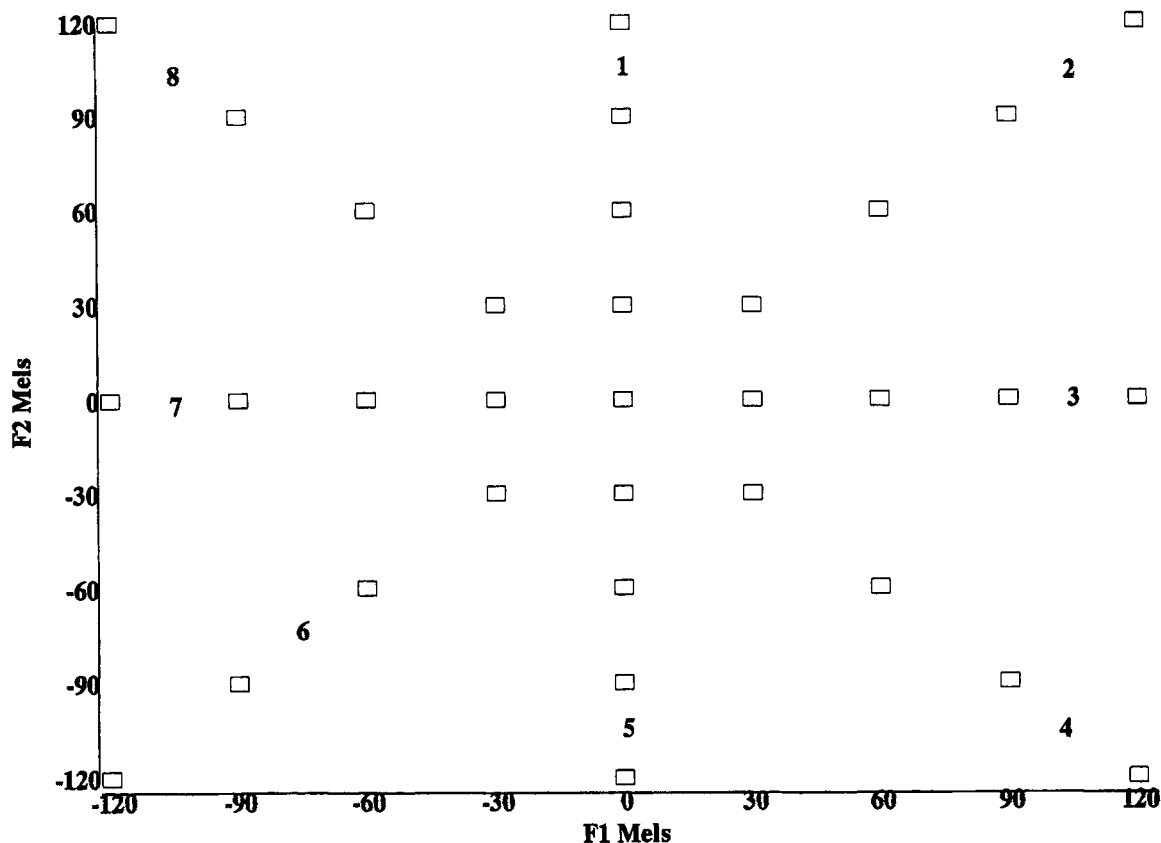


Figure 1. Distribution of the stimuli used in Experiments 1 and 2. The stimuli are arranged around a prototype so that F1 and F2 vary in equal steps of 30 mels. The eight vectors emanate from the central prototype.

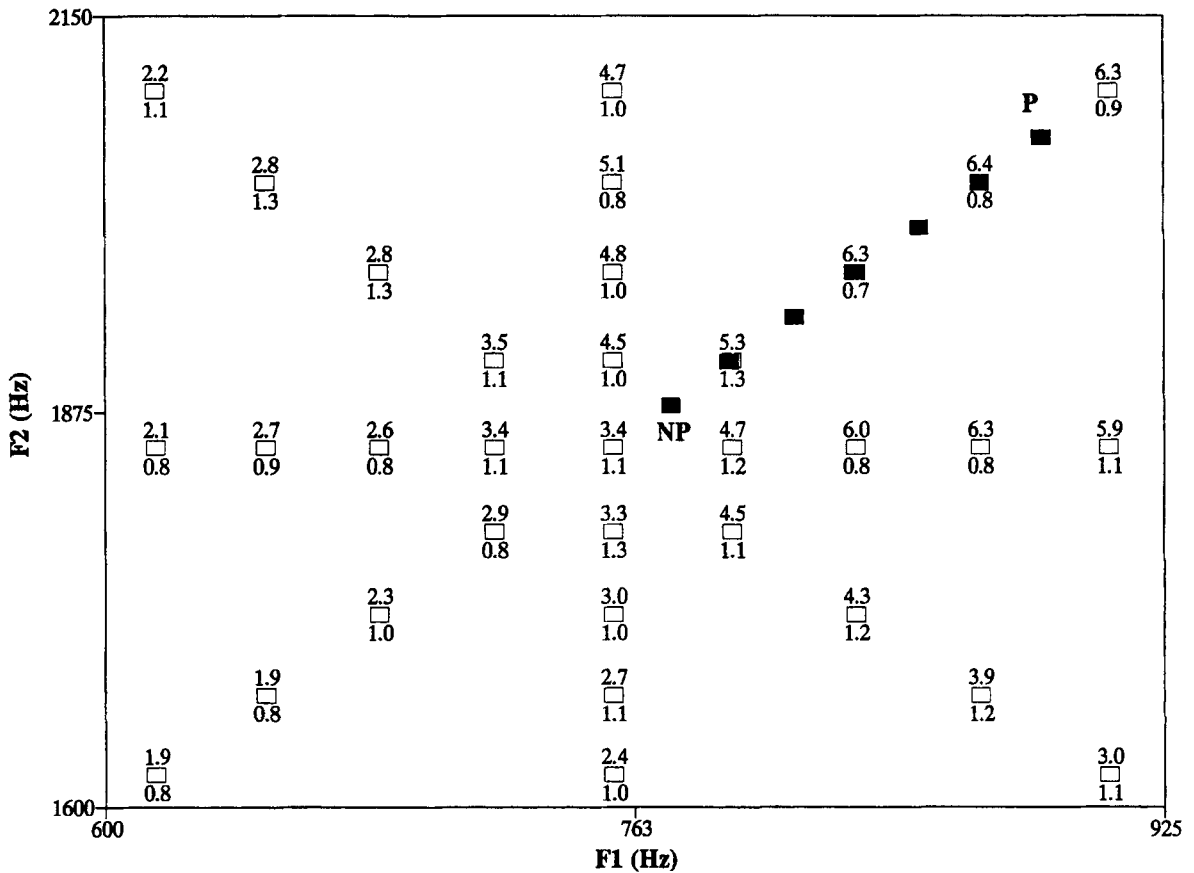


Figure 2. From Experiment 1, the subjects' ratings of each stimulus in the Australian English vowel category /æ/. The upper number placed by each stimulus is the mean; the lower number is the standard deviation. Stimulus coordinates are a function of  $F1$  and  $F2$ , depicted in hertz. The locations of the prototype (P) and the nonprototype (NP) are shown. The filled squares are the locations of the stimuli on the continuum selected for use in Experiment 3.

stimulus with formant values more like the cultivated versions of the vowels, as being the best example. Consequently, the pilot prototype formant values for /i/ and /æ/ were used in the present study. The formant values for /ɔ/, /ɜ/, and /a/ used are those described by Bernard. The formant values of all prototype vowels are presented in Table 1.

The values of the first and second formants ( $F1$  and  $F2$ ) were altered in steps of 30 mels, to form a series of 32 vowels spaced around each prototype. The change in the physical distance between each stimulus was, therefore, constant in terms of  $F1$  and  $F2$ . The mel was used for a number of reasons. The difference limen for the first three formants are similar when formant values are expressed in mels (Flanagan, 1957). The mel scale is believed to equate the perceived magnitude of pitch over a range of frequencies. Using the mel scale to describe the differences in  $F1$  and  $F2$  should preserve equal psychophysical spacing for the formant's pitch (see note 2). The values were calculated, using the formula suggested by Fant (1973,  $y = k \log(1 + f/1,000)$ , where  $y$  is the mel scale value,  $k$  is a constant, and  $f$  is the formant frequency in hertz. The result was 32 stimuli arranged in a series of four squares, each made up of 8 stimuli. Each square became progressively larger in 30-mel steps and was centered on the category prototype. Previous investigators' stimuli (Kuhl, 1991; Lively & Pisoni, 1997; Sussman & Lauckner-Morano, 1995) were situated on circular orbits, 8 stimuli to each, which meant that the  $F1$  and  $F2$  step sizes between stimuli on the diagonal dimensions were different from that of stimuli on the horizontal and vertical axes. The square configuration was thought to be more appropriate, since it maintained a constant step size for the experimental independent variables,  $F1$  and  $F2$ , in all vectors (see Figure 1).

All the stimuli were synthesized with five formants, 16-bit resolution, and at a sample rate of 12 kHz, using a version of the Klatt (1980) speech synthesizer (CSRE45, Avaaz Innovations, Inc., London, ON, 1995). For /i/, /ɔ/, /ɜ/, and /a/, the fundamental frequency contour was the same. It rose from 112 to 132 Hz in the first 100 msec and then fell to 92 Hz over the next 400 msec. The duration of these vowels was 500 msec. Voicing amplitude increased linearly from 0 to 45 dB in the first 40 msec and increased to 53 dB over the next 40 msec to a maximum of 55 dB for 20 msec. The amplitude then fell back to 53 dB for 20 msec, to 45 dB for 380 msec, and then to 40 dB at 500 msec. The vowel /a/ was shorter, being 300 msec long. The fundamental frequency of voicing rose from 112 to 132 Hz in the first 100 msec but then fell to 92 Hz in the last 200 msec. Amplitude of voicing rose to 15 dB by 10 msec, 40 dB by 20 msec, 55 dB by 100 msec, and then fell to 45 dB at 300 msec. The stimuli were balanced for intensity (within 2 dB of each other) by measuring the overall dB(A) value of each stimulus with a sound level meter (Bruel and Kjaer Type 2235), using the A weighted maximum hold function. A final listening check was conducted to ensure equality of loudness.

**Equipment.** The experiment was controlled by computer (Total Peripherals, 386, PC), which also presented the stored, previously synthesized vowel stimuli via a 16-bit digital-to-analog conversion card (Creative Technology SB16). The stimuli were then lowpass filtered at 4200 Hz (Krohn-Hite 3901 elliptical filter) before being directed to the earphones (TDH 39) via a clinical audiometer (Interacoustics AC5). The controlling software randomized the presentation sequence of stimuli within each block, chose the sequence

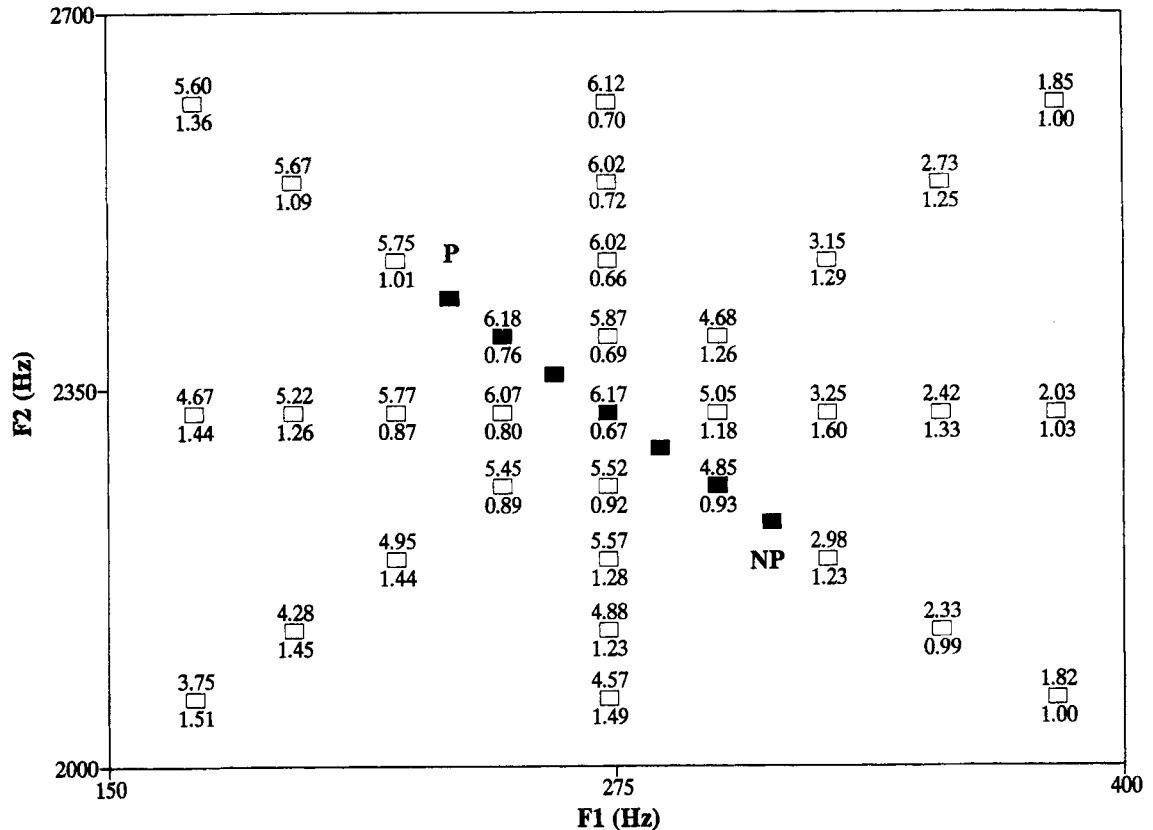


Figure 3. From Experiment 1, the subjects' ratings of each stimulus in the Australian English vowel category /i/. The upper number placed by each stimulus is the mean; the lower number is the standard deviation. Stimulus coordinates are a function of F1 and F2, depicted in hertz. The locations of the prototype (P) and the nonprototype (NP) are shown. The filled squares are the locations of the stimuli on the continuum selected for use in Experiment 3.

of vowel categories for each subject from a Latin square design, and logged all the subjects' responses in a text file.

**Procedure.** The subjects listened to the stimuli presented binaurally via headphones (at 68 dB SPL), seated in a sound-proofed room. The five vowel categories were tested independently, the order taken from a Latin square design. Each subject was presented with all of the 32 vowel stimuli and the prototype once, in one block, as practice. The test session consisted of five blocks in all, with every block comprising one repetition of each stimulus, for a total of 165 trials per category. The subjects were asked to point, using a mouse, to a rating displayed as a 1–7 scale on the computer screen (7 being a *good quality vowel* and 1 being a *poor quality vowel*). For example, the end points of the scale were marked "7 best /ee/" and "1 worst /ee/" with the numbers 2–6 marked on a line between them. Once a subject responded, the next stimulus was presented 500 msec later. The subjects had 5 sec in which to respond; otherwise, a *no response* was recorded, and the next stimulus was presented. After each block of 33 trials, the subjects were given a break from testing until they were ready to proceed. The five ratings for each stimulus were averaged to make a mean rating score for each stimulus. The test session lasted approximately 1 h.

## Results and Discussion

The main finding of this experiment was that listeners were able to rate stimuli as having differing levels of quality, indicating some level of internal structure of the F1–F2 space. This is consistent with previous reports of an internal structure of phonetic categories. Miller and

Volaitis (1989) studied identification and goodness ratings of CVs that varied in VOTs that were representative of the VOTs found when speaking rate was increased. They found that the change in duration by manipulating VOT context for an extreme /pi/–/pi/–/bi/ continuum resulted in a change of identified category membership and perceived goodness.

Figures 2 to 6 show the mean rating scores given to the 32 variants and the prototype in each category in the present study.

It can be seen that, for all categories except for /æ/ and /a/, the stimuli closest to the category center (i.e., our initial best-guess prototype) earned high rating scores. However, these center stimuli were not always given the highest rating, and a number of stimuli in each category attracted equal or higher rating scores. Consider the exceptions. In particular, the preferred stimuli for /æ/ (see Figure 1) were positioned toward the end of the high F1 and F2 space on Vector 2. There are three possible explanations for this. First, the duration of this vowel was shorter than those for the others (300 msec), and duration is an important cue for /æ/ in Australian English (Bernard, 1967; Collins & Blair, 1989; Jones, 1968). The results may have reflected a tradeoff between formant frequencies and duration not seen in the other categories.

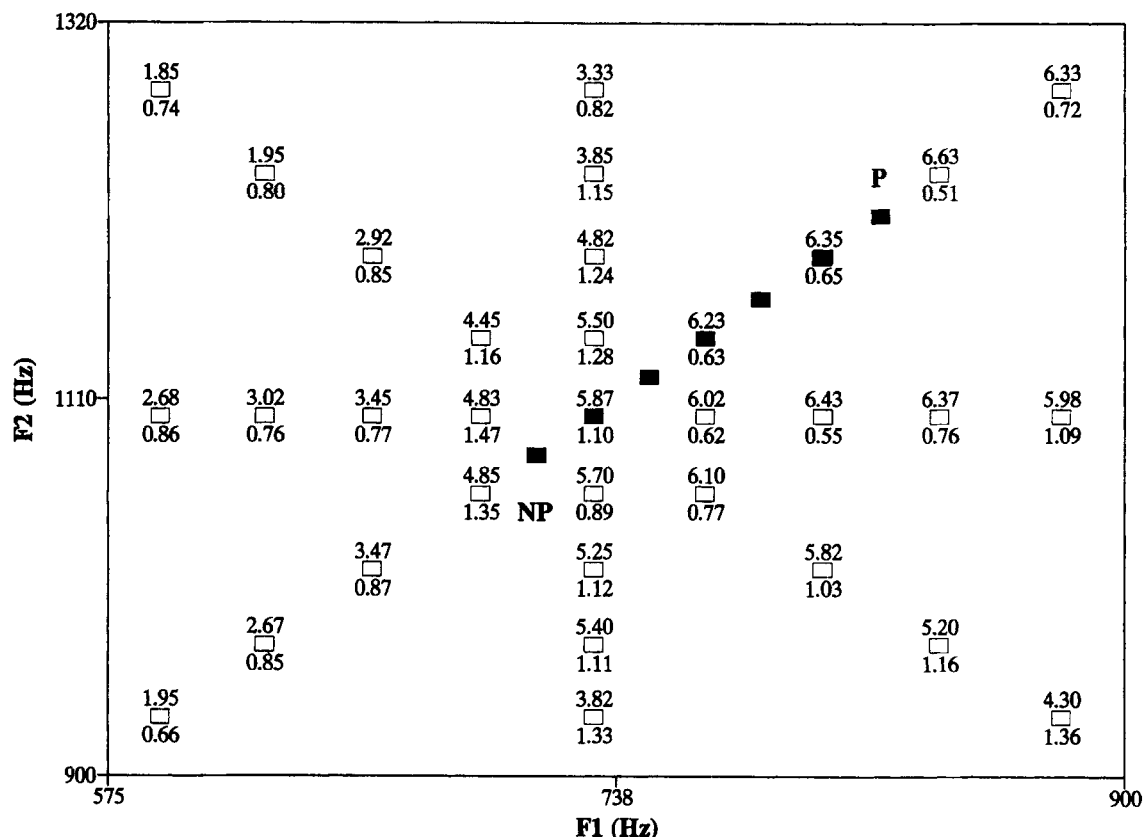


Figure 4. From Experiment 1, the subjects' ratings of each stimulus in the Australian English vowel category /a/. The upper number placed by each stimulus is the mean; the lower number is the standard deviation. Stimulus coordinates are a function of  $F1$  and  $F2$ , depicted in hertz. The locations of the prototype (P) and the nonprototype (NP) are shown. The filled squares are the locations of the stimuli on the continuum selected for use in Experiment 3.

Second, subjects often prefer hyperarticulated versions of vowels when they are presented in isolation. This has been demonstrated for the American vowel /i/ (Johnson, Flemming, & Wright, 1993), and the extent of this preference for any particular vowel is probably dialectally related. The highly rated stimuli in /æ/ sounded as if they were more open, had slightly more nasality, and articulated with more backness than the prototype. These factors may have made them sound more overarticulated than the other categories for Australian listeners. Finally, /æ/ in Australian English is quite variable, with instances of productions approaching /æ<sup>ə</sup>/, /æ:/, and /ɛ/ often being heard (Clark, 1989). These variants tend toward a progressively higher  $F2$  value and, in conjunction with overarticulation, may have been perceived as being good instances of that category. A similar picture is seen in /a/, where the highest ratings were distant from the central prototype, favoring high  $F1$  and  $F2$  values. The vowel /a/ is very stable in Australian English (Clark, 1989) but with some instances leaning toward /eə/ being produced, again a variant where a tendency toward higher  $F2$  values exist. Our rating results for the other vowels are broadly in agreement with previous work in American English, since our initial best-guess prototype for /i/ also attracted a high goodness rating (see Iverson & Kuhl, 1995b; Kuhl,

1991). For /ɔ/ and /ɜ/, more than one stimulus was considered to have high quality ratings that were similar to our best-guess prototype (see also, e.g., Lively & Pisoni, 1997, for their /i/ stimuli).

To assess the performance of different subjects on this task, frequency histograms (number of subjects  $\times$  rating score) for individual stimuli were studied. The subjects' rating scores tended to be more closely clustered where stimulus means were higher (i.e., good quality ratings above 4), and a wider range of rating scores was observed where the mean rating was low. This was taken to reflect a higher level of confidence when rating good stimuli than when rating poor stimuli. A contributing factor to increased standard deviations seemed to be that some individual subjects gave stimuli with a low mean rating extraordinarily high ratings and vice versa.

Kuhl (1991) noted that stimuli situated on the same orbit or radial distance from the prototype attracted similar ratings and that quality declined generally with distance from the central stimulus. In order to investigate this graded structure, separate nonparametric Friedman analyses of variance (ANOVAs) for related samples were applied to the data for the /i/, /ɔ/, and /ɜ/ categories. The /æ/ and /a/ categories were not tested, since it is difficult to discuss symmetry where subjects placed the highest,



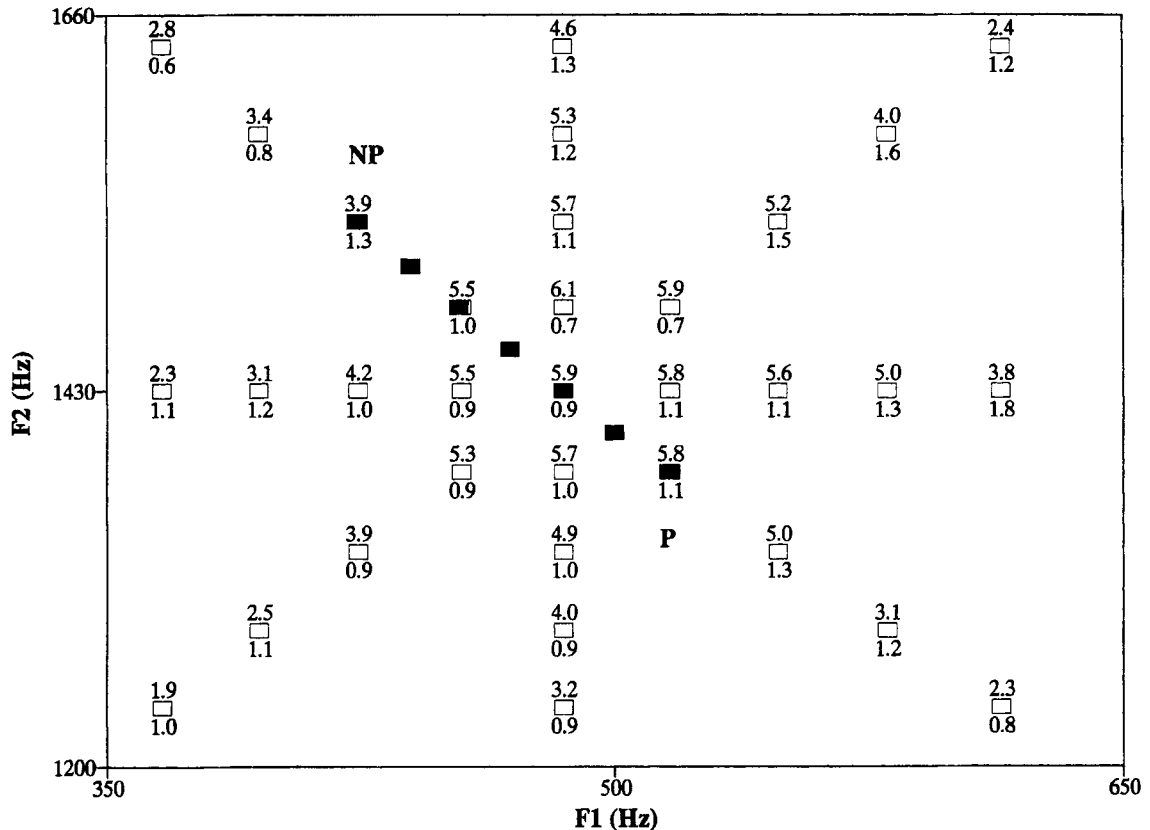


Figure 5. From Experiment 1, the subjects' ratings of each stimulus in the Australian English vowel category / $\epsilon$ /. The upper number placed by each stimulus is the mean; the lower number is the standard deviation. Stimulus coordinates are a function of  $F1$  and  $F2$ , depicted in hertz. The locations of the prototype (P) and the nonprototype (NP) are shown. The filled squares are the locations of the stimuli on the continuum selected for use in Experiment 3.

more prototypical stimuli quite differently from our initial center estimates. The following discussion regarding symmetrical gradation relates to the categories / $i$ /, / $\upsilon$ /, and / $\partial$ /. The mean rating score given to stimuli on the same orbit was the dependent variable, and orbit number (1–4) was the independent variable. Figure 7 shows the mean rating score for each orbit. For these three categories, a significant effect of orbit ( $p < .001$ ) was seen.

Although these results provide evidence of a mean reduction of quality rating as distance increases, scrutinization of Figures 3, 5, and 6 indicates that, in all cases, there are mean group ratings at some point on each orbit where scores are quite disparate. For example, evidence of symmetry might be argued for Orbit 1, for / $i$ /, but not for other orbits (e.g., values on Orbit 2 range from 2.98 to 6.02, with a mean value of 4.75). If the decline in quality was symmetrical, the mean rating scores of the stimuli in each orbit should be the same. In order to test this, further separate nonparametric Friedman ANOVAs for related samples were applied to the eight stimuli on each orbit of the categories / $i$ /, / $\upsilon$ /, and / $\partial$ /. In each of these categories, the mean quality rating of stimuli on the same orbit were highly significantly different at the  $p < .001$  level. Lively and Pisoni (1997) found a similar result with a replication of Kuhl's (1991) prototype and nonprototype / $i$ / stimuli.

It seems, then, that a systematic change in  $F1$  and  $F2$  is not paralleled by a systematic change in quality. Changes in  $F1$  values alone invariably produced a wider range of quality ratings than did changes in  $F2$  alone. Since cues provided by  $F1$  are important in vowel identification (Schouten & Van Hessen, 1992), it would be expected that within-category quality judgments would be affected by manipulation of a spectral cue that influenced vowel identity. Indeed, there is evidence that the physical differences of  $F1$  and  $F2$  can be mapped successfully onto a two-dimensional perceptual vowel space (Kewley-Port & Atal, 1989). This suggests a defined relationship between the physical and the perceptual spaces. Kuhl's (1991) data imply that this relationship is more of a one-to-one function than the present study and others would suggest. Certainly, there is no logical reason to suppose that a phonetic category "center" or prototype has to be in the physical center of a category.

In summary, the internal structure of vowel categories can be described in terms of graded goodness. However, the symmetrical gradation and invariance observed by Kuhl (1991) is not apparent in our data. This finding is consistent with previous work (Lively & Pisoni, 1997). It appears that the measure of goodness apportioned to stimuli by listeners is related to formant values, but the

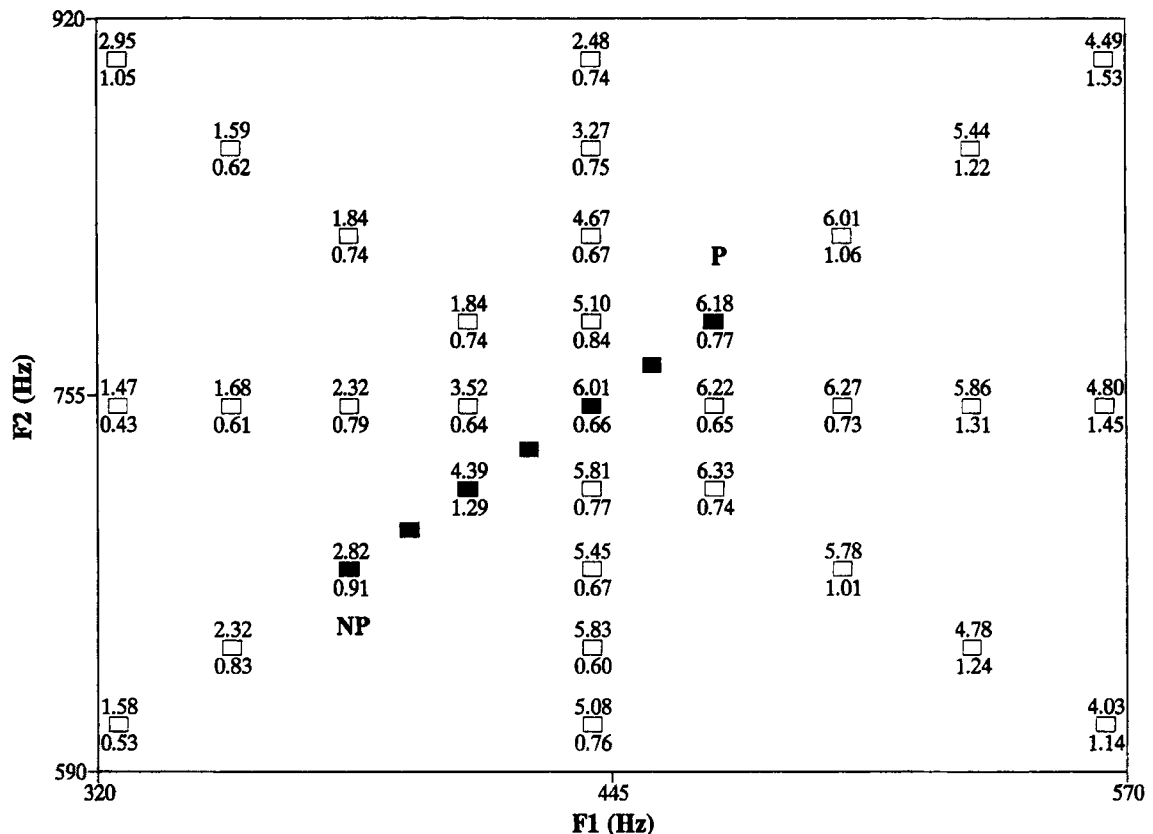


Figure 6. From Experiment 1, the subjects' ratings of each stimulus in the Australian English vowel category /ɔ/. The upper number placed by each stimulus is the mean; the lower number is the standard deviation. Stimulus coordinates are a function of  $F1$  and  $F2$ , depicted in hertz. The locations of the prototype (P) and the nonprototype (NP) are shown. The filled squares are the locations of the stimuli on the continuum selected for use in Experiment 3.

function relating the physical to the perceptual (phonetic) space is not simple.

## EXPERIMENT 2

The results from Experiment 1 provided evidence of an internal structure related, to some extent, to perceived quality and  $F1$  and  $F2$  coordinates. Therefore, it would be expected that higher rated stimuli should also be identified as within-category stimuli. However, this may not be the case with low-rated stimuli. A stimulus may have attracted a mean low rating because it was outside of the category, as was the case for Kuhl's (1991) nonprototype /i/ stimulus. In Experiment 2, the subjects were asked to identify each stimulus in the five Australian English vowel categories as belonging or not belonging to that category, in order to establish category membership of the test stimuli. A second purpose was to identify a non-prototype stimulus with a low rating and a high or, at least, moderate identification score and a prototype stimulus with a high rating and a high identification score, in each category. These stimuli could then be used to investigate discrimination associated with the *perceptual magnet effect* in Experiment 3.

## Method

**Subjects.** The 16 adults who participated in Experiment 1 took part in Experiment 2.

**Procedure.** The stimuli, the equipment, the procedure, and the test environment were the same as those used in the previous experiment, except that, rather than rating each stimulus, the listeners were required to identify each stimulus. For example, they had to decide whether the stimulus they had just heard was /i/ as in the word *peep*. Responses were made by clicking a mouse on one of two boxes on a computer screen labeled "/ee/" or "not-/ee/" (for the /i/ category). The percentage of identification (i.e., percentage of /a/ for vowel /a/) was recorded and averaged across the five blocks to form a mean percentage of identification score for each subject, stimulus, and category.

## Results and Discussion

The mean percentage *in-category* scores for all the subjects for each vowel category are presented in Figures 8–12. The main result of Experiment 2 was that listeners were able to identify a substantial number of the variant vowel stimuli in each category as belonging to that category.

For /i/, /ɜ/, and /ɔ/, the chosen central, prototypical stimulus in each case was identified as belonging to the category approximately 97% of the time. This suggests that the initial choices of prototype formant values were

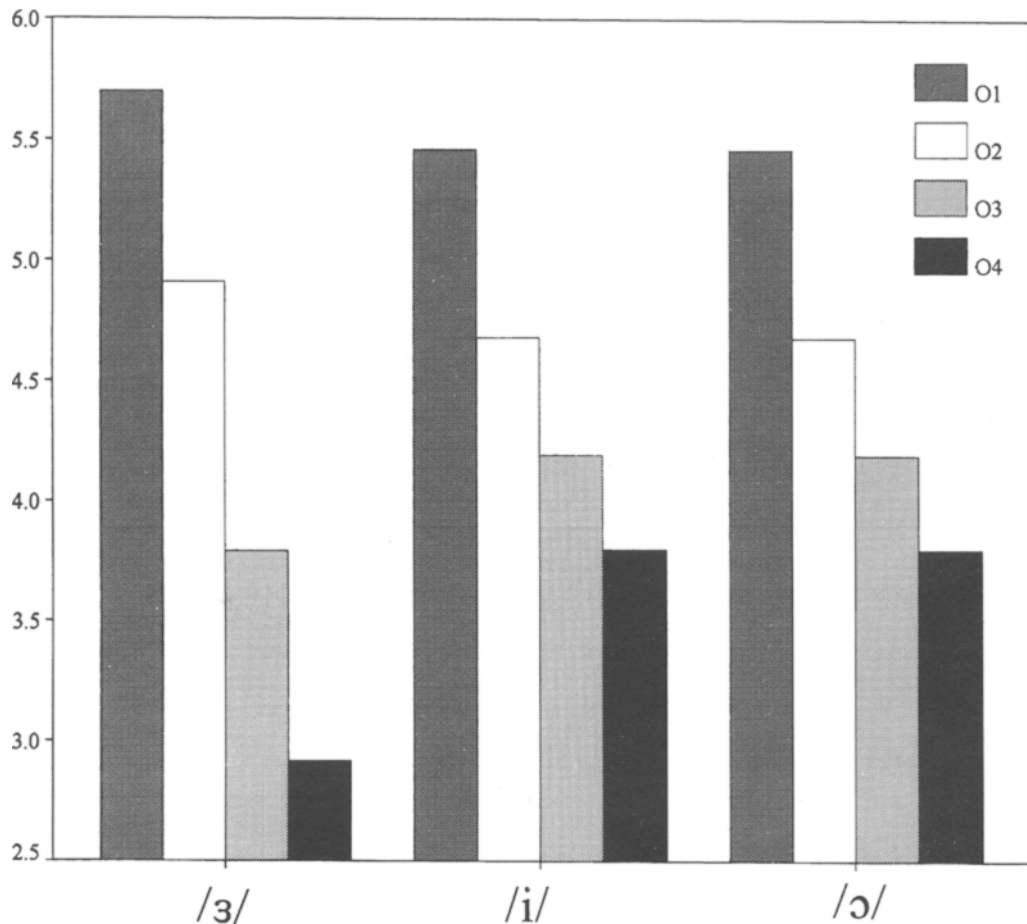


Figure 7. From Experiment 1, the subjects' quality ratings of stimuli from the same orbit averaged to give a mean value for Orbits O1–O4 in each Australian English vowel categories, /ɜ/, /i/, and /ɔ/.

highly representative of the category. In addition, percentage of identification decreased as distance from the center increased. The points in  $F1$  and  $F2$  space, where identification approached 50%, define the category boundaries. The central stimuli of /æ/ and /a/ were identified 37% and 62% of the time, respectively, and received goodness scores of 3.4 and 5.87, respectively, suggesting some disagreement between subjects' choice of a best instance and the characteristic formant values produced in Australian English measured by Bernard (1989). How reliable and relevant Bernard's acoustic specifications of Australian English in the 1960s are to Australian English in the 1990s is of interest. There is evidence of sociolinguistic change in Australia, leading to *general* Australian being spoken more predominantly (Horvath, 1985). In view of this, some disagreement would be expected between subjects' ratings and Bernard's production data. Nevertheless, a good instance (prototype) and a poor instance (nonprototype) stimulus could be identified, from the data, even though they were at locations not predicted by speech production data for these vowels.

The variation seen in the percentage of identification scores reflects subjects' high confidence in responding to stimuli well within categories and outside of categor-

ies but low confidence for stimuli near the 50% boundary. This trend was seen for all the categories. A strong relationship was observed between subjects' ratings (in Experiment 1) and percentage of identification scores for all the vowel stimuli. Since correlation analysis revealed highly significant Spearman's correlation coefficients (.91–.97,  $p < .005$ ), it would seem that listeners' identification judgments, like their rating judgments, rely on the spectral cues that arise from  $F1$  and  $F2$  changes. Listeners had no option other than to identify stimuli as either /æ/ or not /æ/ (for example); therefore, poor quality ratings were reflected in low percentage in-category identification scores.

The present data show that some examples of the stimuli distributions were not identified as belonging to the category in question. Of interest is that this study's 120-mel stimulus in /i/ was not considered by listeners to be an /i/. This is consistent with other reports (Iverson & Kuhl, 1995b; Sussman & Lauckner-Morano, 1995). In a similar experiment, Lively and Pisoni (1997) noted that, for /i/, Kuhl's (1991) 120-mel stimulus was identified quite differently, depending on whether the stimulus was presented in the context of the prototype or of the nonprototype distribution of stimuli. In Experiment 2, category member-

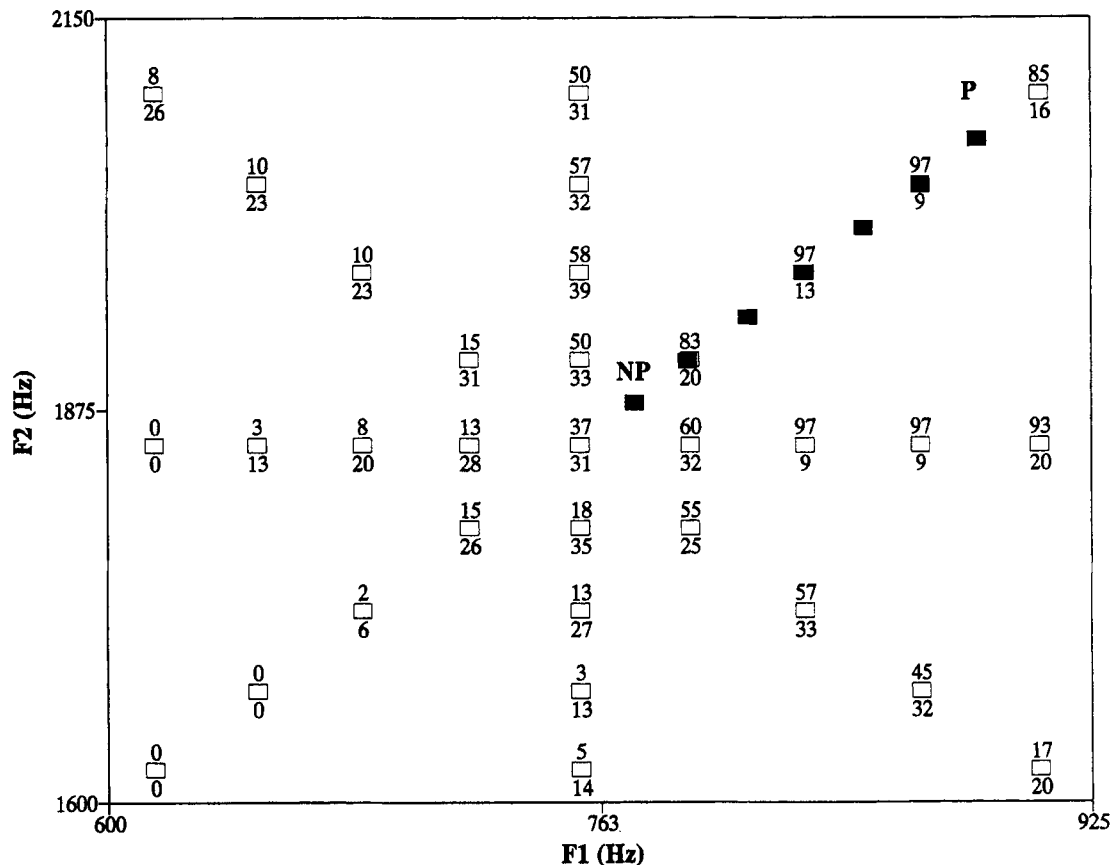


Figure 8. From Experiment 2, the subjects' percentage identification score for each stimulus in the Australian English vowel category /æ/. The upper number placed by each stimulus is the mean; the lower number is the standard deviation. The stimulus coordinates are a function of  $F1$  and  $F2$ , depicted in hertz. The locations of the prototype (P) and non-prototype (NP) are shown in each category. The filled squares are the locations of the stimuli on the continuum selected for use in Experiment 3.

ship was established for stimuli in each vowel category within the context of its 32-stimulus array. Initial comparison of the results from Experiments 1 and 2 enabled a number of prototype and nonprototype stimuli to be selected. Part of Experiment 3 included identification of a subset of these 32 stimuli in order to establish whether set size would have an effect on subjects' labeling of identical stimuli. This smaller set of stimuli (7 in each category) was used to test for the perceptual magnet effect in a discrimination task carried out in Experiment 3.

### EXPERIMENT 3

In this experiment, discrimination of the selected prototype and nonprototype stimuli from variants synthesized between them was investigated. The first and main objective was to demonstrate poorer discrimination performance with a prototype as a referent than with a nonprototype as a referent. This experiment was designed to facilitate discrimination and to better control for the effects of stimulus set context. It was therefore thought to be a fairer test of the perceptual magnet effect than investigations of previous studies.

Second, the influence of listeners' phonetic experience was considered. Kuhl's (1991) subjects were experienced in phonetics, whereas Sussman and Lauckner-Morano (1995) and Lively and Pisoni (1997) used subjects untrained in phonetics. Iverson and Kuhl (1995b) tested a mixture of phonetically trained and untrained subjects in their rating and identification tasks but untrained subjects for their discrimination task. In the present study, a phonetically experienced and a phonetically naive group of subjects were tested. The objective was to investigate the hypothesis that training in phonetics can change subjects' listening strategy or their internal representation of vowel prototypicality. Therefore, the plausibility of inferring magnet effects from discrimination data might be questionable. A point to note here is that, if this perceptual effect is dependent on experience, exactly how robust is it likely to be?

### Method

**Subjects.** Thirty-two adult subjects (10 males and 22 females), 18–29 years of age (mean = 22.4), took part in this experiment. None had taken part in Experiment 1 or 2. The group was drawn from undergraduate and postgraduate students of the University of Queensland. Half of the subjects had taken courses in phonetics, transcription, speech pathology, and linguistics; the remainder were

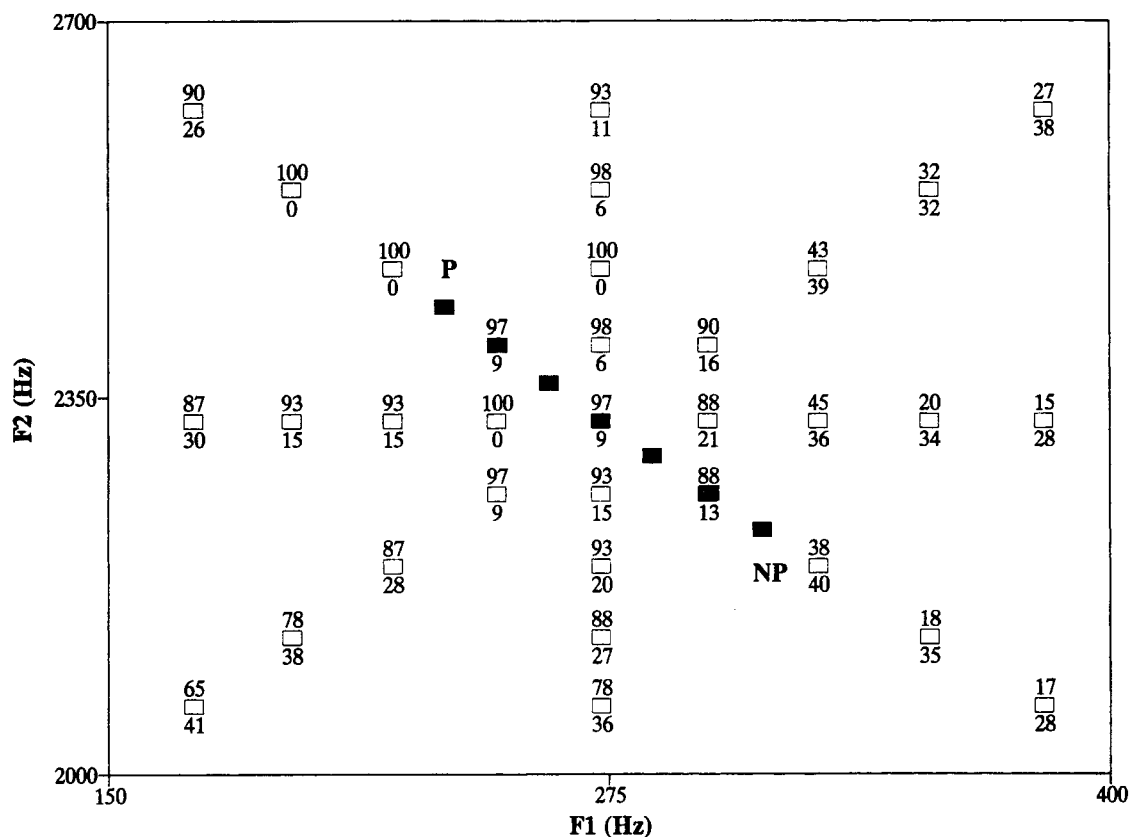


Figure 9. From Experiment 2, the subjects' percentage identification score for each stimulus in the Australian English vowel category /i/. The upper number placed by each stimulus is the mean; the lower number is the standard deviation. The stimulus coordinates are a function of  $F_1$  and  $F_2$ , depicted in hertz. The locations of the prototype (P) and non-prototype (NP) are shown in each category. The filled squares are the locations of the stimuli on the continuum selected for use in Experiment 3.

naive in these areas. All were native Australian English speakers and were right-handed. All the subjects passed a hearing screen at 20 dBHL at octave intervals between 250 and 8000 Hz. They had no reported history of hearing or speech disorder.

**Stimuli.** From the pool of 32 stimuli for each category used in Experiments 1 and 2, a prototype, a nonprototype, and five variants were selected. In order to fit these variants into an equal number of five psychoacoustic steps on a continuum between the prototype and the nonprototype stimuli, the step size was reduced to 15 mels, with new stimuli synthesized in the intermediate positions. By examining the rating and identification scores from Experiments 1 and 2, a continuum of the variants along one diagonal or another was chosen (see the filled blocks marked on Figures 2–6 and 8–12). The five-stimulus continuum was bounded by a prototype stimulus and a nonprototype stimulus.

The main criteria for choosing the prototype and nonprototype stimuli were that they received identification scores that were approximately 40% apart and rating scores that were at least 2 points apart, and that they were separated by six 15-mel steps on any stimulus vector with the same direction. In some cases, these values were interpolated between the stimuli in order to choose an appropriate prototype or nonprototype stimulus that allowed consistent mel intervals. The resulting prototype stimuli were identified as within-category stimuli more than 90% of the time and were rated 5.8 and above. The nonprototype stimuli were identified between 42% and 63% of the time and were rated between 2.8 and 4.3. The nonprototype identification scores given that were below 50% were for the vowels /ɜ/ (42%) and /ɔ/ (46%).

**Procedure.** The equipment used in Experiment 3 was identical to that in Experiments 1 and 2. Discrimination of the five categories of vowel stimuli (/i/, /æ/, /a/, /ɜ/, and /ɔ/) was tested in two conditions. In the first condition, each of the five vowel stimuli was compared with the prototype. In the second condition, each of the five vowel stimuli was compared with the nonprototype selected from that category. A *same/different* paradigm was used throughout.

**Same/different paradigm.** On each trial, either the prototype or the nonprototype stimulus (depending on the condition) was paired with one of the variants from the continuum or with itself. The ISI was 250 msec. The subjects sat looking at a computer screen on which two boxes, one marked *same* the other marked *different*, were displayed. On listening to a trial, the subjects were instructed to use the computer mouse to select the *same* box if they considered the trial to consist of two identical stimuli or the *different* box if not. Responses were logged by the computer. If no response was made after 4 sec, the next trial was presented and a *no response* recorded. The subjects were asked to respond as quickly as possible, since their RT would be logged. Feedback was given on each trial for both hits and correct rejections. There were 120 trials in all, presented in 12 blocks of 10 trials, with *same* trials being presented with a probability of .5. Within each block, each of the five stimuli from a continuum was presented once, in a *different* trial.

Two extra blocks were presented for practice. Trials were randomized within each block, and the order of block presentation was randomized. All stimuli, presentation order, and response logging were under computer control, running CSRE45 software (Avaaz Innovations, Inc.).

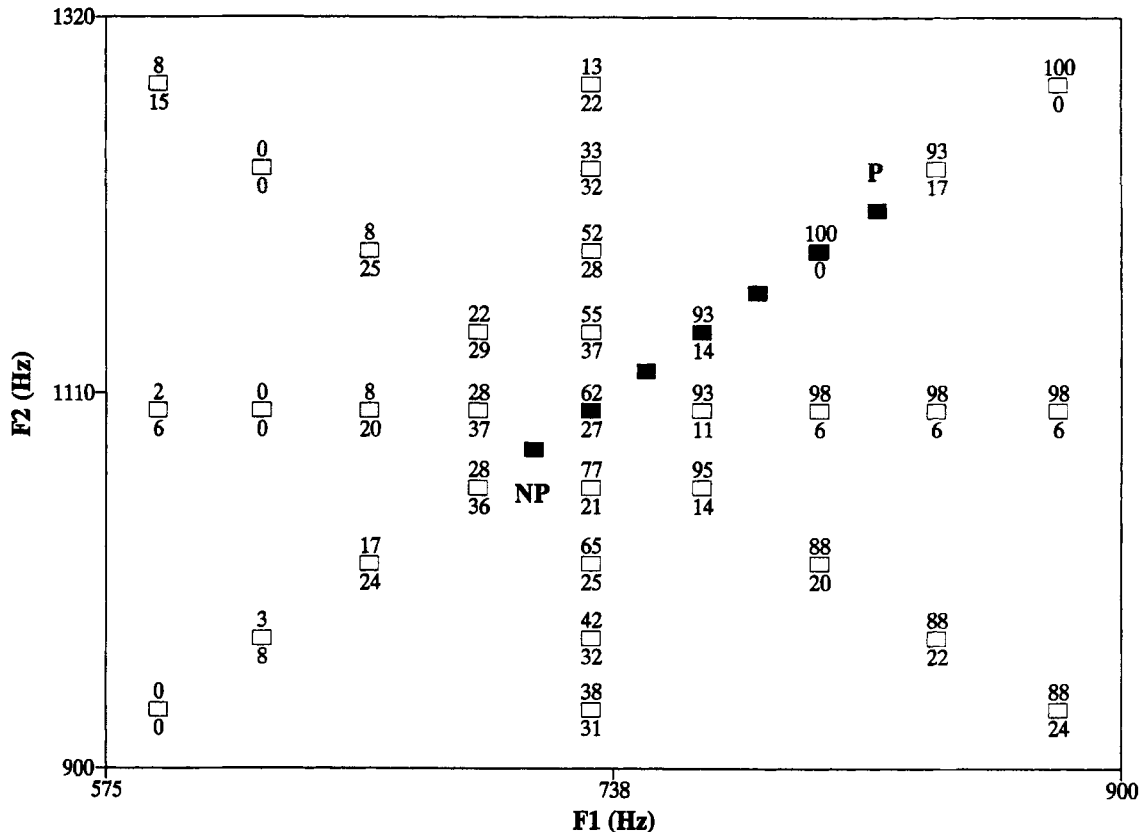


Figure 10. From Experiment 2, the subjects' percentage identification score for each stimulus in the Australian English vowel category /a/. The upper number placed by each stimulus is the mean; the lower number is the standard deviation. The stimulus coordinates are a function of  $F1$  and  $F2$ , depicted in hertz. The locations of the prototype (P) and nonprototype (NP) are shown in each category. The filled squares are the locations of the stimuli on the continuum selected for use in Experiment 3.

From the collected data, hit (*different* when stimuli were different) and false alarm (*different* when stimuli were the same) rates were analyzed, and the unbiased discrimination measure  $d'$  (Macmillan & Creelman, 1991) was calculated. In addition, the mean response RTs for hits for each stimulus on a continuum were calculated.

**Identification.** The data from Experiments 1 and 2 were used to construct the vowel continua used in Experiment 3. In the selection process, it was apparent that the nonprototype stimuli were not always identified as belonging to a category 50% or more of the time, and therefore, discrimination may have been based on across-boundary decisions. However, owing to different stimulus set sizes and context influences, the identification data in Experiments 1 and 2 may not have reflected similar identification scores with the smaller stimulus set size (7 as opposed to 32 stimuli) in Experiment 3. In addition, some of the intermediate stimuli chosen were extrapolated and not implicitly identified in Experiment 2. In view of this, 14 of the subjects (7 naive and 7 experienced) who took part in the third discrimination experiment also identified the same stimuli, using the procedure described in Experiment 2. The results are reported below.

## Results and Discussion

**Identification Data.** The identification results are shown in Table 2. In Experiments 1 and 2, we tested naive listeners and selected the materials for Experiment 3 on the basis of their results. Since Sussman and Lauckner-Morano (1995) found evidence that naive and experienced groups differed in their choice of prototypes, separate

nonparametric Friedman ANOVAs were carried out on data from each vowel category. Factors were group (naive vs. experienced) and identification score for each stimulus. No significant differences of group membership were seen at the  $p < .05$  level. The cause of the disparity between Sussman and Lauckner-Morano's data and the present data is not particularly clear. By comparing the present study's naive listeners in Experiments 1 and 2 with the experienced listeners in the identification part of Experiment 3, we see that different levels of linguistic experience appear not to change perceptions of category membership.

The second point of interest is that all the subjects identified stimuli in the smaller set differently than when exposed to a much larger stimulus set. This was much more obvious around the nonprototype end of the continua, implying that, given less choice, listeners tend to accept poorer quality category members as being within category. As in the second experiment, a number of stimuli were identified as within-category 95% or more of the time. These were generally within 30 mels of the chosen prototype. Subsequently, a satisfactory range between prototype and nonprototype remained in all continua.

The third important point is that the subjects who had discriminated the smaller set of Australian vowels now



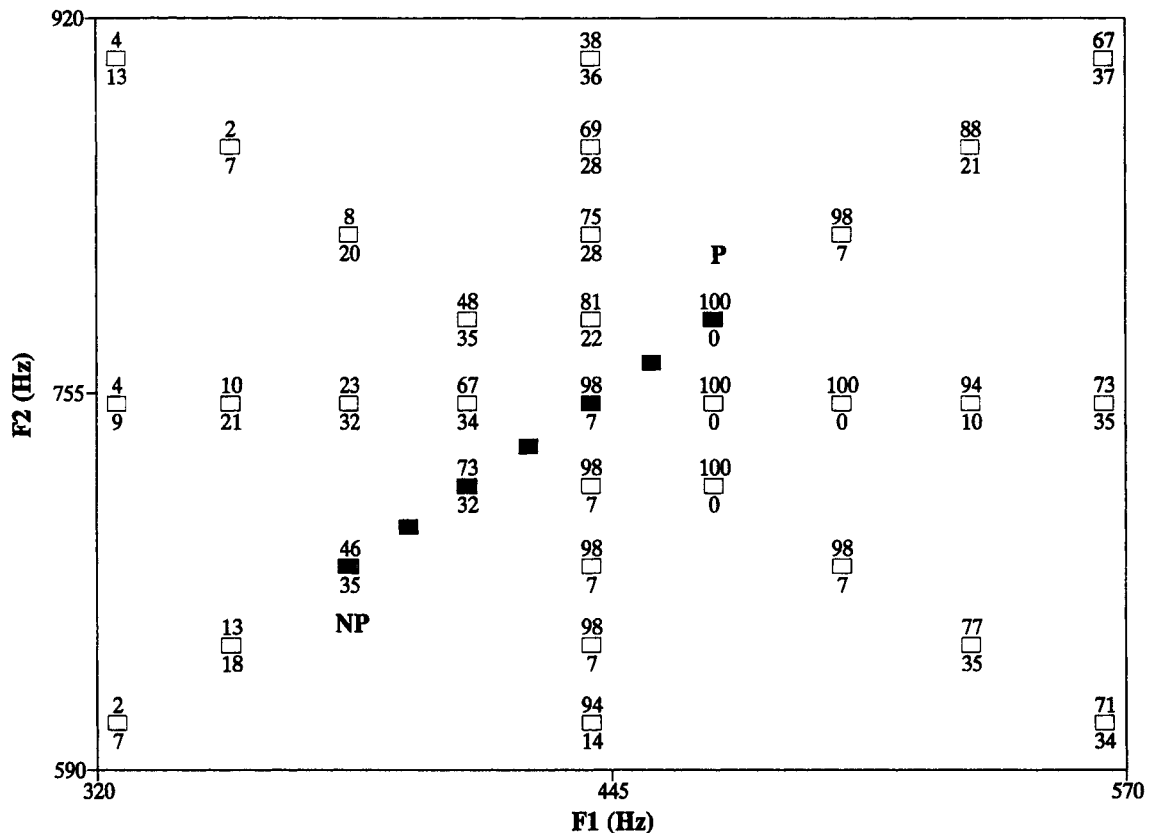


Figure 12. From Experiment 2, the subjects' percentage identification score for each stimulus in the Australian English vowel category /ɔ/. The upper number placed by each stimulus is the mean; the lower number is the standard deviation. The stimulus coordinates are a function of  $F_1$  and  $F_2$ , depicted in hertz. The locations of the prototype (P) and nonprototype (NP) are shown in each category. The filled squares are the locations of the stimuli on the continuum selected for use in Experiment 3.

is similar in each condition. This is not the case for /ɔ/ and /i/. Investigation of the simple effects associated with condition in these categories revealed a significant difference in performance at the N2P2 distance for both /ɔ/ and /i/ ( $F = 32.4, p < .001$ , and  $F = 16.63, p < .001$ , respectively). For /ɔ/, this difference in performance is in the wrong direction to support a perceptual magnet effect (i.e., performance in the prototype condition was better than that in the nonprototype condition). However, performance for /i/ was poorer for the P condition and, therefore, is in the right direction to support a perceptual magnet effect. These differences undoubtedly contributed to the overall interaction of the main effects. On their own, they provide minimal support for a perceptual magnet effect, since a robust effect should be present in all the categories.

A comparison of the discrimination data for /i/ from this study with discrimination data for /i/ from Iverson and Kuhl (1995b) is possible. Iverson and colleagues tested discrimination of /i/ variants corresponding to the high  $F_2$ /low  $F_1$ , low  $F_2$ /high  $F_1$  diagonal of Kuhl's (1991) stimuli. Their "right of prototype" and "left of prototype" stimuli correspond to this study's continua of stimuli between the prototype and the nonprototype. Iverson and Kuhl (1995b) found that their subjects' performance was not significantly poorer (at the  $p < .05$  level,

although it was just significant at the  $p = .11$  level) when discriminating stimuli right of their prototype from the prototype stimulus than when discriminating stimuli left of their nonprototype from the nonprototype stimulus. Thus, no substantial evidence of a shrinking of perceptual space consistent with magnet effects was shown. Iverson and Kuhl (1995b) explain that this effect may be less prevalent in adults, since previous work with children (Kuhl, 1991) had shown a much larger effect. In the present study, no supportive evidence for a perceptual magnet effect was found in any of the discrimination data for the five vowel categories, including /i/.

Further comparison with Sussman and Lauckner-Morano's (1995) data is also possible. Sussman and col-

Table 2  
Identification Scores for the Continua for the Five Australian Vowels Used in Experiment 3

Continuum	/ɔ/	/ɜ/	/a/	/æ/	/i/
DSP	94	100	97	97	96
DS1	93	97	97	94	90
DS2	93	100	94	99	84
DS3	81	99	89	94	76
DS4	57	94	86	86	63
DS5	54	87	66	69	51
DNP	54	69	61	53	51



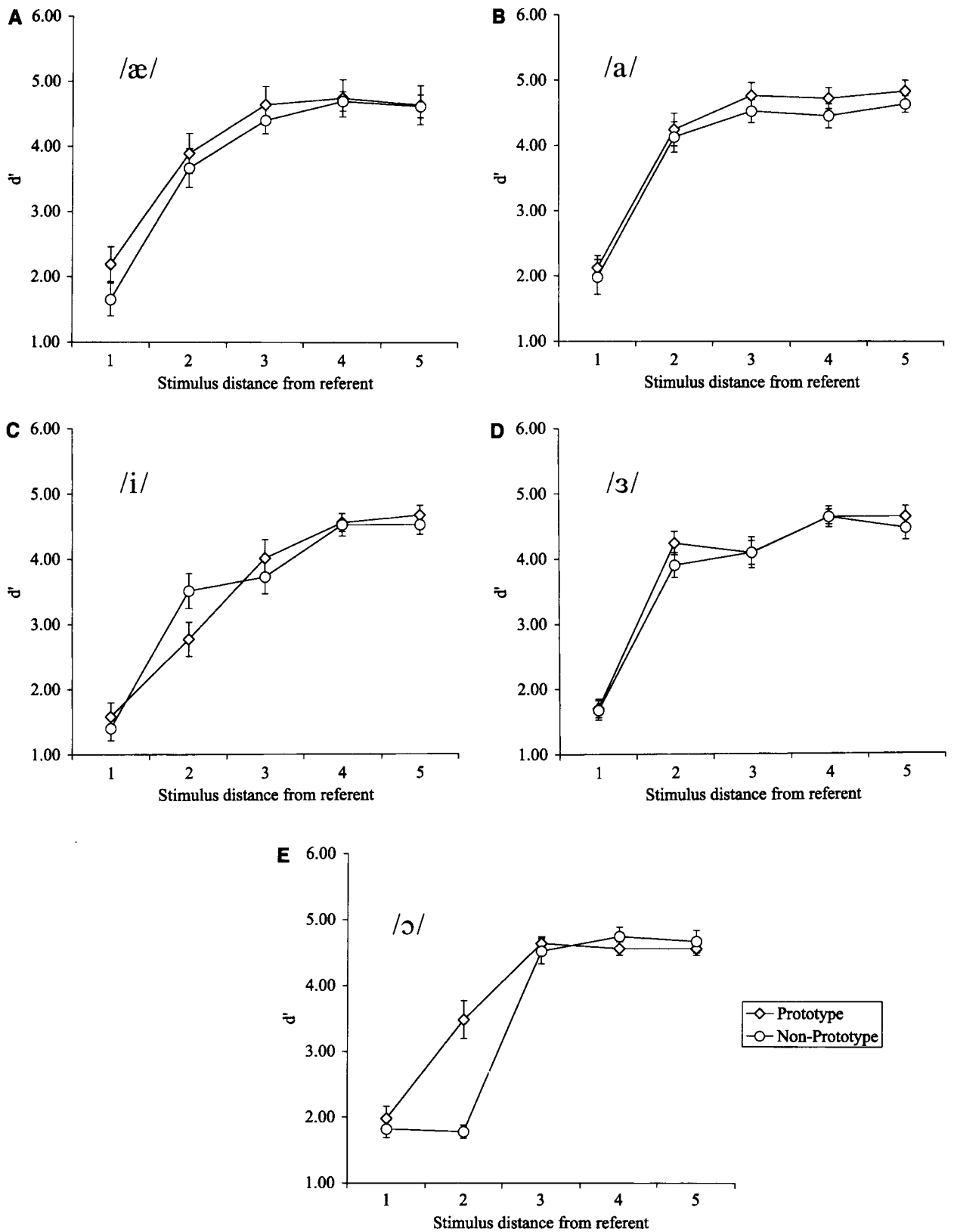


Figure 13. Results from Experiment 3 for the five Australian English vowels tested. The vertical axis is mean  $d'$ , depicting discrimination performance. The five intervals in physical space between the prototype and the nonprototype referents are shown on the horizontal axis.

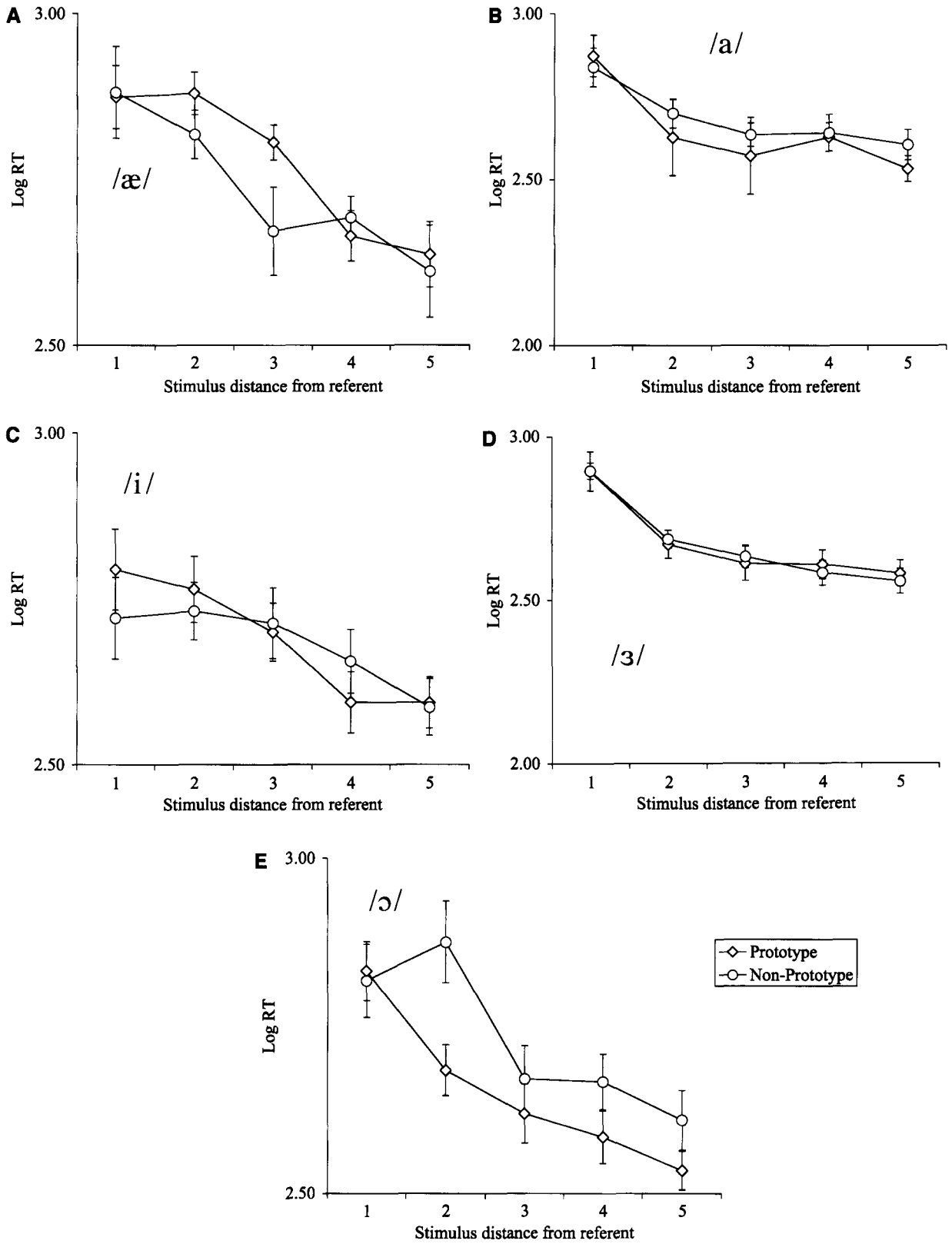


Figure 14. Results from Experiment 3 for the five Australian English vowels tested. The vertical axis is the mean  $\log_{10}$ -transformed reaction time (LogRT) for hits; the horizontal axis gives each of the five intervals of physical space between the prototype and the nonprototype referents.

leagues measured discrimination ( $d'$ ) by 40 adults on five different continua also taken from the high  $F2$ /low  $F1$ , low  $F2$ /high  $F1$  diagonal (vector 4) of Kuhl's (1991) stimuli. There are data for two groups from Sussman's study that can be compared with the present data: first, their prototype (P) group, which discriminated between a prototype /i/ and four stimuli separated by 15-mel steps increasing in  $F1$  and decreasing in  $F2$  along Kuhl's Vector 4; second, their 75-up group, which discriminated between a nonprototype stimulus, 75 mels away from the prototype along Kuhl's Vector 4, and four stimuli toward the prototype in 15-mel steps. Sussman's P and 75-up groups are, therefore, analogous to the present study's /i/ prototype condition and nonprototype condition, respectively. Sussman and Lauckner-Morano found that discrimination was significantly poorer in the group that had the prototype for a referent than in the group that had a nonprototype, the 75-up stimulus, as a referent. In the present study, no such differences were found.

**Reaction time for hits.** In view of the failure to demonstrate a definitive magnet effect by measuring discrimination, it was thought that studying subjects' RTs in each of the prototype and nonprototype conditions might provide evidence for the perceptual magnet effect. RTs have been shown to be more sensitive to within-category acoustic differences than are discrimination measures (Pisoni & Tash, 1974). Multiple linear regression analysis measured the mahalanobis distance for each data point in the RT data, and there were no significant outliers exceeding the chi-squared limit ( $\chi^2 = 19.46, p < .01$ ). Some minor evidence of skewness was present in the RT data; therefore the  $\log_{10}$  transform was taken and used in this analysis. In the following text, RT means the  $\log_{10}$ -transformed RT. The data collected from the two groups (naive and non-naive subjects) were analyzed by running an ANOVA (group  $\times$  condition). Again, no significant differences were found between the groups, and the data were pooled for subsequent analysis.

The mean RT was calculated for each subject, for each of the five acoustic distances from the prototype and nonprototype conditions in each vowel category (see Figure 14). The perceptual magnet effect predicts that, in the prototype condition, discrimination should be poorer than in the nonprototype condition. Therefore, increased RTs where the prototype is a referent were expected. A fully repeated measures ANOVA, with condition (prototype or nonprototype) and distance (N1-P5) as factors, was applied to the RT data in each category, separately. A significant main effect of condition ( $F = 3.85, p = .059$ ) and an interaction between condition and distance ( $F = 2.90, p < .05$ ) was seen for /ɔ/ only. As was expected, there was a general trend in all the vowel categories for RT to increase significantly ( $p < .001$ ) as the acoustic distance became smaller—that is, as the task became more difficult. This indicates that, for categories other than /ɔ/, the rate at which RT varied with distance was similar, irrespective of whether the prototype or the nonprototype was the referent. Investigation of the simple effects associated with condition in /ɔ/ revealed that subjects' RTs were significantly longer in

the nonprototype condition than in the prototype condition for the N1P1 distance only. The perceptual magnet effect would predict longer RTs for the prototype condition, since other members of the stimulus set are rendered more similar and, therefore, less discriminable.

In summary, the important result is that RTs to the stimuli in the *same/different* task were independent of referent quality, and, therefore, the data do not provide evidence for the perceptual magnet effect.

## CONCLUSIONS

The results of the present investigation do not support the presence of a perceptual magnet effect in the five Australian English vowels studied. In Experiment 1, a graded structure was observed in all the categories. This is consistent with previous data (Kuhl, 1991; Lively & Pisoni, 1997; Sussman & Lauckner-Morano, 1995). However, in the categories where there were highly rated central stimuli (e.g., /i/, /ɔ/, and /ɜ/), a symmetrical degradation in quality ratings, as described in Kuhl's (1991) data, was not seen. Stimulus ratings were not invariant, as Kuhl had suggested. In addition, comparison of the identification data in Experiment 2 with the identification data in Experiment 3 support the contention that stimulus set size influences judgments of category membership. These findings are consistent with previous research (Lively & Pisoni, 1997). In Experiment 3, neither RT data nor  $d'$  measures could reveal a perceptual magnet effect. This was so even though test stimuli were within-category, free from the influences of context, and presented in a paradigm designed to promote discrimination of spectral cues.

## REFERENCES

- BERNARD, J. R. (1967). Length and the identification of Australian English vowels. *Journal of the Australasian Universities Language & Literature Association*, 27, 37-58.
- BERNARD, J. R. (1989). Quantitative aspects of the sounds of Australian English. In P. Collins & D. Blair (Eds.), *Australian English: The language of a new society* (pp. 187-204). St. Lucia, Brisbane, Australia: Queensland University Press.
- CLARK, J. E. (1989). Some proposals for a revised phonetic transcription of Australian English. In P. Collins & D. Blair (Eds.), *Australian English: The language of a new society* (pp. 205-213). St. Lucia, Brisbane, Australia: Queensland University Press.
- COLLINS, P., & BLAIR, D. (1989). *Australian English: The language of a new society*. St. Lucia, Brisbane, Australia: Queensland University Press.
- EIMAS, P., & CORBIT, J. (1973). Selective adaptation of linguistic feature detectors. *Cognitive Psychology*, 4, 99-109.
- FANT, G. (1973). *Speech sounds and features*. Cambridge, MA: MIT Press.
- FLANAGAN, J. L. (1957). Estimates of the maximum precision necessary in quantizing certain dimensions of vowel sounds. *Journal of the Acoustical Society of America*, 29, 533-534.
- FLEGE, J. E., SCHMIDT, A. M., & WHARTON, G. (1996). Age of learning affects rate-dependent processing of stops in a second language. *Phonetica*, 53, 143-161.
- FUJISAKI, H., & KAWASHIMA, T. (1969). On the modes and mechanisms of speech perception. *Annual Report of the Engineering Research Institute* (University of Tokyo, Faculty of Engineering), 28, 67-73.
- GRIESER, D., & KUHL, P. K. (1989). Categorization of speech by infants: Support for speech-sound prototypes. *Developmental Psychology*, 25, 577-588.

- HORVATH, B. M. (1985). *Variation in Australian English: The sociolects of Sydney*. Cambridge: Cambridge University Press.
- IVERSON, P., & KUHL, P. K. (1995a). 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.
- IVERSON, P., & KUHL, P. K. (1995b). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *Journal of the Acoustical Society of America*, **97**, 553-562.
- JOHNSON, K., FLEMMING, E., & WRIGHT, R. (1993). The hyper-space effect: Phonetic targets are hyperarticulated. *Language*, **69**, 505-528.
- JONES, A. I. (1968). *Phonetics: A phonological introduction*. Sydney: University of Sydney.
- KEWLEY-PORT, D., & ATAL, S. B. (1989). Perceptual differences between vowels located in a limited phonetic space. *Journal of the Acoustical Society of America*, **85**, 1726-1740.
- KLATT, D. H. (1980). *Speech perception: A model of acoustic phonetic analysis and lexical access*. Hillsdale, NJ: Erlbaum.
- 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). Psychoacoustics and speech perception: Internal standards, perceptual anchors, and prototypes. In L. A. Werner & E. W. Rubel (Eds.), *Developmental psychoacoustics* (pp. 293-332). Washington, DC: American Psychological Society.
- KUHL, P. K. (1993). Early linguistic experience and phonetic perception: Implications for theories of developmental speech perception. *Journal of Phonetics*, **21**, 125-139.
- KUHL, P. K. (1994). Learning and representation in speech and language. *Current Opinion in Neurobiology*, **4**, 812-822.
- KUHL, P. K., & MELTZOFF, A. N. (1996). Infant vocalizations in response to speech: Vocal imitation and developmental change. *Journal of the Acoustical Society of America*, **100**, 2425-2438.
- 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.
- LIVELY, S. E., & PISONI, D. B. (1997). On prototypes and phonetic categories: A critical magnet effect in speech perception. *Journal of Experimental Psychology: Human Perception & Performance*, **23**, 1665-1679.
- MACMILLAN, N. A., & CREELMAN, C. D. (1991). *Detection theory: A user's guide*. New York: Cambridge University Press.
- MASSARO, D., & COHEN, M. (1991). Integration versus interactive activation: The joint influence of stimulus and context in perception. *Cognitive Psychology*, **23**, 558-614.
- MEDIN, D. L., ALTOM, M. W., & MURPHY, T. D. (1984). Given versus induced category representations: Use of prototype and exemplar information in classification. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **10**, 333-352.
- MEDIN, D. L., & SCHAFFER, M. M. (1978). Context theory of classification theory. *Psychological Review*, **85**, 207-238.
- MILLER, J. L. (1977). Properties of feature detectors for VOT: The voiceless channel of analysis. *Journal of the Acoustical Society of America*, **62**, 641-648.
- MILLER, J. L. (1994). On the internal structure of phonetic categories: A progress report. *Cognition*, **50**, 271-285.
- MILLER, J. L., CONNINE, C. M., SCHERMER, T. M., & KLUENDER, K. R. (1983). A possible auditory basis for internal structure of phonetic categories. *Journal of the Acoustical Society of America*, **73**, 2124-2132.
- MILLER, J. L., & EIMAS, P. D. (1996). Internal structure of voicing categories in early infancy. *Perception & Psychophysics*, **58**, 1157-1167.
- MILLER, J. L., & VOLAITIS, L. E. (1989). Effect of speaking rate on the perceptual structure of a phonetic category. *Perception & Psychophysics*, **46**, 505-512.
- MITCHELL, A. G., & DELBRIDGE, A. (1965). *The speech of Australian adolescents*. Sydney: Angus & Robertson.
- NOSOFSKY, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, **115**, 39-57.
- NOSOFSKY, R. M. (1987). Attention and learning processes in identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **13**, 87-108.
- NOSOFSKY, R. M. (1991). Stimulus bias, asymmetric similarity, and classification. *Cognitive Psychology*, **23**, 94-140.
- PISONI, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics*, **13**, 253-260.
- PISONI, D. B., & TASH, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics*, **15**, 285-290.
- 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.
- 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.

## NOTES

1. According to prototype theory, categories are represented by the central tendency of the category members along each attribute. This category representation, in terms of average or typical attributes, is referred to as the *prototype*. For categorization, an incoming stimulus is considered to be a category member if it is judged to be more similar to that category's prototype than to other category prototypes (Medin, Altom, & Murphy, 1984, p. 334).

In contrast, exemplar models are based on information that is actively retrieved from individual members (exemplars) of a category stored in memory. The stimulus to be categorized acts as a retrieval cue that interrogates the membership of one or more exemplars. Classification is then based on the membership of the retrieved exemplars, rather than on summary information about typical attributes (Medin & Schaffer, 1978).

It is not within the scope of this article to discuss the relative merits of each model. However, theoretically, both can account for the issue with which this article is concerned, which is that not all instances of a category are perceived as being equally good category members (see, e.g., Nosofsky, 1986, 1987). Kuhl (1991) refers to prototypes as a way of describing an internal representation of speech abstracted by a listener and does not intend to make a distinction between an exemplar and a prototype model of categorization. The present article has the same intention.

2. The perceptual magnet effect predicts differences in discrimination that are nonlinearly related to changes in the physical dimensions of a stimulus set. In order to show evidence of this perceptual warping, the physical dimension must be understood. Otherwise, any nonlinearity seen in the perceptual data cannot be directly attributed to perception but may be related to nonlinearity in the physical dimension. In this article and in others (e.g., Kuhl, 1991; Iverson & Kuhl, 1995b; Lively & Pisoni, 1997; Sussman & Lauckner-Morano, 1995), the formant frequencies of synthetic vowels were manipulated to form a physical dimension of vowels that differed only in the first and second formants. Since these changes reflect systematic simulated alterations of the human vocal tract, the influences on the overall spectra of the stimuli on the physical dimension is also relatively systematic. Spectral analysis of the test materials shows an analogous systematic change of formant peaks. Given that pitch is a cue for discrimination between two stimuli in the same category and with formants being at the most some 75 mels apart, the mel scale would seem a more equable metric than, say, a frequency scale. We assume, then, that, in the absence of empirical evidence to the contrary, our stimuli (and those of other studies) are physically equally spaced on the mel scale.