Published in final edited form as:

J Exp Psychol Hum Percept Perform. 1997 December ; 23(6): 1665–1679.

On Prototypes and Phonetic Categories: A Critical Assessment of the Perceptual Magnet Effect in Speech Perception

Scott E. Lively and

Human Factors and Media Prototyping, Ameritech, Hoffman Estates, Illinois

David B. Pisoni

Department of Psychology, Indiana University Bloomington

Abstract

According to P. K. Kuhl (1991), a perceptual magnet effect occurs when discrimination accuracy is lower among better instances of a phonetic category than among poorer instances. Three experiments examined the perceptual magnet effect for the vowel /i/. In Experiment 1, participants rated some examples of /i/ as better instances of the category than others. In Experiment 2, no perceptual magnet effect was observed with materials based on Kuhl's tokens of /i/ or with items normed for each participant. In Experiment 3, participants labeled the vowels developed from Kuhl's test set. Many of the vowels in the nonprototype /i/ condition were not categorized as /i/. This finding suggests that the comparisons obtained in Kuhl's original study spanned different phonetic categories.

Prototype theory has had an important influence on theorizing and experimentation in many subdisciplines of cognitive and perceptual psychology (Smith & Medin, 1981). The theory proposes that categories are organized around abstract summary representations stored in long-term memory. Representations are composed of ideal features or dimensional values for objects in the category (Oden & Massaro, 1978; Smith & Medin, 1981). Thus, prototypes are abstracted by the perceiver rather than realized in the physical environment. The appeal of this approach to categorization is that prototypes can serve as the cognitive reference points against which actual items are judged (Rosch, 1975). Objects that are similar to the prototype are judged to be better instances of the category than items that are less similar to the prototype. This suggests that members of a category form a gradient of typicality in which some items are more prototypical than others (Lakoff, 1987; Rosch, 1975). For example, robins and ostriches are both birds. However, robins are better examples of birds because robins are more similar to other birds. Thus, robins are better approximations of the prototype for birds because robins possess more of the critical features for determining "birdness."

A number of researchers have recently argued that phonetic categories are organized around prototypes (Grieser & Kuhl, 1989; Hoffman, 1973; Kuhl, 1991; Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; J. L. Miller & Volaitis, 1989; Oden & Massaro, 1978; Samuel, 1982; Volaitis & Miller, 1992). Prototype theories are attractive to researchers working on theoretical issues in speech perception because prototypes provide one way of dealing with

Copyright 1997 by the American Psychological Association, Inc.

Correspondence concerning this article should be addressed to Scott E. Lively, who is now at Ameritech Product Management, 2C65E, 2000 W. Ameritech Center Drive, Hoffman Estates, Illinois 60196. Electronic mail may be sent via Internet to scott.e.lively@ameritech.com.

Experiments 1 and 2 were reported at the 125th meeting of the Acoustical Society of America, spring 1993.

the lack of acoustic–phonetic invariance in the acoustic signal. It is well known that speech signals are highly variable and are affected by many variables, such as vocal tract size, speaking rate, syllable position, emotional state, and ambient speaking and listening conditions (Lisker & Abramson, 1970; Lively, Pisoni, Summers, & Bernacki, 1993; Streeter, MacDonald, Apple, Krauss, & Galotti, 1983; Summerfield, 1981; Summers, Pisoni, Bernacki, Pedlow, & Stokes, 1988; Williams & Stevens, 1969). Despite this variability, listeners appear to extract invariant phonetic categories from the acoustic signal and display perceptual constancy over a wide range of conditions. Prototype theories attempt to solve the problem of perceptual invariance by assuming that listeners compare the incoming signal to an idealized form stored in long-term memory. If the signal is sufficiently similar to the prototype, then the input signal is accepted as a member of a phonetic category. Thus, perceptual invariance is not found in the acoustic signal, but rather it is achieved during the process of categorization by comparison with idealized forms or representations stored in long-term memory.

Several sources of evidence suggest that some members of a phonetic category are processed in qualitatively and functionally different ways than other members of the same category. Qualitative evidence comes from experiments in which participants are asked to rate the goodness of synthetic test materials drawn from one or more phonetic categories. For example, J. L. Miller and Volaitis (1989) and Volaitis and Miller (1992) found that participants rated some stop consonants as better examples of their phonetic categories than others. In a cross-linguistic study of vowel perception, Hoffman (1973) found that Japanese and English listeners rated some examples of /i/ as being more representative of their phonetic categories than others. Furthermore, both groups of listeners in her study selected the same example as the best or “focal” instance of the test set. Variations in apparent goodness observed by Miller and Volaitis and by Hoffman have been taken to indicate that not all members of a phonetic category approximate an internal standard or prototype to the same degree. In other words, category membership is graded.

The results of perceptual rating experiments suggest that some instances of phonetic categories are accepted by listeners as more representative category members than others. Several studies have also investigated the perceptual consequences of these differences in representativeness. Three findings support the claim that highly representative category members are more salient to listeners than less representative category members. First, Pisoni and Tash (1974) showed that synthetic stop consonants that were far from the category boundary were identified faster than stop consonants closer to the category boundary (see also Studdert-Kennedy, Liberman, & Stevens, 1963). Furthermore, discrimination response times were faster for pairs of highly representative examples than for pairs of less representative instances (see also Eimas & Miller, 1975; Pisoni & Tash, 1974; Repp, 1976). Second, J. L. Miller (1977) and Repp (1977) found that good instances of stop consonants were more effective dichotic competitors than poor instances of stop consonants (see also Cole & Cooper, 1977). Third, Samuel (1982) and J. L. Miller, Connine, Schermer, and Kluender (1983) reported that central category members were more effective adaptors in a selective adaptation paradigm than adaptors either closer to or farther from the phonetic category boundary. Taken together, these three sets of results suggest that highly representative members of phonetic categories are processed more efficiently than less representative examples.

More recent findings concerning the role of phonetic prototypes in speech perception have focused on the “perceptual magnet effect.” Kuhl (1991) reported that some instances of the vowel /i/ are judged better than other instances. Moreover, discrimination between a good example of /i/ and similar instances was more difficult than discrimination between a poor example of /i/ and instances located at equal psychophysical distances (see also Renda,

Hawks, & Klich, 1995). Iverson and Kuhl (1995) have extended Kuhl's (1991) findings by using signal detection analysis and a subset of Kuhl's original stimuli: They found that *d*'s were lower for test items that were close to tokens that received high goodness ratings than for test items that were close to tokens that received low goodness ratings. They also reported that test items that received high goodness ratings tended to cluster closer together in a multidimensional perceptual space more than items that received low goodness ratings. Iverson and Kuhl (1995) argued that these results indicate that good members of categories tend to draw similar items toward themselves to a greater degree than poor members of a category (see Nosofsky, 1986, for the original formulation of this proposal).

Kuhl (1991) demonstrated that the perceptual magnet effect is also species-specific: Both human adults and human infants, but not rhesus monkeys, show evidence of a magnet effect. Rather, monkeys' discrimination is determined solely by the psychophysical spacing of the vowels in the test set, suggesting the absence of any categorization. Phonetic prototypes appear to be formed very early in human development (Grieser & Kuhl, 1989). Kuhl et al. (1992) reported that infants as young as 6 months show a perceptual magnet effect for vowels of their native language, but not for vowels of a foreign language.

Although Kuhl's findings with infants are robust, careful examination of her results obtained with adults suggest that the magnet effect is weak (Kuhl, 1991): Although a large effect was found for vowels close to their respective referents, very small effects were obtained for vowels further from the referents (0.3% to 2.5% misses or failures to discriminate a comparison vowel from a reference vowel). The purpose of the present investigation was to assess the claim that category members are less discriminable from test items that closely approximate a prototype than from items that do not closely approximate a category prototype. Given that Kuhl's findings with adult listeners were not very strong, it is important to replicate these findings before drawing any strong conclusions about the relative discriminability of vowels within the same phonetic category or the role of phonetic prototypes in speech perception.

In Experiment 1, participants rated the goodness of a set of synthetic vowels developed from Kuhl's (1991) original test materials. The results of this experiment provide a measure of how representative each example of /i/ in the test set is of the category as a whole. In Experiment 2, the functional consequences of differences in representativeness were examined in more detail using test materials designed after Kuhl's (1991) synthetic tokens of /i/ and participants' self-selected best examples of the category /i/. The experimental task was the same—different discrimination procedure used by Kuhl (1991, 1992). Finally, in Experiment 3, categorization of the vowels synthesized from Kuhl's original parameters was assessed. It is critical to the logic underlying the perceptual magnet effect that all of the tokens in the test set are from the same phonetic category. However, few data have yet been published demonstrating that all of the vowels used in Kuhl's earlier studies were, in fact, perceived as /i/s (see recent articles by Iverson & Kuhl, 1995; Sussman & Lauckner-Morano, 1995). Because of the importance of prototype theory in the field of categorization and the recent claims made by Kuhl about how prototypes affect the categorization process, we were interested in conducting a more detailed investigation of the perceptual magnet effect.

Experiment 1: Vowel Rating

Two results are necessary to demonstrate a perceptual magnet effect in speech perception (Kuhl, 1991). First, some members of a phonetic category must be judged to be better examples of a phonetic category than others. Second, instances that approximate an idealized prototype should be less discriminable than category members that do not closely match the prototype. Kuhl (1991) established the first condition by having participants rate

variants of a synthetic vowel/i/for goodness in relation to a hypothetical internal standard. Variants in the prototype condition were generated around the average first and second formant values reported for male speakers by Peterson and Barney (1952). In the nonprototype condition, participants rated stimuli that surrounded an arbitrarily selected synthetic/i/. Variants and center vowels in each condition differed from each other acoustically in the same way in terms of the same psychophysical distances as indexed by a mel scale.

Three major findings were obtained in Kuhl's (1991) first experiment. First, Peterson and Barney's mean/i/was given the highest goodness rating in both experimental conditions. Second, ratings in the prototype condition decreased symmetrically with increasing distance from the highest rated vowel. Finally, a lack of symmetry was observed in the nonprototype condition: Tokens near Peterson and Barney's mean/i/were given high ratings, whereas tokens that were farther away were given lower ratings. Taken together, these findings suggest that Peterson and Barney's mean/i/approximates listeners' internal standards across a range of acoustic variation.

The present experimental procedure and test materials were designed to replicate Kuhl's (1991) earlier study. We expected that her major findings would be obtained, namely, that not all members of the phonetic category/i/would be treated as perceptual equals. Such a result would meet the first criterion for demonstrating a perceptual magnet effect in speech perception and would provide the basis for additional experiments exploring the details of this phenomenon.

Method

Participants—Participants in the present experiment were 72 undergraduates enrolled in an introductory psychology course at Indiana University. Listeners were given course credit for their participation. Participants reported no history of any speech or hearing disorders at the time of testing.

Materials—The test items, which were modeled closely after the synthesis parameters described by Kuhl (1991), are displayed schematically in Figure 1. Two sets of vowels were synthesized using the cascade configuration of Klatt's (1980) parallel-cascade speech synthesizer implemented on a VAX-3100 workstation. All items were steady-state synthetic vowels that were 510 ms in duration. In both test sets, the fundamental frequency had a rise-fall contour. At the onset of the stimulus, F0 was 112 Hz. Over the first 100 ms, F0 rose linearly to 132 Hz but fell to 92 Hz over the remaining 410 ms of the stimulus. The third, fourth, and fifth formants were held constant at 3010 Hz, 3300 Hz, and 3850 Hz, respectively, across all vowels in each set. Formant bandwidths and amplitude contours were set to the default values recommended by Klatt(1980).

The first (F1) and second (F2) formants were manipulated to create the prototype, the nonprototype, and all of their respective variants. Variants were 30, 60, 90, or 120 mels in Euclidean distance from their respective referents. Kuhl (1991, 1992) and Fant (1973) have provided the motivation for using vowels scaled in mel space. Variants lay along one of four diameters in which either only F1 changed, only F2 changed, both F1 and F2 changed in a positively correlated manner, or both F1 and F2 changed in a negatively correlated manner. The four Euclidean distances, combined with the four combinations of changes in F1 and F2, formed four "orbits" surrounding the center vowel in each condition. The first and second formants of the prototype vowel were set to the mean values of/i/reported by Peterson and Barney (1952) for male native speakers of English (F1 = 270 Hz and F2 = 2290 Hz, respectively). The first and second formants of the nonprototype were set to values

120 mels away from the prototype on the negatively correlated diameter ($F1 = 347$ Hz and $F2 = 2102$ Hz, respectively).

Procedure—Participants were placed in groups of 6 or fewer. Listeners were seated in individual sound-treated cubicles. Each cubicle was equipped with a CRT monitor, a set of TDH-39 matched and calibrated headphones, and a seven-button response box. The monitor and the response box were interfaced to a PDP 11/34 laboratory computer. The computer controlled stimulus presentation and response collection.

Listeners were randomly assigned to one of four groups. Half of the participants rated the prototype-centered test set for goodness, and the other half rated the nonprototype-centered test set. In addition, half of the participants had the word “peep” written on their response boxes and were told to base their ratings on what they imagined a good example of “peep” to sound like. The remaining participants did not see “peep” on their response boxes and were simply instructed by the experimenter to imagine the word “peep” and to make their judgments against the imagined syllable on each trial.

On each trial of the experiment, the CRT monitor in front of each participant displayed the prompt “Get Ready” for 500 ms prior to stimulus presentation. After the prompt disappeared, a single vowel was presented over the headphones. No auditory referent was provided. Listeners rated each item on a scale of 1 to 7. Participants were instructed to give poor instances of /i/ low ratings (i.e., 1 or 2) and good instances high ratings (i.e., 6 or 7). Listeners were also encouraged to use the entire range of ratings on the scale. Endpoints on the response boxes were labeled as “very poor” or “very good,” respectively. Participants rated each of the 33 variants five times each for a total of 165 trials. Listeners were given a short break half way through the experimental session. Each session lasted approximately 20 min.

Results

A mean rating for each orbit (distance from the center vowel) was obtained from each participant by averaging the ratings from all of the vowels in that orbit. These means were submitted to an analysis of variance (ANOVA). Vowel set (prototype vs. nonprototype) and instruction set (visual “peep” vs. no “peep”) were between-subjects variables. Vowel orbit or distance from the center vowel was treated as a within-subjects variable.

The main findings of the experiment are displayed in Figure 2. Only a significant main effect for orbit was obtained, $F(3, 204) = 39.18$, $p < .01$. Ratings were generally highest for vowels in the orbit closest to the center and decreased systematically with increasing distance from the center. This effect was consistent across both the prototype and nonprototype conditions. As Figure 2 shows, in the prototype condition, the vowel with the highest $F2$ value was given the highest goodness rating. In contrast, in the nonprototype condition, Peterson and Barney’s (1952) mean /i/ was given the highest rating. All of the vowels in the prototype condition that were greater in $F2$ than the center vowel, but had the same $F1$ value, were given high ratings. In the nonprototype condition, variation in $F2$ alone did not play as prominent a role in controlling ratings. Vowels in the prototype condition were given a restricted range of ratings. A much larger range of ratings was obtained for stimuli in the nonprototype condition.

Kuhl (1991) reported remarkable consistency across her listeners in terms of which vowel in the prototype condition was given the highest goodness rating. In the present experiment, many of the vowels in the prototype condition (21 of 33) were rated as the best /i/ by at least one participant. Although many vowels were rated as the best example of the category /i/, the majority of participants (65%) rated four vowels as the best example of /i/ in the prototype

condition. The remaining 17 vowels were each rated as the best example by three or fewer participants. These findings indicate that, although there is some agreement among participants, subjective ratings of vowel quality may be much more variable than the findings reported by Kuhl (1991).

Discussion

Two aspects of the present results are relevant to an evaluation of the proposed role of prototypes in speech perception. First, all of the vowels in the test series were not rated equally. Rather, some instances were rated as better examples of the category/i/ than others. This finding was particularly apparent in the nonprototype condition. High ratings were given to tokens on the left side of the diameter that varied only in F2, whereas much lower ratings were given to vowels on the right side.

Second, although the present findings demonstrate that the vowels differed in terms of their perceived goodness, the pattern of results did not replicate the pattern reported by Kuhl (1991), although we used similar stimuli and procedures. One major difference is that Peterson and Barney's (1952) mean vowel for adult males was not selected by participants in the prototype condition of the present experiment as the best example of/i/. Instead, participants selected an extreme vowel in F1–F2 space to be more representative of/i/ than a more central instance. Johnson (1989) reported a similar result in an experiment in which participants selected the best instances for vowel categories from a large set of items drawn from different phonetic categories.

Another difference between Kuhl's (1991) results and the present findings concerns the stability of prototypes across participants and conditions. Kuhl reported that goodness ratings were highly consistent across participants: All participants selected the same vowel as the most representative example of the category. In the present experiment, however, many of the vowels in the prototype condition were given the highest rating by at least one participant. Part of the discrepancy in the results may be due to differences between the participant groups tested in the two experiments. Kuhl stated that her participants "had some training in phonetics" (p. 94). In the present experiment, our listeners were naive to both synthetic speech and acoustic-phonetics. This difference between participant groups may also explain the findings that the overall ratings in the present experiment were generally lower than those reported by Kuhl (1991; see Kuhl et al., 1992, for a similar finding; Hillenbrand & Gayvert, 1993).

Another difference between Kuhl's results and the present findings concerns the consistency of ratings across conditions. Kuhl reported that the same vowel, Peterson and Barney's mean/i/, was given the highest goodness ratings in both the prototype and nonprototype conditions. This result suggests that the "focal vowel" was stable across acoustic contexts (Hoffman, 1973). We did not observe a similar finding in the present experiment. Rather, the most representative vowel shifted as a function of acoustic context. This finding suggests that participants' ratings of steady-state synthetic vowels may change across contexts. Similar sensitivities to context have been shown by J. L. Miller and Volaitis (1989) and Volaitis and Miller (1992), who reported that goodness ratings for stop consonants vary with changes in speaking rate.

Changes in goodness ratings across contexts cannot be accurately assessed by the present methods, however, because a given participant only heard variants from the prototype or nonprototype condition. Consistency was assessed more accurately in Experiment 2 by having participants pick their own best examples from the prototype-centered test set used in Experiment 1. After a best instance was selected from the prototype-centered set, variants were synthesized around that vowel and participants were asked to rate the new items for

goodness. If vowel ratings are consistent across contexts, then participants should select the same vowel as the best example in each acoustic context. However, if participants select different vowels in different contexts then this finding would imply that vowel ratings are not stable and may not accurately reflect the representatives of a given test item.

Experiment 2: Same–Different Discrimination

The results of Experiment 1 demonstrated that members of the category /i/ vary in the degree to which they match a hypothetical internal standard. This finding meets the first criterion for demonstrating a perceptual magnet effect. The second criterion for the magnet effect is that discriminability should vary as a function of rated goodness. Small acoustic differences among category members that approximate an internal standard should be harder to detect than comparable acoustic differences among items that do not approximate a prototype (Kuhl, 1991). This issue was addressed in Experiment 2.

Because participants in Experiment 1 gave many different vowels the highest goodness ratings, it is inappropriate to assume that the item designated as the most representative by Kuhl (1991) is the “best /i/” for participants in our experiments. Therefore, participants’ own best instances were determined in the present investigation and the perceptual magnet effect was examined with test sets that reflected each participant’s own internal category structure (Samuel, 1982). If the perceptual magnet effect depends on the relative goodness of a range of /i/s, then the likelihood of observing the effect should be maximized when participants hear vowels that approximate their own phonetic prototypes.

In the present experiment, participants selected their own best /i/s from the prototype-centered test set using the vowel rating procedure described Experiment 1. Failures to detect changes from a referent to a variant during discrimination were measured in four conditions. Participants heard prototype and nonprototype vowel sets as described previously by Kuhl (1991) and their own prototype and nonprototype vowel sets. We predicted that the perceptual magnet effect would be observed with both sets of test materials. However, the magnitude of the effect should be larger with listeners’ “self-selected” vowels because these stimuli are assumed to reflect participants’ internal category structures, whereas the vowels used by Kuhl were assumed to be a less representative example of this structure for our participant population.

Method

Participants—Participants were 9 Indiana University students who were paid \$10 for their participation in two experimental sessions. Listeners reported no history of speech or hearing disorders. None of the listeners had any prior experience with synthetic speech or any training in acoustic-phonetics. All participants were naive to the purpose of the experiment.

Materials—Two sets of test materials were synthesized for each participant. The first set was identical to the set used in Experiment 1. These were designated as Peterson and Barney (PB)-prototype and PB-nonprototype stimuli. The second set was based on participants’ self-selected prototypes. The procedure for determining listeners’ own prototypes is described below. Thirty-two variants were synthesized around each participant’s highest rated vowel. A vowel that was 120 mels away from the token with the best rating along the negatively correlated dimension was designated as the nonprototype. Thirty-two additional variants were then synthesized around this vowel. Distance relationships and correlational structure for each listener’s test sets were identical to those used in PB-prototype and PB-nonprototype conditions. The only differences between the PB test materials and the

participants' self-selected tokens were the F1 and F2 values. The first and second formant values for each participant's prototype and nonprototype are given in Table 1.

Procedure—Listeners participated individually in the present experiment over a 2-day period. Stimulus presentation and response collection were controlled by a PDP 11/34 computer. Participants listened over TDH-39 matched and calibrated headphones at a fixed and comfortable listening level of approximately 70 dB SPL. During the first experimental session, listeners rated the goodness of the PB-prototype set. The procedure was identical to that used in Experiment 1, except that participants rated each vowel three times rather than five times. The vowel with the highest mean goodness rating was assumed to approximate the participant's prototype. If two vowels received the same rating, the vowel with the faster mean response time was selected as the participants' prototype.

After the rating task was completed, participants took part in a same–different discrimination task that was similar to the one described by Kuhl (1991). During the first experimental session, participants heard the PB-prototype and PB-nonprototype test sets in separate blocks of trials. The prototype referent was Peterson and Barney's mean/i/. The nonprototype referent had a higher F1 and a lower F2 than the prototype referent and was 120 mels from the prototype on the negatively correlated F1–F2 diameter. In each condition, the variants were the 32 stimuli synthesized around the PB-prototype and PB-nonprototype referents.

Prior to the start of each condition, participants were familiarized with a new referent stimulus. The referent was presented 10 times, and participants made no responses. On each trial of the experiment, the referent vowel was presented four times. With probability .5, the referent changed to a variant or remained the same. This vowel was presented three times. Participants pressed a response key if they detected a change from the referent during this interval. If participants correctly detected a change, a light on the response box was illuminated. Feedback was not given for correct rejections, misses, or false alarms. These response contingencies regarding feedback parallel those used by Kuhl (1991) with adults and with infants in the “head turning” analog of the same–different discrimination task. The interval between presentations was 500 ms. The order of trials was randomized with the constraint that no more than three “change” or “no change” trials occurred in a row. Responses were only collected during the intervals in which a change in the referent was possible. Thus, responses made during the four presentations of the referent were not recorded. Furthermore, changes that were detected when a variant changed back to the referent were also not recorded. Because no responses were collected during the presentation of the referent, false alarm rates could not be estimated under these conditions.

Each of the 32 variants in the prototype and nonprototype conditions was presented twice during the experimental session. In addition, 64 “no change” trials were included in each block. Thus, participants heard 128 trials per condition of which half were “change” trials and half were “no change” trials. Participants were given a short break after every 32 trials (approximately 5 min) and a longer break after the completion of each condition. Half of the participants heard the PB-prototype as the referent first, and the other half heard the PB-nonprototype first.

The second experimental session was similar to the first, except that participants heard their own prototypes, nonprototypes, and variants. Order of presentation was counterbalanced across participants. After completing the same–different discrimination task, participants rated the goodness of the vowels from their own prototype condition to assess the consistency of their judgments across different acoustic contexts. The procedure for the final

rating task was identical to the one used to select the prototypes. Each experimental session lasted approximately 1 hr.

Results

The data of interest with regard to the perceptual magnet effect are participants' generalization scores or miss rates. These are test trials on which listeners failed to detect a change from the referent to a variant. For each participant, mean generalization scores were computed in each condition as a function of orbit and were submitted to an ANOVA. Test condition (PB materials vs. self-selected materials), referent type (prototype vs. nonprototype), and orbit or distance from the referent were within-subjects variables. All interactions were examined using Tukey's honestly significant difference statistic.

Perceptual magnet effect—The top panel of Figure 3 displays mean generalization scores as a function of orbit with the PB test materials. The bottom panel shows generalization scores with participants' self-selected materials. As the figure shows, very little evidence of a perceptual magnet effect was obtained under either of these conditions. Generalization decreased significantly as a function of distance from the referent, $F(3, 24) = 145.12$, $p < .01$. The significant three-way interaction of test condition, referent type, and orbit indicated that the PB-prototype produced less generalization than the PB-nonprototype or participants' self-selected prototypes in the first orbit, $F(3, 24) = 3.74$, $p < .05$. In the second orbit, the PB-prototype produced more generalization than participants' self-selected prototypes ($p < .05$). Differences among the remaining conditions were not significant. Generalization in the PB-prototype condition was not correlated with the difference between participants' goodness ratings of the PB-prototype and all other vowels in the PB-prototype set ($r = -.028$).

Symmetry of generalization—In addition to examining generalization as a function of distance, generalization was also examined as a function of diameter. Kuhl (1991) reported that prototypicality influenced discrimination symmetrically around the F1–F2 space for adult listeners. Thus, generalization was approximately equal across each diameter. Symmetry was assessed in the present experiment by calculating the percentage of generalization errors that occurred on each diameter in each condition. If generalization is symmetric across the F1–F2 space, then equivalent amounts of generalization should be observed across each diameter.

Figure 4 shows generalization as a function of diameter in each test condition. The solid line parallel to the x-axis indicates the amount of generalization predicted along each diameter by the symmetry assumption. As the figure shows, generalization was observed along the diameter that varied only in F2 than along any other diameter.

Consistency of goodness ratings—The results of the vowel rating task with participants' own prototype set were compared to the results of the rating task with the PB-prototype materials to examine the stability of category structures across different acoustic contexts. If vowel ratings are stable measures of the representativeness of vowels, then participants' ratings should be relatively unaffected by changing the acoustic range of the test materials.

Figure 5 displays the results of the analysis of stability across contexts. The solid lines on the figure represent the diameters of the PB-prototype set. Closed symbols represent participants' self-selected prototypes from the PB-prototype set, whereas open symbols represent participants' highest rated instances from their own prototype sets. Ovals that surround an open and closed symbol represent a single participant. For each listener, a shift

in the highest rated instance was observed across stimulus contexts. The mean size of the shift was 80 mels in Euclidean distance. The Appendix shows participants' mean goodness ratings for each vowel in their self-selected prototype test sets.

Discussion

In the present experiment, the perceptual magnet effect was examined using vowels based on stimulus parameters described by Kuhl (1991) or based on participants' own self-selected vowels. The experimental task was modeled closely after the same-different discrimination procedure Kuhl used with adult listeners. We failed to find a perceptual magnet effect with either set of test materials. When participants heard test materials based on Peterson and Barney's mean/i/, the prototype produced significantly less generalization than the nonprototype at the 30-mel distance. At 60 mels, more generalization was observed in the PB-prototype condition than the PB-nonprototype condition. However, the difference was not statistically significant. When participants heard their own self-selected test materials, generalization did not vary as a function of the prototypicality of the referent. These results suggest that the perceptual magnet effect may be difficult to replicate and that generalization over variants of/i/may not necessarily involve the use of prototypes in vowel perception.

In addition to failing to replicate the general pattern of a perceptual magnet effect, we also failed to replicate a more specific finding reported by Kuhl (1991). Kuhl reported symmetric generalization around the F1–F2 space for adult listeners. In the present experiment, generalization was highly asymmetric. Participants were unable to detect changes on a much higher percentage of trials when only F2 of the variant differed from the referent. These results suggest that changes in F2 were much harder to detect than changes in F1 alone or changes when F1 and F2 were correlated. This finding was somewhat surprising given that the vowels were originally scaled to lie at equal psychophysical intervals (see Kuhl, 1992). The pattern of results observed here indicates that the underlying psychological distance between variants may not be equated by independently scaling formant values in mels. Rather, more sophisticated scaling techniques may be required to ensure equal stimulus spacing with complex, multidimensional stimuli such as these synthetic vowels. For example, it may be important to manipulate vowel formants relationally, rather than in an absolute, independent manner (see J. D. Miller, 1989; Syrdal & Gopal, 1986). For the purpose of the present experiment, however, the assumption that independently manipulating different acoustic parameters by using a mel scale will produce similar perceptual effects may not be justified.

A final result from the present experiment concerns the consistency of the vowel ratings: Shifts in the highest rated instance were observed as a function of changes in acoustic context (see also J. L. Miller & Volaitis, 1989; Volaitis & Miller, 1992). These shifts tended to be confined to the diameter that varied only in F2. Five of the 9 participants selected a vowel with a higher F2 than their self-selected prototypes. Three listeners selected a vowel with a lower F2 than their self-selected prototypes. Taken together, these results suggest that goodness ratings are very sensitive to changes in acoustic context. The issue of consistency and its implications for the role of phonetic prototypes in speech perception are addressed later in the General Discussion.

Experiment 3: Vowel Categorization

In the first two experiments described, we assumed that all of the vowels in the PB-prototype and PB-nonprototype test sets were perceived as/i/s. However, few data have been presented demonstrating that participants do, in fact, unambiguously categorize all of the vowels in each set as/i/. Recently, Iverson and Kuhl (1995) found that PB-nonprototype and vowels with higher F1 values and lower F2 values than the PB-nonprototype are not

categorized unambiguously as/i/. The possibility that some of the vowels used in the previous experiments were not perceived as/i/s was examined further in the present experiment. Participants categorized tokens from the PB-prototype and PB-nonprototype test sets, as well as vowels from two new test sets, as the vowel in “beat,” “bait,” “bet,” or “bit.” The relevant data for the present investigation were the percentage of trials on which vowels from the PB-prototype and PB-nonprototype test sets were categorized as the vowel from “beat.” If all of the vowels are not perceived as/i/s, then it is possible that the weak evidence of a magnet effect obtained in Experiment 2 is simply an artifact of cross-category comparisons among different vowels: In the PB-prototype condition, participants would be more likely to make within-category discriminations than in the PB-nonprototype condition. Therefore, discriminations in the former case would be more difficult than in the latter case.

Method

Participants—Participants were 17 undergraduates enrolled in an introductory psychology class at Indiana University. Listeners were given course credit for their participation. No participants reported any history of speech or hearing disorders. Participants did not have any prior experience with synthetic speech or training in acoustic-phonetics. All participants were naive to the purpose of the experiment.

Materials—Two sets of test materials were synthesized for use in the present experiment. The first set of vowels was identical to the test materials used in the PB-prototype and PB-nonprototype conditions of Experiments 1 and 2. The second set of tokens was composed of 66 additional vowels. These vowels were divided into two subsets of 33 tokens each. Both subsets of new vowels were synthesized using the same physical relationships as the vowels from the PB-prototype and PB-nonprototype conditions. However, the formant values for the center vowels in the two new conditions were changed. For one subset of tokens, the F1 and F2 values of the center vowel were set to the values of/e/reported for a male speaker by Klatt (1980). The center vowel in the second subset was 120 mels in F1–F2 space from the center vowel in the first subset along the negatively correlated diameter. F1 of the center vowel in the second subset was higher than F1 of the center vowel in the first new subset. F2 was lower in the center vowel in the second subset than F2 of the center vowel in the first subset. F3, F4, and F5 for all vowels in the two new subsets were set to 2520 Hz, 3300 Hz, and 3850 Hz, respectively. The F0 contours and durations for all vowels used in the present experiment were the same as the stimuli used in Experiments 1 and 2. Figure 6 shows the relationship between the vowels in the PB-prototype and PB-nonprototype sets and the new vowels added in the present experiment.

Procedure—Listeners participated in groups of 5 or fewer. Participants sat in sound-attenuated cubicles that were equipped with TDH-39 matched and calibrated headphones and four-button response boxes. Buttons on the response boxes were labeled “beat,” “bait,” “bet,” and “bit.” Stimulus presentation and response collection was controlled by a PDP 11/34 computer. Participants heard the entire ensemble of 132 vowels during the experiment. On each trial, participants heard a randomly selected vowel from the test set and were asked to categorize the token by pressing one of the appropriately labeled buttons on the response box. Participants were given a maximum of 4 s to respond but were encouraged to respond as quickly and accurately as possible. The intertrial interval was 500 ms. Participants categorized each vowel three times for a total of 396 trials. The experiment lasted approximately 45 min.

Results

The critical results in the present experiment concern the categorization of the vowels from the PB-prototype and PB-nonprototype test sets. Figure 7 shows the percentage of trials on

which each vowel in the PB-prototype and PB-nonprototype sets were categorized as/i/. Categorization results for the vowels based on Peterson and Barney's mean/e/are not displayed because they are not directly relevant to the logic of the present investigation.

As shown in this figure, most of the vowels in the PB-prototype condition were categorized as/i/. Only one vowel in this condition was not categorized as an/i/on more than 50% of the trials. In contrast, many of the vowels in the PB-nonprototype condition were not reliably categorized as/i/. All of the vowels with an F1 greater than the PB-nonprototype were categorized as/i/on less than 50% of the trials. Indeed, many of these vowels were almost never labeled as/i/s. The pattern of results demonstrates that there are at least two distinct perceptual categories of vowels in the PB-prototype and PB-nonprototype sets and that not all responses involve the vowel category/i/.

Discussion

The results of the present experiment indicate that some of the vowels in the test set based on Kuhl's (1991) stimulus parameters were not perceived as members of the phonetic category/i/. Indeed, even the PB-nonprototype was categorized as/i/on only 54% of the trials, averaged across the prototype and nonprototype conditions. These findings suggest that participants were making cross-category comparisons on many of the trials of the same-different discrimination task in Experiment 2. The pattern of results raises several questions about the basis of the perceptual magnet effect. Recall that one of the critical assumptions of the methodology used by Kuhl to measure the magnet effect was that all of the test stimuli were perceived as members of the same phonetic category. The results of the present experiment are inconsistent with this important assumption. Some stimuli in the nonprototype condition are not perceived as/i/, as originally claimed by Kuhl (1991).

Iverson and Kuhl (1995) have recently reported a similar finding with regard to a reduced set of vowels in the PB-prototype and PB-nonprototype test sets. They examined goodness ratings and identification responses for vowels lying along the negatively correlated F1–F2 diagonal. Their vowel set included all vowels along this diagonal in the PB-prototype and PB-nonprototype sets. They found that all of the vowels with higher F1s and lower F2s than the PB-nonprototype were given relatively low goodness ratings (<3.5 on a 7-point scale) and were categorized as/i/on fewer than 50% of the experimental trials. They argued that these findings were surprising, given that all of Kuhl's (1991) participants reported hearing tokens from only one phonetic category. However, Iverson and Kuhl downplayed the importance of these results by emphasizing the context-sensitive nature of vowel identification (Fry, Abramson, Eimas, & Liberman, 1962; Nearey, 1989; Stevens & Öhman, 1969). Thus the present findings are consistent with Iverson and Kuhl's recent results and suggest that participants may have been making cross-category comparisons in the PB-nonprototype condition of Experiment 2.

Taken together, the present findings and the most recent results of Iverson and Kuhl (1995) suggest that the basis of the perceptual magnet effect may not be related to the rated goodness of stimulus tokens in the acoustic test set. Instead, the decreased ability of participants in the original Kuhl (1991) investigation to discriminate among vowels in the prototype condition, relative to the nonprototype condition, may be an artifact related to the range of vowel categories that listeners are exposed to during an experimental session. In the prototype condition, participants made only within-category discriminations. In the nonprototype condition, in contrast, participants made within-category comparisons on some trials and between-category comparisons on other trials. This difference in the distribution of trials requiring within-category and between-category discriminations suggests a plausible alternative explanation for the perceptual magnet effect that does not rely on the relative goodness or representativeness of the vowel tokens. The perceptual magnet effect may

reflect differences in category membership rather than differences in discriminability among items of the same phonetic category.

General Discussion

Three experiments were conducted to examine and assess the robustness of the perceptual magnet effect, which has been proposed as evidence of the use of phonetic prototypes in speech perception. Two general claims were evaluated in the present set of experiments. The first claim is that some members of a phonetic category approximate a hypothetical internal standard or prototype more closely than others. The second claim is that discriminability is affected by the goodness or representativeness of the category members. Specifically, vowels that approximate an idealized prototype should be more difficult to discriminate than vowels that do not approximate the category prototype. Taken together, these two claims define the perceptual magnet effect in speech perception (Kuhl, 1991).

The results of the present experiments provide support for the first assumption concerning phonetic prototypes. In Experiment 1, participants rated the goodness of two sets of synthetic vowels that were all supposed to be from the same phonetic category/i/. In both the prototype and nonprototype conditions, participants provided a range of goodness ratings. In the prototype condition, the vowel with the highest F2 value was given the highest rating. In the nonprototype condition, Peterson and Barney's (1952) mean/i/for adult males was given the highest rating. These results demonstrate that the items in each set varied in the degree to which they approximated an internal standard or prototype: Some category members were judged better than others.

However, goodness ratings were not stable across changes in acoustic context. In Experiment 1 and the rating portion of Experiment 2, the vowel that was given the highest goodness rating was not invariant, but shifted with changes in acoustic context. These findings are problematic for any version of prototype theory which claims that phonetic prototypes are perceptual representations that stand for the category as a whole. If this were the case, the same vowel should be the most representative vowel across changes in acoustic context. For example, in Experiment 2, only a different subset of the same category was sampled when the acoustic context changed. Thus, the representation of the vowel category should have remained the same. We did not obtain this predicted pattern of results in the present experiments.

Several other recent studies have also demonstrated that goodness ratings vary with changes in acoustic context. J. L. Miller and Volaitis (1989) and Volaitis and Miller (1992) have shown that speaking rate exerts a strong influence on the internal structure of phonetic categories that are primarily differentiated by changes in voice onset time. Taken together with the present findings, these results suggest that the goodness of a category member is a *relational property* rather than an absolute one (see also Joos, 1948; Ladefoged, 1967; Ladefoged & Broadbent, 1957). When the acoustic context changes, the relative perceptual salience of an item also changes. These findings suggest that although perceptual rating experiments with synthetic speech stimuli may provide some information about the relative internal structure of a phonetic category, the observed perceptual structure is very sensitive to changes in contextual variables. Thus, it is important to qualify the results of category rating experiments by noting cautiously that this methodology only measures the goodness of a given set of acoustic test materials. Results based on a small sample of tokens may not be representative of the category as a whole and may not generalize to other stimulus materials or test contexts.

The second claim addressed by the present experiments concerned the relative discriminability of synthetic vowels that varied in goodness. In Experiment 2, the perceptual

magnet effect was examined in a same–different discrimination task using vowels synthesized according to the stimulus parameters provided by Kuhl (1991) and vowels selected by the participants themselves. Little evidence was found to support a role of phonetic prototypes in speech perception. Indeed, in some conditions, poor examples of the category/i/ modeled after Kuhl's vowels produced more generalization in Experiment 2 than better instances of the same category. The results are exactly the reverse of what would be expected if there were a tendency to reduce discriminability among better quality speech tokens within a category, as a hypothetical perceptual magnet effect predicts.

The reason for our failure to replicate a perceptual magnet effect is unclear at this time. One possibility to consider is that dialect differences between participants in Kuhl's experiment and listeners in the present experiments may have affected the outcome of these experiments. If dialect differences were responsible for the differences among the participants, there would be no reason to suspect that a prototype derived from Peterson and Barney's average/i/ would be the most representative example for participants in the present experiments. In fact, the results of Experiment 1 indicated that participants varied in their choices of the best/i/. In contrast, Kuhl's participants consistently selected Peterson and Barney's (1952) mean vowel as the best exemplar of their/i/ category.

The possibility that dialect differences could be used to account for the present results can be largely discounted, however, by the results of Experiment 2. In that experiment, participants selected their own best instances from the vowels derived from the synthesis parameters used by Kuhl (1991). Thus, the test materials were custom tailored to suit the categories of individual participants. A perceptual magnet effect was also not observed under these conditions either. Generalization did not vary as a function of test condition (prototype vs. nonprototype) but was determined by the psychophysical distance between the referent and the variants. This pattern of results suggests that dialect differences cannot be used to account for the present failure to observe a perceptual magnet effect for/i/.

Another potential explanation for our failure to replicate Kuhl's (1991) findings is that the participants in the two experiments differed with respect to their knowledge of acoustic-phonetics and prior exposure to synthetic speech sounds. Kuhl (1991) reported that all of her adult participants had some training in phonetics. Participants in the present experiments were naive. It is possible that this difference may account for the discrepancies between the two studies. However, Kuhl's own results with infants render this possibility very unlikely. She reported that infants show a very strong perceptual magnet effect for vowel sounds in their own language (Kuhl, 1991; Kuhl et al., 1992). Furthermore, if the infant data are set aside for a moment, the generalizability and theoretical importance of the perceptual magnet effect in adults can be questioned if only participants who have some special training or prior phonetic knowledge demonstrate it under highly constrained experimental conditions.

The results of two recent studies also call into question the robustness of the perceptual magnet effect. First, Renda, Hawks, and Klich (1995) reported a perceptual magnet effect for/i/ only when they used the identical waveforms used in Kuhl's (1991) original study. However, they were unable to obtain the effect for the vowel/e/. Their failure to find an effect for/e/ was interesting because they used a smaller psychophysical distance between vowels in the/e/ set than in the/i/ set. They also indicated that many of the vowels in the prototype and nonprototype/e/ sets were not perceived as/e/ by at least one of their listeners. This condition favors finding a spurious magnet effect because participants may be making cross-category comparisons for the nonprototype condition. They concluded that the perceptual magnet effect may not occur for all vowels and may not be as robust a phenomenon as Kuhl has suggested.

Second, recent findings reported by Sussman and Lauckner-Morano (1995) also call into question the robustness of the perceptual magnet effect. They used a change/no-change discrimination paradigm with short interstimulus intervals and a modified subset of Kuhl's (1991) original stimuli that varied in F1 and F2. Their analysis of participants' misses did not show evidence of a perceptual magnet effect. When they computed d' 's, they obtained some weak evidence for a magnet effect: Participants' d' 's were lower when making comparisons against Kuhl's best test item than when making comparisons against a test item that was 75 mels away. This evidence is questionable, however, because the stimulus that was 75 mels from Kuhl's best test item was only identified as /i/ 52% of the time. When participants made comparisons against a stimulus that was identified as /i/ more reliably, only weak evidence of a perceptual magnet effect was obtained. Based on additional findings that showed that participants were better at making discriminations when F1 decreased between the standard and the comparison stimulus, Sussman and Lauckner-Morano (1995) concluded that auditory sensitivity may play an important role in the perceptual magnet effect. This conclusion and the observation that participants may be making cross-category comparisons call into question the robustness of the perceptual magnet effect and its generalization to other contexts.

In addition to examining the general pattern of the perceptual magnet effect, some of the fine-grained details of the present experiments should also be considered as possible reasons for failing to replicate Kuhl's (1991) results. For example, across each of the test sets used in the present experiments, F2 played a prominent role in determining the internal structure of the phonetic category. In Experiments 1 and 3, the vowel with the most extreme F2 value was given the highest goodness rating and was categorized as /i/ on the highest percentage of trials. In Experiment 2, the highest percentage of errors was observed along the diameter that varied only in F2. Determining the precise role of F2 in vowel categorization is beyond the scope of the present report. However, one possibility for the difficulty in discriminating changes in F2 may be that F2 and F3 combine perceptually to form a "center of gravity" because of their relatively close spacing in the vowel /i/ (Christovich & Lublinskaya, 1979).

Lacerda (1993) has recently reported a similar failure of discrimination among vowels that vary in F2. He examined vowel discrimination in younger and older infants and found that his listeners could discriminate tokens that varied in F1 but could not reliably discriminate items with the same psychophysical spacing that varied only in F2. Taken together, Lacerda's results and the present findings suggest that changes in the F2 of vowels are more difficult to discriminate than changes in F1, both for infants and adults. Obviously, further research is needed to localize this effect.

Historically, research on phonetic prototypes in speech perception has been concerned with determining the internal structure of phonetic categories and relating this internal structure to a variety of perceptual effects (Kuhl, 1991; Kuhl et al., 1992; J. L. Miller & Volaitis, 1989; Samuel, 1982; Volaitis & Miller, 1992). The challenge of this work is to define precisely what is meant by a phonetic prototype. Within the field of categorization research in cognitive psychology, the term "prototype" is generally taken to be a summary representation for the category as a whole (Smith & Medin, 1981). It is the cognitive reference point against which all potential category members are judged (Rosch, 1975). Prototypes have been defined in a number of ways, such as ideal instances (Rosch, 1975), composite representations of multiple exemplars (McClelland & Rumelhart, 1985), statistical averages and other measures of central tendency (Smith & Medin, 1981), and ideal values along stimulus dimensions (Oden & Massaro, 1978). Implicit in each of these definitions is the assumption that a prototype is represented in long-term memory as an abstraction and that it serves to represent the category as a whole.

The problem in applying the term *prototype* to phonetic categories in speech perception is that the acoustic-phonetic form of an utterance is highly dependent on the context in which it is uttered. Yet, according to the view that is at least implicit in a phonetic prototype theory, the same prototype must be applied in each case because it is the idealized representation for the category as a whole. This approach is assumed in spite of the fact that the internal structure of many phonetic categories changes as a function of contextual variables, such as speaking rate, vocal tract length, or phonetic context (J. L. Miller & Volaitis, 1989; Volaitis & Miller, 1992). The issue of whether a phonetic prototype represents a summary representation of a category or whether it simply means a good example of a category in a particular context requires more investigation. The qualitative notion of a “phonetic prototype” needs to be more rigorously defined before it can be considered a useful theoretical construct for researchers working in speech perception. At the very least, a theory of phonetic prototypes needs to specify the nature of the mental representation and how the prototype is adapted to function in different phonetic contexts (Jusczyk, 1993b).

For example, adapting prototype theory to deal with contextual shifts in goodness ratings may involve the postulation of some sort of perceptual normalization mechanism. However, if a normalization mechanism is incorporated, then some theoretical account must be offered as to why and how item-specific information, such as voice and speaking rate, is stored in memory and used to facilitate the recognition of new items (see Goldinger, 1992; Palmeri, Goldinger, & Pisoni, 1993; Lively, Logan, & Pisoni, 1993; Nygaard, Sommers, & Pisoni, 1992, 1994; Pisoni, 1992). A great deal of research from both the visual and auditory domains suggests that participants retain highly detailed information about the surface forms of objects they are exposed to in experimental contexts. Furthermore, much of this information appears to be applied to the recognition of new items without conscious awareness (Jacoby & Brooks, 1984; Kolers, 1976; Schacter & Church, 1992). If it is the case that highly detailed information from previously experienced instances is brought to bear on the recognition of new items, then the assumption of idealized, highly abstract, prototypical representations for phonetic categories may have to be abandoned in favor of multiple context-sensitive representations (Hintzman, 1986; Jusczyk, 1992, 1993a). This issue remains an important topic for further theoretical investigation in speech perception, as well as other areas of research on categorization.

In summary, the results of three perceptual experiments using synthetic vowels demonstrate that some members of the phonetic category /i/ are treated as more representative of the category as a whole than others. However, discriminability of specific instances was not affected by the goodness of the category members. The perceptual magnet effect was not observed with vowels based on synthesis parameters used by Kuhl (1991) or with participants' own self-selected prototypes. The failure to replicate Kuhl's results suggests that the perceptual magnet effect may not be a very robust phenomenon. Taken together, the results of the present series of experiments call into question the use of the perceptual magnet effect as evidence for the role of phonetic prototypes in speech perception. A more precise definition of a phonetic prototype will be necessary before it will be a useful theoretical construct in the field of speech perception and spoken language processing.

Acknowledgments

This work was supported, in part, by National Institutes of Health (NIH) Training Grant DC-00012-15 and NIH Research Grant DC-00111-17.

We wish to thank Peter W. Jusczyk, Joan Sussman, and Robert Remez for their helpful comments on this work and Matt Peuquett for his assistance in collecting the data.

References

- Christovich LA, Lublinskaya VV. The “center of gravity” effect in vowel spectra and the critical distance between the formants: Psychophysical study of perception of vowel-like stimuli. *Hearing Research*. 1979; 1:1985–195.
- Cole RA, Cooper WE. Properties of frication analyzers for [j]. *Journal of the Acoustical Society of America*. 1977; 62:177–182.
- Eimas, PD.; Miller, JL. Auditory memory and the processing of speech (Developmental studies of speech perception: Progress Report No 3, 117–135). Providence, RI: Walter S. Hunter Laboratory of Psychology, Brown University; 1975.
- Fant, G. Speech sounds and features. Cambridge, MA: MIT Press; 1973.
- Fry DB, Abramson AS, Eimas PD, Liberman AM. The identification and discrimination of synthetic vowels. *Language and Speech*. 1962; 5:171–189.
- Goldinger, SD. Words and voices: Implicit and explicit memory for spoken words (Research on speech perception Technical Report No 7). Bloomington, IN: Indiana University; 1992.
- Grieser D, Kuhl PK. Categorization of speech sounds by infants: Support for speech sound prototypes. *Developmental Psychology*. 1989; 25:577–588.
- Hillenbrand J, Gayvert RT. Identification of steady-state vowels synthesized from Peterson and Barney measurements. *Journal of the Acoustical Society of America*. 1993; 92:668–674. [PubMed: 8370872]
- Hintzman DL. “Schema abstraction” in a multiple trace memory model. *Psychological Review*. 1986; 93:411–428.
- Hoffman, MAB. Focal vowels. Paper presented at the 48th meeting of the Linguistic Society of America; San Diego, CA. 1973.
- Iverson P, Kuhl PK. Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *Journal of the Acoustical Society of America*. 1995; 97:553–562. [PubMed: 7860832]
- Jacoby, LL.; Brooks, LR. Nonanalytic cognition: Memory, perception, and concept learning. In: Bower, G., editor. *The psychology of learning and motivation*. Vol. 18. New York: Academic Press; 1984. p. 1-47.
- Johnson, KA. On the perceptual representation of vowel categories (Research on Speech Perception, Progress Report No 15, 343–358). Bloomington, IN: Speech Research Laboratory, Indiana University; 1989.
- Joos M. Acoustic phonetics. *Language*. 1948; 24:1–136.
- Juszyk, PW. Developing phonological categories from the speech signal. In: Ferguson, CA.; Menn, L.; Stoel-Gammon, C., editors. *Phonological development: Models, research, implications*. Timonium, MD; York: 1992. p. 17-64.
- Juszyk PW. From general to language-specific capacities: The WRAPSA model of how speech perception develops. *Journal of Phonetics*. 1993a; 21:3–28.
- Juszyk PW. Some reflections on developmental changes in speech perception and production. *Journal of Phonetics*. 1993b; 21:109–116.
- Klatt DH. Software for a parallel/cascade formant synthesizer. *Journal of the Acoustical Society of America*. 1980; 67:971–995.
- Kolers PA. Reading a year later. *Journal of Experimental Psychology: Human Learning and Memory*. 1976; 2:554–565.
- Kuhl PK. Human adults and human infants show a “perceptual magnet effect” for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*. 1991; 50:93–107. [PubMed: 1945741]
- Kuhl, PK. Psychoacoustics and speech perception: Internal standards, perceptual anchors, and prototypes. In: Werner, LA.; Rubel, EW., editors. *Developmental psychoacoustics*. Washington, DC: APA Press; 1992. p. 293-332.
- Kuhl PK, Williams KA, Lacerda F, Stevens KN, Lindblom B. Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*. 1992; 255:606–608. [PubMed: 1736364]

- Lacerda F. Sonority contrasts dominate young infants' vowel perception. *Journal of the Acoustical Society of America*. 1993; 93:2372.
- Ladefoged, P. Three areas of experimental phonetics. London: Oxford University Press; 1967.
- Ladefoged P, Broadbent D. Information conveyed by vowels. *Journal of the Acoustical Society of America*. 1957; 29:98–104.
- Lakoff, G. *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press; 1987.
- Lisker, L.; Abramson, AS. The voicing dimension: Some experiments in comparative phonetics. *Proceedings of the 6th international conference of phonetic sciences*; Prague: Academia; 1970. p. 563-567.
- Lively SE, Logan JS, Pisoni DB. Training Japanese listeners to identify English/r/and/l/: II. The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America*. 1993; 94:1242–1255. [PubMed: 8408964]
- Lively SE, Pisoni DB, Summers VW, Bernacki RH. Effects of cognitive workload on speech production: Acoustic analyses and perceptual consequences. *Journal of the Acoustical Society of America*. 1993; 93:2962–2973. [PubMed: 8315159]
- McClelland JL, Rumelhart DE. Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General*. 1985; 114:159–188. [PubMed: 3159828]
- Miller JD. Auditory-perceptual interpretation of the vowels. *Journal of the Acoustical Society of America*. 1989; 85:2114–2134. [PubMed: 2659639]
- Miller JL. Properties of feature detectors for VOT: The voiceless channel of analysis. *Journal of the Acoustical Society of America*. 1977; 62:641–648. [PubMed: 903506]
- Miller JL, Connine CM, Schermer TM, Kluender KR. A possible auditory basis for internal structure of phonetic categories. *Journal of the Acoustical Society of America*. 1983; 73:2124–2133. [PubMed: 6875098]
- Miller JL, Volaitis LE. Effects of speaking rate on the perceptual structure of a phonetic category. *Perception & Psychophysics*. 1989; 46:505–512. [PubMed: 2587179]
- Nearey TM. Static, dynamic, and relational properties in vowel perception. *Journal of the Acoustical Society of America*. 1989; 85:2088–2113. [PubMed: 2659638]
- Nosofsky RM. Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*. 1986; 115:39–57. [PubMed: 2937873]
- Nygaard, LC.; Sommers, MS.; Pisoni, DB. Effects of speaking rate and talker variability on the representation of spoken words in memory. *Proceedings of the 1992 International Conference on Spoken Language Processing*; Edmonton, Alberta, Canada: University of Alberta; 1992. p. 209-212.
- Nygaard LC, Sommers MS, Pisoni DB. Speech perception as a talker-contingent process. 1994 Manuscript submitted for publication.
- Oden GC, Massaro DW. Integration of featural information in speech perception. *Psychological Review*. 1978; 85:172–191. [PubMed: 663005]
- Palmeri TJ, Goldinger SD, Pisoni DB. Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1993; 19:309–328.
- Peterson GE, Barney HL. Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*. 1952; 24:175–184.
- Pisoni, DB. Talker normalization in speech perception. In: Tohkura, Y.; Vatikiotis-Bateson, E.; Sagisaka, Y., editors. *Speech perception, production and linguistic structure*. Tokyo: IOS Press; 1992. p. 143-151.
- Pisoni DB, Tash J. Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics*. 1974; 15:285–290.
- Renda SC, Hawks JW, Klich R. An investigation of the perceptual magnet effect in adults. *Journal of the Acoustical Society of America*. 1995; 97(Suppl 5):3420.

- Repp, BH. Posner's paradigm" and categorical perception: A negative study (Status report in speech research). Vol. SR-45/46. New Haven, CT: Haskins Laboratories; 1976. p. 153-161.
- Repp BH. Dichotic competition of speech sounds: The role of acoustic stimulus structure. *Journal of Experimental Psychology: Human Perception and Performance*. 1977; 3:37-50.
- Rosch E. Cognitive reference points. *Cognitive Psychology*. 1975; 7:532-547.
- Samuel AG. Phonetic prototypes. *Perception & Psychophysics*. 1982; 31:307-314. [PubMed: 7110883]
- Schacter DL, Church BA. Auditory priming: Implicit and explicit memory for words and voices. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1992; 18:915-930.
- Smith, EE.; Medin, DL. Concepts and categories. Cambridge, MA: Harvard University Press; 1981.
- Stevens KN, Öhman SEG. Crosslanguage study of vowel perception. *Language and Speech*. 1969; 12:1-23. [PubMed: 5789292]
- Streeter LA, MacDonald NH, Apple W, Krauss RM, Galotti KM. Acoustic and perceptual indicators of emotional stress. *Journal of the Acoustical Society of America*. 1983; 73:1354-1360. [PubMed: 6853847]
- Studdert-Kennedy M, Liberman AM, Stevens KN. Reaction time study to synthetic stop consonants and vowels at phoneme centers and at phoneme boundaries. *Journal of the Acoustical Society of America*. 1963; 35:1900.
- Summerfield AQ. Articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception & Performance*. 1981; 7:1074-1095. [PubMed: 6457109]
- Summers WV, Pisoni DB, Bernacki RH, Pedlow R, Stokes M. Effects of noise on speech production: Acoustic and perceptual analyses. *Journal of the Acoustical Society of America*. 1988; 84:901-916. [PubMed: 3183208]
- Sussman J, Lauckner-Morano VJ. 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*. 1995; 96:539-552. [PubMed: 7860831]
- Syrdal AK, Gopal HS. A perceptual model of vowel recognition based on the auditory representation of American English vowels. *Journal of the Acoustical Society of America*. 1986; 79:1086-1100. [PubMed: 3700864]
- Volaitis LE, Miller JL. Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories. *Journal of the Acoustical Society of America*. 1992; 92:723-735. [PubMed: 1506527]
- Williams CE, Stevens KN. On determining the emotional state of pilots during flight: An exploratory study. *Aerospace Medicine*. 1969; 40:1369-1372.

Figure 1. Formant frequencies for Kuhl's (1991) synthetic speech stimuli plotted in F1–F2 space. The values are scaled in terms of mels. F1 = first formant; F2 = second formant.

Figure 2. Goodness rating data from Experiment 1. The prototype condition is displayed in the upper left panel, and the nonprototype condition is displayed in the lower right panel. PB = Peterson and Barney (1952).

Figure 3.

Generalization scores from Experiment 2. The top panel shows performance with test stimuli synthesized from Kuhl's (1991) test parameters; the bottom panel shows generalization with participants' self-selected stimuli. PB = Peterson and Barney (1952).

Figure 4.

Percentage of generalization errors attributable to each diameter (D) in each condition of Experiment 2. PB = Peterson and Barney (1952); F1 = first formant; F2 = second formant.

Figure 5.

Shift in participants' highest rated exemplars across stimulus conditions in Experiment 2. PB = Peterson and Barney (1952); F1 = first formant; F2 = second formant.

Figure 6.

Formant frequencies for vowels in the PB-prototype set, PB = nonprototype set, and the two subsets of /e/ used in Experiment 3. PB = Peterson and Barney (1952); F1 = first formant; F2 = second formant.

\$watermark-text

Table 1

First (F1) and Second (F2) Formant Values for Each Participant's Prototype and Nonprototype

Participant	Prototype		Nonprototype	
	F1 in Hz	F2 in Hz	F1 in Hz	F2 in HZ
S1	270	2502	347	2302
S2	270	2430	347	2234
S3	308	2388	387	2194
S4	324	2290	404	2102
S5	308	2195	387	2012
S6	251	2242	327	2057
S7	233	2195	308	2013
S8	251	2239	327	2148
S9	233	2388	308	2194

Appendix

Participants' Goodness Ratings of Self-Selected Test Items

Orbit/Vector	S1	S2	S3	S4	S5	S6	S7	S8	S9
Center	2.00	6.00	5.67	4.00	4.67	4.67	5.67	5.67	6.00
o1v1	2.67	6.33	6.67	4.33	6.00	4.33	5.67	4.67	6.67
o1v2	4.33	5.00	5.33	4.33	6.00	5.00	6.00	4.67	3.67
o1v3	3.00	5.00	3.67	4.00	5.67	4.33	4.33	4.67	4.67
o1v4	2.67	6.00	4.67	5.33	6.00	4.33	5.00	4.67	3.33
o1v5	2.67	4.33	4.67	4.67	5.67	5.67	4.00	3.67	2.67
o1v6	2.00	5.33	3.67	5.33	5.33	5.00	3.33	1.67	1.67
o1v7	3.33	4.33	5.33	5.33	5.00	3.33	5.00	3.00	4.00
o1v8	2.00	4.00	5.00	5.00	5.33	5.33	4.67	3.00	3.00
o2v1	3.00	5.67	6.33	5.33	4.67	4.00	5.67	5.67	4.67
o2v2	4.00	5.00	6.33	4.00	6.00	3.33	4.67	5.00	4.33
o2v3	3.33	2.00	2.33	3.67	4.67	2.33	2.00	1.33	1.33
o2v4	3.67	6.00	4.33	5.00	6.33	5.33	3.33	2.67	1.67
o2v5	1.67	4.67	3.67	4.67	5.33	5.33	4.33	4.33	2.00
o2v6	2.00	3.00	3.00	4.33	4.67	2.50	2.00	3.33	1.33
o2v7	1.67	2.33	1.67	3.67	5.00	4.67	3.33	4.33	3.33
o2v8	3.00	4.67	3.00	3.00	5.00	4.33	4.67	3.33	4.00
o3v1	4.00	6.33	6.67	2.00	5.33	4.67	6.33	5.00	7.00
o3v2	4.00	1.00	3.33	2.00	5.00	2.33	3.00	5.33	1.33
o3v3	3.00	1.00	2.00	1.33	3.00	1.00	1.00	1.33	1.67
o3v4	2.00	3.00	3.00	2.33	5.00	3.00	2.00	2.00	1.00
o3v5	2.00	4.33	3.33	5.67	2.67	4.33	2.33	1.33	3.67
o3v6	2.00	2.67	1.67	4.33	3.00	3.33	1.33	2.00	1.33
o3v7	2.00	3.33	1.00	3.67	4.00	3.67	3.00	3.00	1.00
o3v8	2.00	2.67	3.00	2.67	5.00	3.00	4.00	4.33	2.00
o4v1	6.33	5.67	7.00	4.67	6.00	4.67	7.00	5.33	6.67
o4v2	3.67	1.00	3.67	1.33	2.67	1.00	1.00	1.67	2.33
o4v3	2.67	1.00	1.00	2.67	1.67	1.00	1.00	1.00	2.00
o4v4	3.33	2.67	2.67	2.67	2.33	1.00	2.00	1.67	1.00

\$watermark-text

\$watermark-text

\$watermark-text

Orbit/Vector	S1	S2	S3	S4	S5	S6	S7	S8	S9
o4v5	2.33	4.33	2.67	5.67	3.67	5.00	3.33	1.67	3.00
o4v6	1.33	3.33	1.00	5.67	4.00	2.67	1.00	2.00	1.33
o4v7	1.00	4.00	1.00	4.00	3.67	1.67	2.33	3.67	2.67
o4v8	1.00	3.33	2.67	2.33	3.67	2.33	3.00	4.67	2.67