

Depolarizing the perceptual magnet effect

Andrew J. Lotto^{a)}

*Department of Psychology and Parmlly Hearing Institute, Loyola University Chicago,
6525 N. Sheridan Road, Chicago, Illinois 60626*

Keith R. Kluender and Lori L. Holt

*Department of Psychology, University of Wisconsin—Madison, 1202 W. Johnson Street, Madison,
Wisconsin 53706*

(Received 1 October 1997; revised 6 February 1998; accepted 9 March 1998)

In recent years there has been a great deal of interest in demonstrations of the so-called “Perceptual-Magnet Effect” (PME). In these studies, AX-discrimination tasks purportedly reveal that discriminability of speech sounds from a single category varies with judged phonetic “goodness” of the sounds. However, one possible confound is that category membership is determined by identification of sounds in isolation, whereas, discrimination tasks include pairs of stimuli. In the first experiment of the current study, identifications and goodness judgments were obtained for vowels (/i/–/e/) presented in pairs. A substantial shift in phonetic identity was evidenced with changes in the context vowel. In a second experiment, listeners participated in an AX-discrimination task with the vowel pairs from the first experiment. Using the contextual identification functions from the first experiment, predictions of discriminability were calculated using the classic tenets of Categorical Perception. Obtained discriminability functions were well accounted for by predictions from identification. There was no additional unexplained variance that required the proposal of “perceptual magnets.” These results suggest that PME may be nothing more than further demonstration that general discriminability is greater for cross-category stimulus pairs than for within-category pairs. © 1998 Acoustical Society of America.

[S0001-4966(98)05606-9]

PACS numbers: 43.71.An, 43.71.Es [WS]

INTRODUCTION

One of the more fruitful areas of speech research in the last decade has been the study of the internal structure of phonetic categories.¹ Instead of the traditional concentration on the ability of listeners to distinguish phonetically relevant contrasts, this research has focused on functional relations of speech sounds within a single phonetic distribution. As may be expected, it has been demonstrated that not all exemplars of a distribution identified by a single phoneme are considered to be equally good representatives of that phoneme (Grieser and Kuhl, 1989; Miller and Volaitis, 1989; Kuhl, 1991; Volaitis and Miller, 1992). More surprising are results suggesting that perceived “category goodness” influences the discriminability of exemplars. Relatively good exemplars appear to be harder to discriminate from neighboring tokens; whereas, relatively poor exemplars are easier to discriminate from neighboring tokens. Because the perceptual space appears to “shrink” around good members of a category, this finding has been termed the Perceptual-Magnet Effect (PME; see Kuhl, 1991, 1993; Kuhl *et al.*, 1992). Kuhl (1991) suggests that the prototype (judged as best representative) vowel acts as a magnet drawing in vowel exemplars and increasing similarity between the magnet and other members of the category.

A typical demonstration of PME is presented by Iverson and Kuhl (1995). In their first experiment, a series of vowels

varying in first ($F1$) and second ($F2$) formant frequency were presented to adult listeners for phonetic identification and goodness judgments. This vowel series (Fig. 1) contained equal mel-spaced steps and was equivalent to the diagonal from the /i/ distributions utilized by Grieser and Kuhl (1989) and Kuhl (1991). Subjects identified each vowel, presented in isolation, as either /i/ or /e/. Vowels 1 through 9 (as labeled in Fig. 1) were each identified as /i/ over 50% of the time and were considered by Iverson and Kuhl to all be appropriate members of the phonetic category /i/. Subjects also judged the goodness of each vowel token as a representative of /i/. There was a clear gradient in the goodness judgments with those tokens with higher $F2$ and lower $F1$ frequencies being given higher average goodness ratings. Iverson and Kuhl designated vowel 5 as the prototype (**P**) vowel and vowel 9 as the nonprototype (**NP**). Of course, “prototype” is a label with some theoretical content as, for example, in the categorization work of Rosch (1975) and Posner and Keele (1968, 1970). Moreover, Kuhl (1992, 1993) has suggested that internal prototypes may play a key role in human speech-sound categorization. However, in this case, the “prototype” vowel refers simply to a relatively good exemplar and “nonprototype” refers to a poorer exemplar. In fact, **P** did not receive the highest goodness rating in this series, but was chosen because it had $F1$ and $F2$ values identical to the “prototype” used in earlier experiments by Kuhl (1991; Grieser and Kuhl, 1989). These $F1$ and $F2$ values were also, not coincidentally, the values reported by

^{a)}Electronic mail: alotto@luc.edu

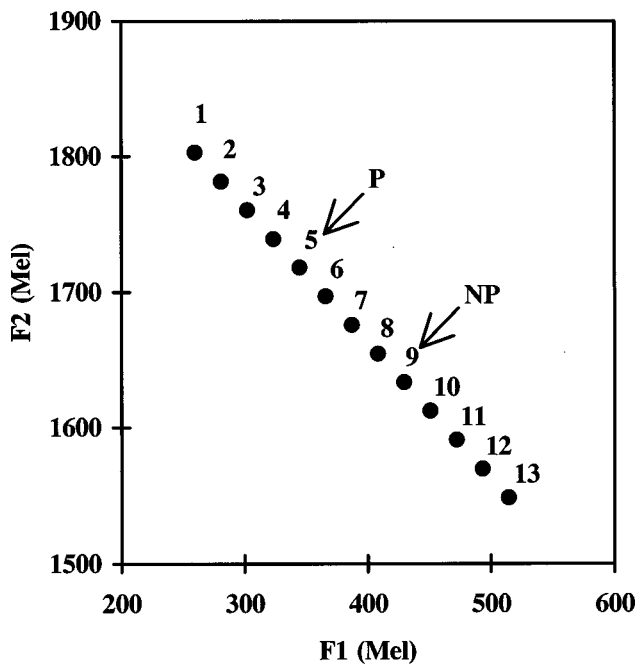


FIG. 1. Formant-frequency values in mel for stimuli used in Iverson and Kuhl (1995) and the present study. **P** is the "prototype" vowel and **NP** is the "nonprototype" as specified in Kuhl (1991). Vowels 1 and 13 are not used in the present design.

Peterson and Barney (1952) for the average production of /i/ by 33 adult male speakers.

In the second Iverson and Kuhl (1995) experiment, subjects participated in an AX (same/different) discrimination task. In one block of trials, subjects attempted to discriminate **P** from its six nearest neighbors (vowels 2, 3, 4, 6, 7, and 8) and, in another block, discriminated **NP** from its neighbors (vowels 6, 7, 8, 10, 11, and 12). Because the distance between vowels was presumably psychophysically equal (30 mels), one might predict that **P** and **NP** would be equally discriminable from their neighbors. However, the results of a signal-detection analysis (Macmillan and Creelman, 1991) indicated that subjects were better at distinguishing **NP** from surrounding vowels. Even when one disregards discriminations across the category boundary (i.e., **NP** vs vowels 10, 11, and 12), there still appears to be a relation between discriminability and category goodness. Discrimination scores (d') were lower for pairings of **P** with vowels that received higher goodness ratings (vowels 2, 3, and 4) than for pairings of **P** with poorer exemplars (vowels 6, 7, and 8). These intra-category discrimination differences define PME.

The fecundity of this effect in terms of research and theory is notable. Demonstrations of perceptual magnet effects have been described for several phonetic categories (e.g., Davis and Kuhl, 1994; Iverson *et al.*, 1994; Iverson and Kuhl, 1996); the effect has been elicited from infant listeners (Grieser and Kuhl, 1989; Kuhl *et al.*, 1992; Polka and Werker, 1994); it has been modeled by various computational "neural network" approaches (Guenther and Gjaja, 1996; Lacerda, 1998); and has served as the foundation of a developmental speech-perception theory (Kuhl, 1992, 1993).

Despite the comprehensive nature of this work, there have been some concerns raised about the demonstrations of

PME. One of these concerns is that not all stimuli in these experiments are actually members of the same phonetic category. Lotto *et al.* (1996) reported data from an experiment in which phonetically naive adults were presented with the series from Iverson and Kuhl (1995) and were asked to identify the vowel as /i/, /æ/, /e/, /ɪ/, /ɛ/, /ʊ/, or "none of the above." The vowel stimulus **NP** was identified as /i/ only 7.3% of the time. Similar results were obtained by Sussman and Lauckner-Morano (1995), who presented the entire /i/ distribution from Kuhl (1991) to phonetically-trained listeners for identification. The same **NP** vowel was identified as /i/ only 8% of the time. Thus it appears that infants and adults in the studies of Grieser and Kuhl (1989) and Kuhl (1991) were actually displaying better discrimination for *inter*-category pairings (more likely with **NP**) than for *intra*-category pairings (those containing **P**). This result would not be novel; increased discriminability of cross-boundary contrasts is one of the classical findings of speech perception as a defining feature of "Categorical Perception" (Liberman *et al.*, 1957; Stevens *et al.*, 1969; Pisoni, 1971; Wood, 1976).²

This concern about stimulus set is exacerbated by an aspect of the methodology used in the design of Iverson and Kuhl (1995) and which is typical of PME experiments. As described above, vowels were presented in isolation for identification and goodness rating, but were presented as pairs in the discrimination task. This is problematic because there is ample evidence that identification of vowels can be affected by the presence of contextual sounds (e.g., Fry *et al.*, 1962; Eimas, 1963; Nearey, 1989). In particular, Thompson and Hollien (1970) demonstrated that the identification of an ambiguous vowel can be shifted to /i/ when preceded by a good exemplar of /ε/ and can be shifted to /ε/ when preceded by a good /i/. If this applies to the Iverson and Kuhl (1995) experiment, then vowels identified as /i/ in isolation may be perceived as /ε/ when presented in context during discrimination. This shift is more likely to happen for boundary stimuli. Therefore, **NP** may be perceived as /ε/ when presented with vowels 6, 7, or 8 and as /i/ when presented with vowels 10, 11, or 12. Thus all of these discriminations would be between phonetic categories. This possibility makes the interpretation of the data from Grieser and Kuhl (1989), Kuhl (1991), Kuhl *et al.* (1992), and Iverson and Kuhl (1995) more problematic.

To analyze this possibility, experiments 1 and 2 in this paper are replications of the first two experiments of Iverson and Kuhl (1995) except that identification and goodness ratings (experiment 1) are elicited in the same contexts as are present in the AX-discrimination experiment. In fact, in the present identification and discrimination tasks, experimental design and stimulus presentation are exactly the same; only the response labels differ. Thus the identification responses will presumably be a good representation of the perceived phonetic identities of the vowels during the discrimination task. In this way, one is able to determine if these patterns of discriminability really hold for stimuli from a single category.

I. EXPERIMENT 1

The design of this experiment is nearly identical to the AX-discrimination task constructed by Iverson and Kuhl (1995). However, instead of responding to differences between stimuli, subjects identify or rate the category goodness of one member of the stimulus pair. If context affects the perceived identity and category goodness of a vowel, then one would expect differences between these data and those collected by Iverson and Kuhl (1995) for stimuli presented in isolation.

A. Methods

1. Subjects

Ten listeners participated in the identification task and twenty-one different listeners participated in the category-goodness rating task. All subjects were students at the University of Wisconsin-Madison and reported that they were native English speakers with normal hearing.

2. Stimuli

The stimuli were synthesized at a 10-kHz sampling rate with the cascade formant synthesizer described by Klatt (1980) using the parameter values from Iverson and Kuhl (1995). Each vowel was 435-ms long and had steady-state frequency values of 3010, 3300, and 3850 Hz for the third through fifth formants. Frequency values for $F1$ and $F2$ varied to produce a Euclidean distance of 30 mel between each vowel as shown in Fig. 1. $F1$ -frequency varied from 197 to 429 Hz and $F2$ -frequency varied from 1925 to 2489 Hz. The values for **P** were 270 and 2290 Hz for $F1$ and $F2$, respectively. For **NP** these values were 347 and 2102 Hz. For all vowels, f_0 rose from 112 to 130 Hz over the first 100 ms and then declined to 92 Hz over the final 335 ms of the vowel.

All stimuli were matched in rms energy. Following D/A conversion (Ariel DSP-16), stimuli were low-pass filtered (Frequency Devices 677, cutoff frequency: 4.8 kHz) prior to being amplified (Stewart HDA4), and played over headphones (Beyer DT-100) at 75 dB SPL.

3. Procedure

In each experimental session, one to three subjects were tested concurrently in single-subject sound-attenuated booths (Suttle Equipment). There were four blocks of trials in a session. In two of the blocks, **P** (vowel 5) was paired with itself or with one of the neighboring stimuli (vowels 2, 3, 4, 6, 7, 8). In the other two blocks, **NP** (vowel 9) was paired with itself or with its neighbors (vowels 6, 7, 8, 10, 11, 12). In each block there were 30 *same* trials containing the standard (**P-P** or **NP-NP**), five *same* trials for each of the six neighbor stimuli (2-2, 3-3, 4-4, etc.) and 10 *different* trials for each of the comparison stimuli (e.g., P-2, P-3, 6-P, or NP-6, 10-NP, NP-8). In half of the *different* trials, **P** or **NP** were presented first in the pair. For one block each of the **P** and **NP** conditions, subjects either identified or gave a goodness rating for the first vowel in the pair; during the other block they identified or rated the second vowel. The order of

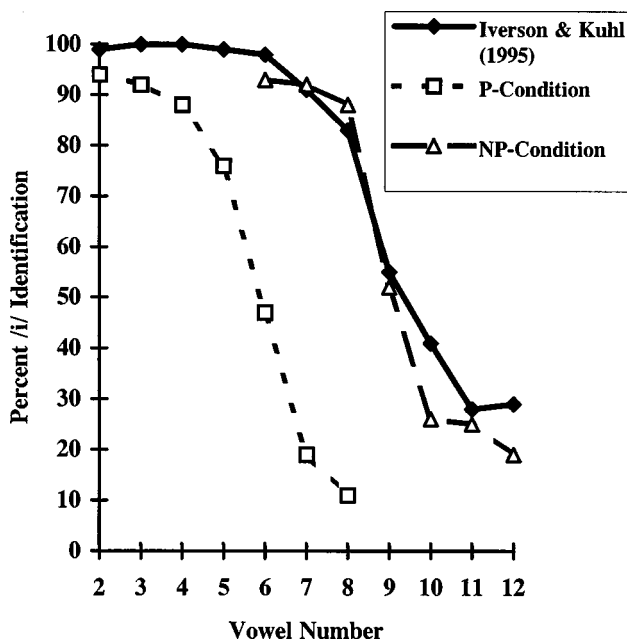


FIG. 2. Identification functions for stimuli presented in context of **P** and **NP**. For comparison, identification function of stimuli presented in isolation from Iverson and Kuhl (1995, numbers taken from their Fig. 2) are also presented.

presentation of the **P** and **NP** blocks and the order of instructions to label the first or second stimulus were counterbalanced across subjects.

On each trial, a randomly determined pair of stimuli were played with an ISI of 250 ms. Subjects in the identification task were instructed to label either the first or second stimulus as the vowel in “he” or as the vowel in “hay” by pressing an appropriately labeled button on a response box. Subjects in the goodness-rating task were asked to evaluate the “goodness” of the first or second vowel as an example of the vowel in “he.” These ratings were made by pressing one of seven buttons on a response box labeled from “very good” (7) to “very poor” (1). Stimulus pairs were presented about once every four seconds.

Prior to the four blocks of 120 stimuli, two short training blocks were presented. Subjects heard each of the 19 pair types (7 *same* pairs and 12 *different* pairs) but did not make any responses. This allowed subjects to become familiar with the range of the stimuli in context. Training and testing together lasted approximately 45 min.

B. Results and discussion

1. Identification task

Identification functions in the context of **P** and **NP** are presented in Fig. 2 along with the identification data from Iverson and Kuhl (1995) for stimuli presented in isolation. Identification percentage is collapsed across presentation order. For example, identifications of vowel 10 in the **NP** condition are averaged across five identifications of vowel 10 followed by **NP** and five identifications of vowel 10 preceded by **NP**. Data for **P** (vowel 5) and **NP** (vowel 9) are for identifications of repeated stimuli (e.g., identify **P** preceded by **P**).

TABLE I. Mean percent /i/ identifications and goodness ratings for vowels 6, 7, and 8 in **P** and **NP** contexts. These are the vowels which appear in both conditions. Matched-sample *t* tests comparing the differences between contexts are also displayed. *=significant for two-tailed test at alpha of 0.05.

Vowel	P condition identification	NP condition identification	<i>t</i> test df(9)	P goodness	NP goodness	<i>t</i> test df(20)
6	47%	93%	4.01*	4.13	5.41	3.66*
7	19%	92%	6.03*	2.92	5.26	6.28*
8	11%	88%	7.45*	2.40	4.66	6.57*

It is clear from Fig. 2 that identification of a vowel is affected considerably by context. Identification functions for the **P** condition differed substantially from the data reported by Iverson and Kuhl (1995) for stimuli presented in isolation. The shift in identifications with changes in context are similar to those presented in earlier demonstrations of vowel contrast (e.g., Fry *et al.*, 1962; Thompson and Hollien, 1970). Vowels presented with a relatively good exemplar of /i/ (**P** condition) were identified as /i/ less often than when presented in isolation. Of particular interest is the identification of vowel 6, 7, and 8, because these were presented in context of both **P** and **NP**. Table I presents the mean percentage of /i/ identification for each vowel in each context along with results of tests of the significance of the effects of context. All three vowels changed identity from /i/ (average /i/ identification=91%) in the **NP** condition to /e/ (average /i/ identification=25.67%) in the **P** condition.

Also of importance to demonstrations of PME is the perceived identity of **P** and **NP** themselves in the context of other vowels. Table II displays the mean percentage of /i/ identifications for **P** and **NP** in all contexts collapsed across presentation order. Context effected a change in perceived identity for both vowels. In the context of better /i/ representatives (vowels 6, 7, and 8), **NP** was perceived most often to be /e/ (average /i/ identification=27.33%). In the context of poor /i/ representatives (vowels 10, 11, and 12), **NP** was labeled more often as /i/ (average /i/ identification =66.67%). The percentage of /i/ identifications for **P** was also affected by context, but was above 50% for all contexts except for vowel 2 (48%).

TABLE II. Mean percent /i/ identification for **P** (vowel 5) and **NP** (vowel 9) in the context of their neighboring vowels. **P** was identified in the context of vowels 2 through 8 and **NP** was identified in the context of vowels 6 through 12.

Context vowel	P identification	NP identification
2	48	...
3	52	...
4	69	...
5	76	...
6	79	23
7	82	23
8	84	36
9	...	52
10	...	56
11	...	62
12	...	82

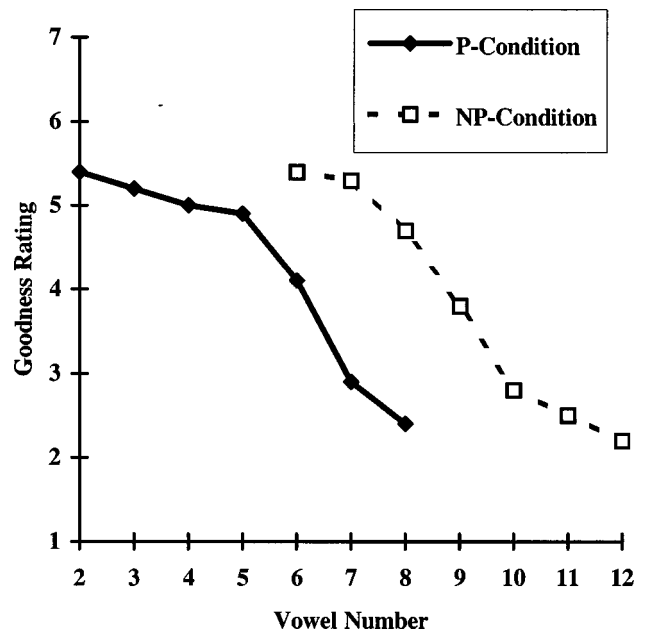


FIG. 3. Category goodness ratings for stimuli presented in context of **P** and **NP**. Scale goes from 7 (very good /i/) to 1 (very poor /i/).

2. Goodness judgments

One subject's data was omitted from the analysis because the subject only used one rating for all of the vowels in the experiment. Mean goodness judgments collapsed across the remaining 20 subjects' data are presented in Fig. 3 for both **P** and **NP** contexts. Perceived category goodness also appears to be highly dependent on vowel context. Again, vowels 6, 7, and 8 were rated in both contexts and significantly different ratings were obtained for the **P** and **NP** contexts (Table I). Vowels presented in the context of a relatively good /i/ exemplar (**P**) were judged to be poorer members of the category /i/ (mean rating of 2.7) than when those same vowels were presented with **NP**, a relatively poor exemplar (mean rating of 5.3).

This context sensitivity of perceived vowel identity and category goodness makes PME demonstrations (e.g., Grieser and Kuhl, 1989; Kuhl, 1991; Kuhl *et al.*, 1992; Iverson and Kuhl, 1995) difficult to interpret. For example, results from the second experiment of Iverson and Kuhl (1995) indicate that discriminability of **NP** from neighboring vowels is higher than discriminability of **P** from its neighbors. Complicating these results, the data from the current experiment suggest that all of the comparisons involving **NP** are *inter*-categorical and most of the comparisons with **P** are *intra*-categorical. When **NP** is paired with vowels 6, 7, or 8, **NP** is identified as /e/ (73% of the time) and the other vowel is perceived as /i/ (91% of the time). On the other hand, when **NP** is paired with the vowels 10, 11, or 12, **NP** is identified as /i/ (67% of the time) and the paired vowel is identified as /e/ (77% of the time; all of these numbers represent averages). That is, listeners are always asked to discriminate an /i/ token from an /e/ token during **NP** "different" trials. During blocks of **P** stimuli, discriminations are always made between two /i/ tokens, except when **P** is paired with vowel 2. If the results of the identification task of experiment 1 are reasonable estimates of perceived identity during the dis-

crimination task, then the obtained pattern of discrimination results can be described as follows: Discriminability of two vowels judged to be from different phonetic categories is higher than discriminability of two vowels judged to be from the same phonetic category. This pattern of results is exactly that described classically as Categorical Perception.

If, in fact, PME does reduce to a demonstration of phonetic Categorical Perception, then one can predict quantitatively the differences in discrimination scores between **NP** and **P** conditions. Past research on categorical perception of speech sounds has shown that discrimination can be fairly well modeled by assuming that discriminability of two speech sounds is solely a function of their perceived phonetic identities. That is, two sounds labeled as different phones can be discriminated; whereas, those labeled equivalently cannot be discriminated (e.g., Liberman *et al.*, 1957; Miyawaki *et al.*, 1975). Experiment 2 was designed to determine if this simple method (along with context effects on identification demonstrated in experiment 1) could account for PME.

II. EXPERIMENT 2

Experiment 2 is a nearly exact replication of the second experiment from Iverson and Kuhl (1995). Listeners were asked to discriminate the same vowels that were presented in experiment 1 in order to investigate the ability of identification percentages to account for the differences of discriminability in **P** and **NP** conditions. Discrimination data from Iverson and Kuhl (1995) were not used because there has been some suggestion that dialect differences and heterogeneous laboratory procedures may have been responsible for some of the past discrepancies in vowel identification (e.g., Sussman and Lauckner-Morano, 1995) and choice of “best” or “prototypical” vowel (e.g., Lively, 1993) with these very same stimuli. Thus listeners from the same subject pool were chosen for the two experiments in this paper. Besides this difference in subjects, the current experiment was designed to be as similar as possible to the discrimination experiment of Iverson and Kuhl (1995).

A. Methods

1. Subjects

Twenty-two University of Wisconsin undergraduate students participated in experiment 2. All reported normal hearing and had learned English as their first language. None of these subjects had participated in experiment 1. While using the same subjects for the identification and discrimination tasks may have made comparisons between the tasks more valid, different subjects were used for the discrimination task to make the results as comparable to previous PME studies as possible. Iverson and Kuhl (1995) used different subjects in their identification and discrimination tasks, and it was decided that giving subjects in the current experiment extra experience with the stimuli (in an identification task) would introduce an undesired confound.

2. Stimuli

Stimuli were identical to those used in experiment 1.

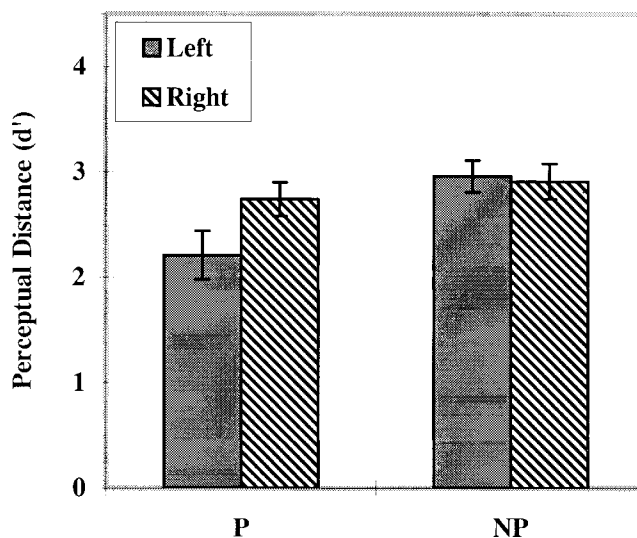


FIG. 4. Average d' scores for experiment 2 separated by standard (**P** or **NP**) and by position of compared stimulus with respect to standard (Left or Right).

3. Procedure

Stimulus presentation was similar to that used in experiment 1, except that there were now only two blocks of stimuli; a 240-trial block of **P** trials and a 240-trial block of **NP** trials. Each block consisted of 120 *same* trials (60 of **P** or **NP** and 10 of each neighboring stimulus) and 120 *different* trials, and order of block presentation was counterbalanced.

After presentation of a vowel pair, subjects pressed one of two labeled buttons to register whether the pair was “same” or “different.” This is slightly different from the response requested by Iverson and Kuhl (1995). Their subjects held down a key during vowel presentation and lifted the response key when they thought that they heard a different pair. This response technique was more similar to the go/no-go task used by Kuhl (1991) for monkeys and human infants.

Prior to each block of test trials, subjects participated in a short training block. These blocks consisted of 12 *same* trials and 12 *different* trials randomly ordered. Subjects received feedback of the correct button response during training. A block of training trials containing **P** stimuli were presented prior to the **P** testing block and a similar block of **NP** stimuli were presented prior to testing with **NP** stimuli. This is also slightly different from Iverson and Kuhl (1995). Their subjects received only one block of training (either **P** or **NP**) prior to testing.

B. Results and discussion

A bias-free measure of discriminability, d' (Macmillan and Creelman, 1991), was calculated for each vowel pairing for each subject. Average d' scores are presented in Fig. 4 and Table III. Statistical analyses of the obtained data replicated all qualitative results from Iverson and Kuhl (1995). Of greatest significance, the PME was apparent in the data. Paired-sample t -test revealed that listeners were significantly worse at discriminating stimuli from **P** than from **NP** [$t(21) = -3.40$, $p < 0.005$]. In addition, discrimination

TABLE III. Average perceptual distance (d') for each comparison in experiment 2. Each comparison stimulus is identified by a distance (in mel) and direction (left or right) of **P** or **NP**.

Stimulus location	30 mel	60 mel	90 mel
Left of P	1.19	2.24	3.20
Right of P	1.60	2.92	3.71
Left of NP	1.86	3.15	3.86
Right of NP	1.76	3.26	3.69

was significantly worse [$t(21)=3.3, p<0.005$] to the left of **P** than to the right of **P**. Iverson and Kuhl (1995) propose that this difference is indicative of PME without the confound of inter-categorical pairs, because vowels 2, 3, 4, 6, 7, and 8 were all labeled as /i/ over 80% of the time in their identification task. Results of experiment 1 raise doubts about the perceived identities of these vowels and it is possible that the difference in discriminations to the right and left of **P** also result from confounds with category membership. This possibility will be analyzed later in this report. For the boundary stimulus **NP**, there was no significant difference [$t(21)=0.31, p=0.76$] between discriminations from the right and left. Also, there was no significant difference [$t(21)=-1.31, p=0.21$] for discriminations of the shared vector (i.e., vowels 6, 7, and 8) in the **P** and **NP** conditions.

Another way of analyzing these discrimination data in light of the purported theoretical basis of PME is offered by Kuhl (1991). The *generalization score* is the percentage of misses on trials containing a nonidentical pair. That is, it is the percentage of “same” responses on trials with two different stimuli. If, in fact, prototypical members of a category act as perceptual magnets, then listeners should produce generalization errors more often when presented **P** and its neighbors than when presented **NP** and its neighbors. This is exactly the pattern of results obtained in the present discrimination task. The generalization score was significantly greater [$t(21)=3.20, p<0.005$] in the **P** condition (mean score=23.60) than in the **NP** condition (mean score=17.99). Additionally, generalization scores were significantly higher [$t(21)=6.24, p<0.0001$] for discriminations to the left side of **P** (mean score=30.53) than for discriminations to the right of **P** (mean score=16.67). This is also indicative of PME.

One concern with these results, however, is the possible confound of intercategory discriminations. One way to examine this confound is to compare the observed generalization scores with scores that are predicted solely on the basis of a simple relation between perceived identity and discrimination. As pointed out by Liberman *et al.* (1957), if it is assumed that listeners can only discriminate speech sounds that differ in their phonemic label, then the percentage of misses should be comparable to the probability of a listener labeling the two (different) vowel sounds equivalently.³ That is, listeners are presumed to be unable to discriminate speech sounds that they identify with the same phonemic label. For the present discrimination task, these computations will reveal the degree of difference in generalization score between **P** and **NP** conditions that arise solely from the differences in perceived identity that were demonstrated in experiment 1.

These predicted generalization scores for the two conditions will not include any variation due to “phonetic goodness” and will simply be based on phonemic identity. If the theory of perceptual magnets is correct, then the obtained generalization scores should be a result of *both* phonemic identity and phonetic goodness. Because, **P** is a “better” member of the category /i/, there should be higher obtained generalization scores for this condition above and beyond those predicted from mere phonemic identity. That is, if PME is due to a shrinking of the perceptual space around the prototype, then the differences between predicted generalization scores for **P** and **NP** conditions should be substantially less than the scores that were obtained. For these comparisons, it is the *differences* in the generalization scores and not the actual scores themselves that are important. Predictions of discriminability using this method understandably underestimate the discriminability of speech sounds (e.g., Miyawaki *et al.*, 1975). If one assumes that other bases for discrimination other than perceived identity (e.g., guessing, spectral difference) are equally potent for the **P** and **NP** conditions, then it is appropriate to compare the predicted differences and the observed differences.

Average percent /i/ and /e/ identifications from experiment 1 for vowels presented in context were used to estimate the probability of perceived identifications. Based solely on predicting discrimination from identification, the predicted generalization score is 44.40 for the **P** condition and 35.83 for the **NP** condition resulting in a predicted difference in generalization scores of 8.57. The obtained difference in generalization scores was 5.61. The difference predicted on the basis of identification functions from experiment 1 is actually larger than the obtained difference. This is opposite the prediction of PME theory. The prototype **P** is *more* discriminable from its neighbors as compared to **NP** than would be expected according to classic findings of Categorical Perception.

Iverson and Kuhl (1995) cite the comparison of discrimination to the left and right of **P** as a measure of PME that is less confounded by the effects of perceived identity. The predicted difference between generalization scores to the left (54.79) and to the right (33.97) of **P** is 20.82. The obtained difference was 13.86. Again, when one uses identifications percentages from a labeling task of vowels in context, the predicted difference in generalizations is *greater* than that actually obtained in a discrimination task. If category goodness plays any role in predicting discriminations it must be in a manner opposite to what has previously been suggested! Good representatives of a vowel category are equally or more discriminable than poorer representatives for within-category comparisons. Following application of time-worn models of Categorical Perception and classical findings of phonetic perception in context, PME theory makes the wrong predictions.

III. GENERAL DISCUSSION

Traditionally, discriminability of speech sounds has been considered solely a function of differences in acoustic structure and phonemic identity. In contrast, PME purportedly demonstrated that the interior structure of a single pho-

netic category was a determinant of discriminability. It is essential to this claim that the speech sounds being discriminated belong to the same phonetic category. Iverson and Kuhl (1995) used stimuli that were labeled in isolation as /i/ more than 50% of the time. However, perceived category membership and representativeness are substantially variable with changes in context. In experiment 1, vowels labeled as good exemplars of /i/ in the context of **NP** were labeled as /e/ when presented with **P**. These results are problematic for interpretations of PME for two reasons. First, if phonetic identity and category “goodness” are so pliable, then it becomes difficult to claim that they indicate the structure of stable mental representations of phonetic categories. The theoretical value of “prototypical” and “nonprototypical” vowels is questionable when their very phonetic identity can be so easily manipulated.

More troublesome for the theory of perceptual magnets is that these shifts in vowel identity are apt to create *inter*-categorical comparisons in a discrimination task. Because the effect of context is contrastive, pairs of vowels will tend to be labeled as different phonemes. This is much more likely to occur for poor exemplars of a category, such as in the **NP** condition, because the identity of these vowels is more ambiguous. It has long been known that cross-boundary discriminations are easier than within-category discriminations. Therefore, one would predict that poor exemplars of a category would be easier to discriminate than relatively good unambiguous exemplars. This is exactly the finding that has been come to be known as PME.

This is all in marked contrast to the generally accepted interpretation of PME as a demonstration of the shrinking of perceptual space around prototypical members of a phonetic category. The results of experiment 2 offer no evidence for a smaller perceptual space or greater similarity amongst good vowel exemplars. In fact, the obtained difference in discriminability of good versus poor tokens was smaller than predicted from identification functions. There is no evidence that the prototype (**P**) served as a “perceptual magnet.” Thus the results do not support the Natural Language Magnet model of phonetic category-acquisition (Kuhl, 1993) or various computational models which result in smaller response spaces around prototypical category members (e.g., Guenther and Gjaja, 1996; Lacerda, 1998).

This is not to say that phonetic-category structure is not an important indicator of processes involved in the perception of speech. In experiment 1, vowels that were judged to be good exemplars when presented in isolation were more effective at changing the identity of stimuli with which they were paired (see Fig. 2). This is similar to previous results demonstrating that some stimuli serve as more effective “adapters” in “selective adaptation” tasks (Miller *et al.*, 1983).

Questions remain as to the genesis and purpose of graded structures within phonetic categories. It is quite possible that categories come to mirror the statistical characteristics of experienced phonetic distributions. Kluender *et al.* (under review) have collected data from birds, humans and computational models responding to vowel distributions, and the data suggest that category structure is a result of the input

distribution itself and its relation to other phonetic categories. It will be beneficial for speech scientists to continue investigating the structure of phonetic categories and their relation to natural speech perception. What is challenged by the present data, however, is whether a link between “goodness” judgments and discriminability is informative about the nature of categories. Once confounding effects are accounted for, putative differences in discriminability are neutralized and perceptual magnets are depolarized.

¹The authors prefer the term “functional equivalence class” instead of “category,” because “category” has a long theoretical history and is used in several different ways that obscure important conceptual differences. However, for ease of reading and to facilitate comparison with other reports, the more familiar term will be used. In this report, “category” will be used to refer to a set of stimuli that are linguistically functionally equivalent for a particular perceiver and to the pattern of responses to those stimuli. That is, a category is defined by identification functions. The term is *not* meant to refer to mental representations of external sets of stimuli which may determine the identification responses, though this is a common use of the term.

²Both “Categorical Perception” and “Perceptual-Magnet Effect” are meant to refer to patterns of results which are obtained using experimental tasks. Both terms have also been used as descriptions of theoretical constructs which are the purported cause of these effects. The use of the terms should not imply that the authors are advocating either theoretical view.

³The assumption that discrimination is based solely on identification is quite different from the assumptions underlying signal-detection analysis. Thus predictions of *d'* scores using the Liberman *et al.* (1957) equations would be inappropriate.

Davis, K., and Kuhl, P. K. (1994). “Tests of the perceptual magnet effect for American English /k/ and /g/,” *J. Acoust. Soc. Am.* **95**, 2976.

Eimas, P. D. (1963). “The relation between identification and discrimination along speech and nonspeech continua,” *Language and Speech* **6**, 206–217.

Fry, D. B., Abramson, A. S., Eimas, P. D., and Liberman, A. M. (1962). “The identification and discrimination of synthetic vowels,” *Language and Speech* **5**, 171–189.

Grieser, D., and Kuhl, P. K. (1989). “Categorization of speech by infants: Support for speech-sound prototypes,” *Develop. Psych.* **25**, 577–588.

Guenther, F. H., and Gjaja, M. N. (1996). “The perceptual magnet effect as an emergent property of neural map formation,” *J. Acoust. Soc. Am.* **100**, 1111–1121.

Iverson, P., Diesch, E., Siebert, C., and Kuhl, P. K. (1994). “Cross-language tests of the perceptual magnet effect for /r/ and /l/,” *J. Acoust. Soc. Am.* **96**, 3228.

Iverson, P., and Kuhl, P. K. (1995). “Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling,” *J. Acoust. Soc. Am.* **97**, 553–562.

Iverson, P., and Kuhl, P. K. (1996). “Influences of phonetic identification and category goodness on American listeners’ perception of /r/ and /l/,” *J. Acoust. Soc. Am.* **99**, 1130–1140.

Klatt, D. H. (1980). “Software for a cascade/parallel formant synthesizer,” *J. Acoust. Soc. Am.* **67**, 971–995.

Kluender, K. R., Lotto, A. J., Holt, L. L., and Bloedel, S. L. (under review). “Role of experience for language-specific functional mappings of vowel sounds.”

Kuhl, P. K. (1991). “Human adults and human infants show a ‘perceptual magnet effect’ for the prototypes of speech categories, monkeys do not,” *Percept. Psychophys.* **50**, 93–107.

Kuhl, P. K. (1992). “Infants’ perception and representation of speech: Development of a new theory,” in *Proceedings of the International Conference on Spoken Language Processing*, edited by J. J. Ohala, T. M. Nearey, B. L. Derwing, M. M. Hodge, and G. E. Wiebe (University of Alberta, Edmonton, Alberta), pp. 449–456.

Kuhl, P. K. (1993). “Innate predispositions and the effects of experience in speech perception: The native language magnet theory,” in *Developmental Neurocognition: Speech and Face Processing in the First Year of Life*, edited by B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. McNeilage, and J. Morton (Kluwer Academic, Dordrecht, The Netherlands), pp. 259–274.

- Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., and Lindblom, B. (1992). "Linguistic experience alters phonetic perception in infants by 6 months of age," *Science* **255**, 606–608.
- Lacerda, F. (1998). "An exemplar-based account of emergent phonetic categories," *J. Acoust. Soc. Am.* **103**, 2980(A).
- Lieberman, A. M., Harris, K. S., Hoffman, H. S., and Griffith, B. C. (1957). "The discrimination of speech sounds within and across phoneme boundaries," *J. Exp. Psychol.* **54**, 358–368.
- Lively, S. E. (1993). "An examination of the perceptual magnet effect," *J. Acoust. Soc. Am.* **93**, 2423.
- Lotto, A. J., Kluender, K. R., and Holt, L. L. (1996). "Effects of language experience on perceptual organization of vowel sounds," paper presented at the Fifth Conference of Laboratory Phonology, Northwestern University, July 6.
- Macmillan, N. A., and Creelman, D. (1991). *Detection Theory: A User's Guide* (Cambridge U.P., New York).
- Miller, J. L., Connine, C. M., Schermer, T. M., and Kluender, K. R. (1983). "A possible auditory basis for internal structure of phonetic categories," *J. Acoust. Soc. Am.* **73**, 2124–2133.
- Miller, J. L., and Volaitis, L. E. (1989). "Effect of speaking rate on the perceptual structure of a phonetic category," *Percept. Psychophys.* **46**, 505–512.
- Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. M., Jenkins, J. J., and Fujimura, O. (1975). "An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English," *Percept. Psychophys.* **18**, 331–340.
- Nearey, T. M. (1989). "Static, dynamic, and relational properties in vowel perception," *J. Acoust. Soc. Am.* **85**, 2088–2113.
- Peterson, G. E., and Barney, H. L. (1952). "Control methods used in the study of the vowels," *J. Acoust. Soc. Am.* **24**, 175–184.
- Pisoni, D. B. (1971). "On the nature of categorical perception of speech sounds," Doctoral thesis, University of Michigan.
- Polka, L., and Werker, J. F. (1994). "Developmental changes in perception of non-native vowel contrasts," *J. Exp. Psychol.* **20**, 421–435.
- Posner, M. I., and Keele, S. W. (1968). "On the genesis of abstract ideas," *J. Exp. Psychol.* **77**, 28–38.
- Posner, M. I., and Keele, S. W. (1970). "Retention of abstract ideas," *J. Exp. Psychol.* **83**, 304–308.
- Rosch, E. H. (1975). "Cognitive representations of semantic categories," *J. Exp. Psychol.* **3**, 193–233.
- Stevens, K. N., Liberman, A. M., Studdert-Kennedy, M., and Ohman, S. E. G. (1969). "Crosslanguage study of vowel perception," *Language and Speech* **12**, 1–23.
- Sussman, J. E., and Lauckner-Morano, V. J. (1995). "Further tests of the perceptual magnet effect in the perception of [i]: Identification and change/no-change discrimination," *J. Acoust. Soc. Am.* **97**, 539–552.
- Thompson, C. L., and Hollien, H. (1970). "Some contextual effects on the perception of synthetic vowels," *Language and Speech* **13**, 1–13.
- Volaitis, L. E., and Miller, J. L. (1992). "Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories," *J. Acoust. Soc. Am.* **92**, 723–735.
- Wood, C. C. (1976). "Discriminability, response bias, and phoneme categories in discrimination of voice onset time," *J. Acoust. Soc. Am.* **60**, 1381–1389.