# An analytical error invalidates the "depolarization" of the perceptual magnet effect

Frank H. Guenther
Department of Cognitive and Neural Systems
Boston University
677 Beacon Street
Boston, MA 02215

and

Speech Communication Group
Research Laboratory of Electronics
Massachusetts Institute of Technology
Email: guenther@cns.bu.edu

## Abstract

In a recent article, Lotto et al. [*JASA*, **103**, pp. 3648-3655, 1998] presented experiments investigating the role played by perceived phonemic identity in demonstrations of decreased discriminability for prototypical vowel sounds. The authors interpreted their results as evidence against shrinkage of perceptual space near vowel category prototypes. In this letter, it is shown that this interpretation is based on a flawed data analysis in which a key confounding term has been neglected.

The perceptual magnet effect (PME) reported by Kuhl (1991) has been a hotly debated topic in the speech perception literature in recent years. According to the Kuhl et al. account, the PME is a case of perceptual space being "shrunk" near prototypical examples of vowel sounds, as compared to near non-prototypical examples, such that the same spectral difference seems smaller near prototypical vowels than near non-prototypical vowels. In a recent JASA article entitled "Depolarizing the perceptual magnet effect", Lotto, Kluender, and Holt (1998) attempt to show that the PME is nothing more than a "further demonstration that general discriminability is greater for cross-category stimulus pairs than for within-category pairs" (p. 3648). In other words, Lotto et al. posit that PME experimental results can be completely explained by a simple assumption, rooted in classical treatments of categorical perception (CP), without reference to shrunken representations of spectral space near prototypical vowels. The assumption is that differences in the ability to discriminate prototypical vs. non-prototypical sounds can be attributed to differences in the probability that two sounds in a discrimination pair are given different phonemic labels by the listener. That is, two non-prototypical vowel sounds are more likely to be given different phonemic labels by the listener, and are thus more likely to be discriminated by the listener, than two prototypical vowel sounds.

In order to assess this claim, Lotto et al. perform some interesting PME experiments that correct for contextual effects on phonemic identification. After analyzing their experimental results, Lotto et al. claim that "Following application of time-worn models of Categorical Perception and classical findings of phonetic perception in context, PME theory makes the wrong predictions" (p. 3653). However, an analytical error in the authors' analysis invalidates this claim. Before investigating this issue, it is useful to define some terminology. In the following, P will be used to refer to the prototype stimulus (i.e., a "good" example of /i/) and NP to the non-prototype (a "bad" example of /i/). The probability of generalizing two sounds (i.e., responding that they are the same when they are actually different) will be denoted by p(gen). Finally, following Lotto et al. (1998), the term "PME theory" will be used to refer to the assertion by Kuhl and colleagues that the PME represents a shrinking of perceptual space in the neighborhood of prototypical sounds and that it is not simply the result of "phonemic identity" processes as described above.

Lotto et al. compare the difference between the generalization of P and NP sounds as measured in their experiment with a predicted difference in generalization of P and NP sounds obtained under the assumption that consideration of phonemic identity alone accounts for the difference. Quoting Lotto et al. (p. 3653):

> "... if PME is due to a shrinking of the perceptual space around the prototype, then the differences between predicted generalization scores [predicted under the assumption that phonetic identity alone is responsible] for P and NP conditions should be substantially less than the scores obtained [in their experiment]. 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 observed differences."

The assumption underlying this comparison can be written mathematically as follows:

$$p(gen) = p(gen\ by\ PI) + p(gen\ by\ spect) \qquad \text{Equation 1}$$

where p(gen by PI) is the probability that the two sounds in a discrimination trial are identified as the same phoneme by the listener (i.e., the probability of generalizing based on phonemic identity), and p(gen by spect) is the probability that the two sounds cannot be discriminated based on spectral distance or other means. Lotto et al. use their identification scores from Experiment 1 to predict the difference in generalization scores for P and NP based on phonemic identity alone:

$$Diff_PI = p(gen\ by\ PI, P) - p(gen\ by\ PI, NP). \qquad \text{Equation 2}$$

where p(gen by PI, P) is the probability of generalization based on phonemic identity in the prototype case, etc. The difference predicted by phonemic identity alone is then compared to the

measured difference in generalization scores, which, from Equation 1, is assumed to correspond to the following:

$$\begin{aligned}\text{Diff_measured} &= p(\text{gen, P}) - p(\text{gen, NP}) \\ &= p(\text{gen by PI, P}) + p(\text{gen by spect, P}) - \\ &\quad [p(\text{gen by PI, NP}) + p(\text{gen by spect, NP})]\end{aligned}$$

$$= \text{Diff_PI} + p(\text{gen by spect, P}) - p(\text{gen by spect, NP}). \qquad \text{Equation 3}$$

Because the difference as predicted from phonemic identity alone (Diff_PI, calculated to be 8.57 by Lotto et al.) is greater than the difference as determined from the measured generalization scores (Diff_measured, which was 5.61 in the Lotto et al. experiment), the authors conclude that p(gen by spect, P) must be less than p(gen by spect, NP), which is the opposite of what is predicted by PME theory: according to this analysis, it is easier to discriminate, based on spectral difference, sounds near the prototype than sounds near the non-prototype. This prompts the authors to claim (p. 3653) that "If category goodness plays any role in predicting discriminations it must be in a manner opposite to what has previously been suggested!"

The following hypothetical example can be used to highlight the problem with the Lotto et al. analysis. Assume that, in the NP case, all test pairs fall into different phonemic categories. That is, p(gen by PI, NP) = 0%. Assume further that all test pairs in the P case fall into the same phonemic category, so p(gen by PI, P) = 100%. For this case, Diff_PI is 100%. Next, assume that PME theory is correct, such that p(gen by spect) = 30% for the NP case and 60% for the P case. That is, perceptual space is warped around the prototype such that the same spectral difference is half as likely to be discriminated near P than near NP. Since all NP stimuli can be discriminated by phonemic identity alone, p(gen, NP) = 0% regardless of p(gen by spect, NP), and since no P stimuli can be distinguished by phonemic identity, p(gen, P) = p(gen by spect, P) = 60%. The value of Diff_measured as calculated from generalization data following the Lotto et al. method thus would be 60%. Since Diff_PI is 100% and Diff_measured is only 60%, using the Lotto et al. logic one would conclude that there is a very strong anti-magnet effect, as Lotto et al. conclude from their experimental results. Clearly this is incorrect in our hypothetical

example since p(gen by spect) is twice as large for the P case than the NP case; that is, there is a strong magnet effect, not a strong anti-magnet effect.

Although this example represents an extreme case for the sake of illustration, the same problem exists for less extreme cases. The problem lies with the assumption summarized in Equation 1. Generalization will take place only if a stimulus pair generalizes based on phonemic *and* based on spectral difference. Equation 1 is thus incorrect, and the following equation should have been used instead:

p(gen) = p(gen by PI *and* gen by spect)

= p(gen by PI) + p(gen by spect) - p(gen by PI *or* gen by spect) Equation 4

This form of the equation follows from a fundamental axiom of probability theory (e.g., Ross, 1984):

p(A *and* B) = p(A) + p(B) - p(A *or* B). Equation 5

The exclusion of the p(gen by PI *or* gen by spect) term completely confounds the Lotto et al. analysis and conclusions. The Lotto et al. comparison of predicted and measured differences in generalization would be valid only if p(gen by PI *or* gen by spect) were the same in the P and NP conditions, since in this case they would cancel each other in the equation for Diff_measured. Given that the Lotto et al. experimental results indicate that p(gen by PI) is 8.57% less for NP than for P, it is not reasonable to assume that this is the case. It can be shown that, for PME theory to be supported by the Lotto et al. results, p(gen by PI *or* gen by spect) need only be 3% less in the NP case than in the P case. Since there is no way to estimate the values of p(gen by PI *or* gen by spect) for the P and NP conditions based on the Lotto et al. experimental results (nor the results of any other experiment of which I am aware), the Lotto et al. results do not provide evidence against PME theory despite the authors' claims to the contrary.

Of course, a lack of evidence against PME theory should not be interpreted as evidence supporting PME theory. It appears clear from Kuhl (1991) and followup studies including Lotto

et al. (1998) that our auditory perceptual space for vowels is warped in the following sense: discriminability as measured by $d'$ is worse for some parts of acoustic space than other parts, and this appears to have something to do with the locations of vowel categories in the listener's language. However, the relative roles of labeling processes, category prototypes, and other influences in this warping remains to be clarified.

**Acknowledgements**

Supported by grant R29 DC02852 from the National Institute on Deafness and other Communication Disorders.

## References

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

Lotto, A.J., Kluender, K.R., and Holt, L.L. (1998). Depolarizing the perceptual magnet effect. *Journal of the Acoustical Society of America*, **103**(6), pp. 3648-3655.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A.M., Jenkins, J.J., and Fujimora, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception and Psychophysics*, **18**, pp. 331-340.

Ross, S. (1976). A first course in probability. New York: Macmillan Publishing Company.