Reply to "An analytical error invalidates the 'depolarization' of the perceptual magnet effect" [J. Acoust. Soc. Am. 107, 3576–3577 (2000)]

Andrew J. Lotto[a)]
Department of Psychology & Parmly Hearing Institute, Loyola University Chicago, 6525 North Sheridan Road, Chicago, Illinois 60626

(Received 24 November 1999; revised 15 December 1999; accepted 28 January 2000)

In a previous paper, Lotto *et al.* [J. Acoust. Soc. Am. **103**, 3648–3655 (1998)] claimed that experiments investigating a perceptual magnet effect (PME) failed to demonstrate a novel effect of goodness judgments on speech sound discrimination due to a confound in the procedure. New analyses of these data and a review of the recent literature fail once again to support the existence of PME. [S0001-4966(00)01105-X]

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

The perceptual magnet effect (PME) has been offered as a novel perceptual effect in work on the discrimination of speech sounds (e.g., Kuhl, 1991; Iverson and Kuhl, 1995). The data from several experiments have been interpreted as an indication that category ''goodness'' influences discrimination above and beyond the effects of spectral distance and phonemic identification. For example, Iverson and Kuhl (1995) demonstrated that vowel tokens that were labeled as ''good'' members of the /i/ category [the **Prototype** (**P**) condition] were harder to discriminate from neighboring exemplars than were members that were considered poor examples of /i/ [the **Nonprototype** (**NP**) condition]. Importantly, these discriminations were all considered to be *intra*category comparisons. Lotto *et al.* (1998) suggested that this experiment contained a confound that made the interpretation in terms of PME problematic. The problem is that identification functions were collected for stimuli presented in isolation and discrimination occurred for stimuli presented in pairs. Lotto *et al.* (1998) demonstrated that the perceived phonemic identities of these stimuli were greatly affected by context in the pair condition and, therefore, that the identification functions obtained in isolation were not good indicators of identity in the discrimination task.

Of greatest theoretical significance was the fact that a shift in vowel identity was more likely to occur for the poor vowel exemplars (i.e., the **NP** condition) than for the good exemplars (i.e., the **P** condition). Thus, discriminations in the **NP** case were more likely to be *intra*categorical and the **P** discriminations were more likely to be *intra*categorical. If this is the case, then the PME is not a novel demonstration. It has long been observed that discriminations between sounds from different phonemic categories are easier than discrimination of sounds from within one speech category (e.g., Liberman *et al.*, 1957; Pisoni, 1971; Wood, 1976).

To see if there was any evidence for an effect of category goodness on discrimination above and beyond the established effects of spectral distance and phonemic identity, Lotto *et al.* examined the *generalization scores* (misses). This measure had been used by Kuhl in previous demonstrations of PME (Grieser and Kuhl, 1989; Kuhl, 1991; Kuhl *et al.*, 1992). The obtained difference in generalization scores for the **P** and **NP** conditions was compared to the predicted difference in generalization scores based solely on phonemic identification. In making this comparison, it was presumed that effects of spectral difference were equivalent (these were experimentally equated in the two conditions by using the mel scale) and that any joint probabilities were equivalent between conditions. Because there is no way in this experimental setup to estimate the joint probabilities (e.g., the probability of discriminating on the basis of phonemic identification *and* perceptual space [goodness] differences), this was considered the most conservative route. The results demonstrated that the obtained scores did not differ from the predicted scores in a way that would serve as evidence for PME. There was no evidence that the differences between the **P** and **NP** conditions were due to category goodness. The conclusion reached was that there was a lack of positive evidence for PME.

Guenther (2000) points out that if the joint probabilities differ in the **P** and **NP** conditions, then the outcomes of generalization difference scores could tell a different story. He presents a clear and powerful example in which a warped perceptual space could lead to generalization difference scores that incorrectly appear to demonstrate no magnet effect. It should be noted that Lotto *et al.* (1998) were not attempting to disprove the existence of PME with the difference equations, but that they were looking for positive evidence for PME using Kuhl's generalization score. However, Guenther's (2000) point is significant for the extensive PME literature. Basically, Guenther's analysis leads one to conclude that the use of generalization scores in all of the PME work is of dubious merit.

The following example elucidates the problem. Presume that the data demonstrate that the probability of discrimination based on phonemic identity is equal in both the **P** and **NP** condition of a PME experiment, let's say, 0.50 in either case. Now, presume that the probability of discriminating based on perceptual space distance (warped in relation to goodness) is 0.60 for the **P** condition and 0.40 for the **NP**

[a)]Electronic mail: alotto@luc.edu

condition. That is, the better members of a category are more easily discriminated from neighbors than are poorer members. This is the opposite of PME: it is an anchoring effect. What will the generalization scores show? This depends on the joint probability of discriminating based on phonemic identification *and* perceptual space distance. Let's say that for the **P** condition the joint probability is 0.40 and that for the **NP** condition it is 0.10. Plug these values into Guenther's equation (???):

$$[1-(0.50+0.60)-0.40]-[1-(0.50+0.40)-0.10]=0.10.$$

The difference in generalization scores for this examples looks like PME: generalization scores are higher for the **P** condition than for the **NP** condition. However, the values presented here are for an underlying perceptual space that is warped in the opposite direction. It appears that if one allows the joint probabilities to differ, then generalization scores can be quite deceiving.

Guenther (2000) suggests that this analysis of joint probabilities invalidates the conclusions of Lotto *et al.* (1998) that there is no evidence for PME. However, Guenther's analysis, in fact, supports this conclusion by pointing out that previous empirical work using generalization scores do not provide compelling evidence for PME. Note that this is true for any experiment that allows cross-category pairs in discrimination. If phonemic identity is equivalent for all stimuli, then the joint probabilities are zero. Lotto *et al.* (1998) demonstrated that the discrimination pairs in Iverson and Kuhl (1995) often contained cross-category comparisons. Furthermore, the strong context effect on perceived identity that Lotto *et al.* demonstrated for stimuli in pairs indicates that many of the demonstrations of PME probably have contained cross-category discriminations. Thus, the analysis of Guenther (2000) does not invalidate the conclusions of Lotto *et al.* (1998), but rather the two works taken together provide a challenge to the entire PME enterprise at least in relation to the use of generalization scores.

I. NEW DATA ANALYSIS

If the generalization score is a dubious measure for testing PME, is there another analysis that will provide information about the effect of goodness ratings on discrimination above and beyond the effects of spectral distance and phonemic identity? One possibility is to model discrimination using multiple regression to see if goodness ratings account for a significant amount of the variance.

To further test for PME, the discrimination data from experiment 2 of Lotto *et al.* (1998) were modeled using stepwise regression with variables representing spectral distance, discrimination predicted from perceived phonemic identity, goodness rating of the standard (P or NP), and average goodness rating of both members of each discrimination pair. In stepwise regression, factors are entered into the model depending on the amount of variance (*R*-squared) accounted for in the dependent variable. A significance test determines which factors get entered into the model with the factors accounting for the most variance being added first. For this analysis, the factors were the following.

(1) Spectral distance: Euclidean distance between the standard and the comparison stimuli in the mel-scaled first formant×second formant space (30, 60 or 90 mel).
(2) Probability that both stimuli in pair will have the same perceived phonemic identity (PI): This is the predicted probability of a miss dependent solely on the identification functions for stimuli presented in pairs from Lotto *et al.* (1998) experiment 1.
(3) Goodness rating for standard: Mean goodness rating for P or NP for the particular pair being discriminated as determined from experiment 1 of Lotto *et al.*
(4) Average goodness rating for pair: The average goodness rating for the two stimuli in the pair as obtained in experiment 1 of Lotto *et al.*

Evidence for PME would correspond to either of these last two factors accounting for a significant amount of the variance in discrimination scores beyond the variance accounted for by the first two factors.

Two models were calculated: one for the dependent variable of hits (i.e., ''different'' responses to stimuli that are in fact different) and one for the bias-free measure of sensitivity, d' (Green and Swets, 1966). For the model of hits, spectral distance was entered, first having accounted for the most variance [R-square$=0.79$, $F(1,10)=36.67$, $p<0.0001$]. The factor PI was then entered, resulting in a model that accounted for a healthy 91% of the variance in hits [$F(2,9)=46.41$, $p<0.0001$]. Both of the goodness factors failed to contribute significantly to the model (p's >0.67; criterion of selection$=0.05$).

The model of d' data followed the same pattern. Spectral distance was entered into the model first [R-square $=0.86$, $F(1,0)=59.97$, $p<0.0001$] followed by PI [full model, R-square$=0.94$, $F(2,9)=72.24$, $p<0.0001$]. The two goodness-rating factors failed to account for any significant added variance (p's>0.19).

These results offer a compelling lack of evidence for PME. In fact, they demonstrate that spectral distance and phonemic identity do a remarkably good job of predicting discrimination for these vowel stimuli (94% of the variance accounted for in d' scores!).

II. CONCLUSIONS

Taken together, these new regression models, the empirical work of Lotto *et al.* (1998), and the analysis by Guenther (2000) offer a sobering picture of PME. As Lotto *et al.* point out, PME has been one of the more fertile areas of recent speech perception research, spawning numerous empirical, theoretical, and modeling papers. However, the evidence underlying this enterprise is tenuous at best.

Guenther (2000) ends with the statement, ''It appears clear from Kuhl (1991) and follow-up studies including Lotto *et al.* (1998) that our perceptual spaces for vowels is warped.'' But the data as they stand do *not* allow for such a conclusion. The regression models indicate that one can account for discrimination by looking at spectral distance and binary phonemic identity decisions without any additional *intra* category information. In addition, there is a growing literature of empirical work that fails to demonstrate perceptual warping for vowel categories when proper controls are taken. For example, Lively and Pisoni (1995) found no difference in P and NP discrimination when individual prototypes were determined for each subject. Renda *et al.* (1995) failed to find PME for the vowel /•/ Kewley-Port and Neel (1998) and Sharma and Dorman (1998) have also demonstrated failures of PME hypotheses. Thus, there does not appear to be sufficient evidence in the literature to back up Guenther's (2000) conclusion[1] and there appears to be little support for the concept of PME.

[1]There *are* models of perceptual space warping associated with categorical perception (e.g., Livingston *et al.*, 1998). This may be a fruitful avenue for the study of categorization effects on perceptual space.

Green, D. M., and Swets, J. A. (**1966**). *Signal Detection Theory and Psychophysics* (Wiley, New York).

Grieser, D., and Kuhl, P. K. (**1989**). ''Categorization of speech by infants: Support for speech-sound prototypes,'' Dev. Psychol. **25**, 577–588.

Guenther, F. H. (**2000**). ''An analytical error invalidates the 'depolarization' of the perceptual magnet effect,'' J. Acoust. Soc. Am. **107**, 3576–3577.

Harnad, S., Hanson, S. J., and Lubin, J. (**1995**). ''Categorical perception and the evolution of supervised learning in neural nets,'' in *Proceedings of the AAAI Spring Symposium on Machine Learning of Natural Language and Ontology*, edited by D. W. Powers and L. Uhr (DFKI, Kaiserslautern, Germany), pp. 65–74.

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.

Kewley-Port, D., and Neel, A. T. (**1998**). ''Relation between discrimination and identification of English vowels,'' J. Acoust. Soc. Am. **103**, 2983.

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., 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.

Liberman, 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., and Pisoni, D. B. (**1995**). ''On prototypes and phonetic categories: A critical assessment of the perceptual magnet effect in speech perception,'' J. Exp. Psych.: Hum. Percept. Perform. **23**, 1665–1679.

Livingston, K. R., Andrews, J. K., and Harnad, S. (**1998**). ''Categorical perception effects induced by categorical learning,'' J. Exp. Psych.: Learning Memory and Cognition **24**, 732–753.

Lotto, A. J., Kluender, K. R., and Holt, L. L. (**1998**). ''Depolarizing the perceptual magnet effect,'' J. Acoust. Soc. Am. **103**, 3648–3655.

Pisoni, D. B. (**1971**). ''On the nature of categorical perception of speech sounds,'' Doctoral Thesis, University of Michigan.

Renda, S. C., Hawks, J. W., and Klich, R. (**1995**). ''An investigation of the perceptual magnet effect in adults,'' J. Acoust. Soc. Am. **97**, 3420.

Sharma, A., and Dorman, M. F. (**1998**). ''Exploration of the perceptual magnet effect using the mismatch negativity auditory evoked potential,'' J. Acoust. Soc. Am. **104**, 511–517.

Wood, C. C. (**1976**). ''Discriminability, response bias, and phoneme categories in discrimination of voice onset time,'' J. Acoust. Soc. Am. **60**, 1381–1389.