1-1-2004

Evaluation of the perceptual magnet effect and categorical perception for musical timbre

Melissa Kay Griffith
University of Nevada, Las Vegas

Follow this and additional works at: <https://digitalscholarship.unlv.edu/rtds>

Repository Citation

Griffith, Melissa Kay, "Evaluation of the perceptual magnet effect and categorical perception for musical timbre" (2004). *UNLV Retrospective Theses & Dissertations*. 1771.
<http://dx.doi.org/10.25669/xnzd-1e8r>

This Thesis is protected by copyright and/or related rights. It has been brought to you by Digital Scholarship@UNLV with permission from the rights-holder(s). You are free to use this Thesis in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you need to obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/or on the work itself.

This Thesis has been accepted for inclusion in UNLV Retrospective Theses & Dissertations by an authorized administrator of Digital Scholarship@UNLV. For more information, please contact digitalscholarship@unlv.edu.

EVALUATION OF THE PERCEPTUAL MAGNET EFFECT AND CATEGORICAL
PERCEPTION FOR MUSICAL TIMBRE

by

Melissa Kay Griffith

Bachelor of Science
University of South Carolina, Upstate
2001

A thesis submitted in partial fulfillment
of the requirements for the

Master of Arts Degree in Psychology
Department of Psychology
College of Liberal Arts

Graduate College
University of Nevada, Las Vegas
May 2005

UMI Number: 1428558

INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

UMI[®]

UMI Microform 1428558

Copyright 2005 by ProQuest Information and Learning Company.

All rights reserved. This microform edition is protected against unauthorized copying under Title 17, United States Code.

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

Thesis Approval

The Graduate College

University of Nevada, Las Vegas

April 8, 2005

The Thesis prepared by

Melissa Kay Griffith

Entitled

Evaluation of the Perceptual Magnet Effect and Categorical Perception
for Musical Timbre

is approved in partial fulfillment of the requirements for the degree of

Master of Arts in Psychology

Examination Committee Co-Chair

Examination Committee Chair

Dean of the Graduate College

Examination Committee Member

Examination Committee Member

Graduate College Faculty Representative

ABSTRACT

Evaluation of the Perceptual Magnet Effect and Categorical Perception for Musical Timbre

by

Melissa Kay Griffith

Dr. Karen Kemtes, Examination Committee Chair
Assistant Professor of Psychology
University of Nevada, Las Vegas

Recent auditory research has raised fundamental questions about the perceptual magnet effect (PME), where discrimination performance is poorer for stimuli that approach best exemplars of a phonetic category. It has been suggested that the effect reflects inter-categorical comparisons, and might not generalize to nonspeech. Three experiments addressed these concerns. In Experiment 1 prototype and non-prototype stimuli were determined from goodness ratings of synthesized (violin) timbres varying in center frequencies of F1 and F2 formants. Experiment 2 evaluated for a PME using discrimination data, and influences from other categories by comparing goodness ratings from stimuli in prototype, non-prototype, and no context conditions. Experiment 3 used labeling and discrimination tasks to assess if categorical perception occurs with timbres. Despite having stimuli reliably identified to be within the intended category, no PME was found. It is suggested that the PME, if a real phenomenon, is too difficult to tease apart from categorization tendencies.

TABLE OF CONTENTS

ABSTRACT.....	iii
LIST OF FIGURES	vi
ACKNOWLEDGMENTS	vii
CHAPTER 1 INTRODUCTION	1
CHAPTER 2 LITERATURE REVIEW	4
Support for the Perceptual Magnet Effect	5
Criticisms of the Perceptual Magnet Effect	12
The Perceptual Magnet Effect and Musical Stimuli	18
Purpose of the Study	24
CHAPTER 3 EXPERIMENT 1	27
Method	27
Results and Discussion	33
CHAPTER 4 EXPERIMENT 2	40
Method	41
Results and Discussion	44
CHAPTER 5 EXPERIMENT 3	56
Method	58
Results and Discussion	60
CHAPTER 6 GENERAL DISCUSSION	65
Comparisons with Speech and Musical Research	67
Theoretical Implications	70
Future Directions	77
APPENDIX I TABLES AND FIGURES	79
APPENDIX II MUSICAL HISTORY QUESTIONNAIRE.....	89
REFERENCES	91

VITA.....	98
-----------	----

LIST OF FIGURES

Figure 1	Values of Stimuli (in Hz) That Were Used in Experiment 1	80
Figure 2	Mean Goodness Ratings and Percentage of Within-category Identifications of All Stimuli Used in Experiment 1	81
Figure 3	Diagonal Vector of Stimuli Used in Experiment 2	82
Figure 4	Mean Goodness Ratings and Within-category Identification Responses for Stimuli in Task With No Context in Experiment 2	83
Figure 5	Mean Goodness Ratings and Within-category Identification Responses for Stimuli in Task With Context in Experiment 2	84
Figure 6	$P(c)_{\max}$ Scores for Each Stimulus and its Corresponding Standard from Experiment 2	85
Figure 7	Stimuli Used in Experiment 3	86
Figure 8a	Experiment 3 Labeling and Discrimination Functions (Subjects 1 – 6)	87
Figure 8b	Experiment 3 Labeling and Discrimination Functions (Subjects 7 – 10)	88

ACKNOWLEDGMENTS

My sincerest gratitude is extended to Dr. Michael Hall for inspiring the project, initially serving as committee chair, guiding me through from start to finish and providing invaluable advice that will continually guide my professional development. Michael Hall's extraordinary dedication and admirable professional ethic has made being his student truly an honor. Special thanks also are extended to Dr. Karen Kemtes for serving as committee chair during the final stages of the project and for the helpful advice and support. I also would like to thank committee member Dr. Gretchen Kambe for the helpful academic and professional guidance. Many thanks are extended to Dr. Alice Corkill for serving on the committee, for the helpful advice and for the support in recruiting subjects that played a vital role in getting this project finished. Many thanks to several graduate students, including Barbara Badgett, Michelle Christenson, Ed Crossman, Miriam Dunbar, Linda Kelso, Matt McCrudden, Karen Reid and Alex Spetariu, and several laboratory members, including Dahlia Espeut, Duncan Leaf, David Menard, Jenny Roche and Chris Speirer for enthusiastically serving as subjects on many occasions. Gratitude also is extended to many introductory psychology students for aiding in data collection. I also would like to thank my fiancé, Brian Gregg, for being my practice audience, for staying up all night to keep my coffee cup full while I worked on this project and for the amazing support that kept me calm throughout this process. I also would like to thank numerous friends and family members (especially my mother) for the encouragement and support.

CHAPTER 1

INTRODUCTION

Recent work in speech perception has indicated that phonetic categories may be organized according to category prototypes (e.g., Grieser & Kuhl, 1989). Category prototypes have been empirically defined as the highest rated example of a phonetic category on a perceptual goodness scale (e.g., Kuhl, 1991). Prototypes have been claimed to be the center of speech categories (Kuhl, 1991). Evidence for this assertion has been provided by the finding that discrimination among speech sounds close to category prototypes is significantly worse than around non-prototypic representations. This claimed shrinking of perceptual space around prototypes has been termed the perceptual magnet effect (see Kuhl, 1991). This effect, if reliable, has important implications for how we categorize speech information. The perceptual magnet effect suggests that category centers are the basis for speech representation. This claim is contrary to previous theories of speech perception, which suggest that category boundaries (see Repp, 1984 for a review), are fundamental for representation.

Though the perceptual magnet effect has received much support (e.g., Aaltonen, Eerola, Hellström, Uusipaikka, & Lang, 1997; Iverson & Kuhl, 1995, 2000), several experiments have been reported that challenge the effect (e.g., Lively & Pisoni, 1997; Lotto, Kluender, & Holt, 1998; Sussman & Lauckner-Morano, 1995). The central

criticism is that the speech stimuli that have been used may not represent a single phonetic category. Thus, better discrimination around non-prototypic exemplars could be claimed to be a result of inter-categorical interference, rather than a perceptual magnet effect based upon the representation of a single phonemic category.

In the present study, three experiments were designed to address the issue of inter-categorical interference in the speech perceptual magnet effect literature. The present investigation attempted to resolve this issue in the speech work by utilizing a stimulus set that represented a broadly defined category with minimal inter-categorical interference to assess a potential magnet effect.

An additional finding that makes claims of a perceptual magnet effect questionable is the failure of experiments utilizing musical stimuli to find the effect (e.g., Acker, Pastore, & Hall, 1995; Schellenberg, 2002). The failure of researchers to provide convincing evidence of a musical perceptual magnet effect (e.g., Barrett, 1998) is problematic because the effect is claimed to be based on experience (Kuhl, 1993). Given that the magnet effect is argued to be a result of experience with sounds, the effect should generalize to nonspeech, musical stimuli. The results of the musical studies have been difficult to interpret, however, because of the use of stimuli that are not analogous to speech stimuli.

The present study was motivated by concerns that the nonspeech (musical) stimuli used thus far to study the perceptual magnet effect are not analogous to vowel stimuli. Thus, one purpose of this study was to determine if a perceptual magnet effect occurs in a musical paradigm using stimuli that are more analogous to vowels than the stimuli that have previously been used. An analysis similar to that which was used to address the

speech perceptual magnet effect issues, as described above, was extended to address the musical concerns.

CHAPTER 2

LITERATURE REVIEW

The perceptual implications for comparing incoming stimuli to a set of stored exemplars and idealized prototypes have motivated much research in speech perception (e.g. Grieser & Kuhl, 1989). It has been well established that within-category stimuli surrounding a prototype seem to be graded, i.e. increasing in perceived goodness as an instance of the category approaches an empirically defined prototype from a set of stimuli (Grieser & Kuhl, 1989; Kuhl, 1991; Miller & Volaitis, 1989; Volaitis & Miller, 1992). It has also been found that prototypic category members are processed faster (e.g., Pisoni & Tash, 1974) and more effectively (e.g., Repp, 1977) than non-prototypic members.

These findings of a graded category structure and more effective processing of prototypical stimuli, suggest that phonemic categories may be organized around prototypes, or best exemplars. Further evidence for this assertion has been demonstrated by Kuhl (1991), who found that stimuli close to prototypes in psychoacoustic space become difficult to discriminate from the prototype as compared to stimuli surrounding non-prototypic representations. The difference in discrimination has been proposed to be a result of prototypical category members attracting surrounding members and shrinking perceptual space. Kuhl coined this warping of perceptual space the “perceptual magnet effect” (PME).

Support for the Perceptual Magnet Effect

Initial evidence for the PME was provided by a study in which Kuhl (1991) synthesized a set of stimuli representing forms of the vowel /i/, as in the word *peep*, by varying the position of the first two formants (F1 and F2). A formant is a frequency band of relatively greater intensity in an acoustic signal. Formants are determined from the characteristics of the resonating body that pressure from the sound source is submitted to. The positions of the first two formants are a defining characteristic of the perceived category of a vowel sound (Handel, 1989). Kuhl's stimuli consisted of a prototype (P) and a nonprototype (NP) as defined by an earlier experiment (Grieser & Kuhl, 1989) in which participants rated the perceived goodness of a set of synthesized /i/ stimuli on a scale of 1 (poor exemplar) to 7 (good exemplar). The P and NP were determined by obtaining the average best and worst ratings, respectively, across participants. Kuhl synthesized variants around the P and NP (32 around each) along vectors of four circular orbits, such that stimuli along any vector were separated by 30-mel intervals across adjacent orbits. [The mel scale is a psychophysical scale that is linear at low frequencies and logarithmic at high frequencies (Fant, 1973)]. The frequencies of the vowel stimuli were converted from Hz to mels because the mel scale is assumed to reflect the pattern of responses on the basilar membrane (Fant, 1973), and thus allows for equal perceptual (pitch) separation between steady-state stimuli.

Kuhl (1991) presented the entire set of stimuli to a group of adult participants from which goodness ratings as members of the category were collected. She found that, similar to Grieser and Kuhl (1989), participants consistently rated certain stimuli as better

in goodness than others and that ratings increased symmetrically as stimuli approached the P. In a subsequent discrimination task Kuhl presented the P with each of its 32 variants, and the NP with its 32 variants, to a group of six-month-old infants, adults, and mature rhesus macaques (ranging from 1 to 3 years old). She found that, in infants and adults, discrimination was significantly poorer around the P than the NP. She concluded that the prototype acts as a perceptual magnet, attracting its surrounding stimuli and shrinking perceptual space. The magnet effect was not obtained with the macaques, suggesting that the effect is unique to humans. Kuhl interpreted the results of the goodness rating and discrimination experiments to mean that phonemic categories have a specific internal structure in long-term memory. Specifically, Kuhl proposed that phonemes are organized around category prototypes. Because the magnet effect was only observed with the human subjects, she also suggested that the formation of category representations may depend upon linguistic experience.

Though the PME relies on category prototypes, Kuhl (1991) does not suggest that the effect is necessarily a reflection of prototype theory. In fact, it has been proposed that exemplar theory, prototype theory or a combination of the two theories can account for the effect (Iverson & Kuhl, 1995; Kuhl, 1993). Exemplar theory (e.g., see Hintzman, 1986; Nosofsky, 1987; for a speech example, see Pisoni, 1993) suggests that we store all individual instances of a category in long-term memory. According to this approach, novel stimuli are categorized by comparing them with the individual exemplars from several categories. The novel stimulus should be identified as a member of the category whose exemplars, on average, provide the closest match. The more representative, or prototypic, a stimulus is, the more similar it will likely be to a large set of individual

exemplars. A cluster of prototypic exemplars would be close together perceptually, and would explain why discrimination performance is poor around the prototype in the PME. Because there are several shared properties with exemplars, a more representative stimulus also will be categorized and accessed faster (i.e., reflecting its reduced distance relative to the category's exemplars).

In contrast, prototype theory claims that we store a category as a single, idealized representation, or prototype, that reflects a mathematical average of all exemplars of that category (e.g., Posner & Keele, 1968; for a speech example, see Samuel, 1982). This theory maintains that incoming stimuli are compared to prototypes from various categories. Just as with exemplar theory, a stimulus will be identified as a member of the category whose prototype is the shortest distance away.

Research has indicated shortcomings in both exemplar and prototype theory as explanations for the nature of representation of perceptual events. Exemplar theory has been criticized because the number of exemplars that contribute to the representation of a single category would be overwhelming and inefficient for storage and recall (Meyer, 1970). Applications of prototype theory to speech perception have likewise been suggested to be inadequate because prototypic representations do not likely contain variable speech information, such as that from different speakers (Iverson & Kuhl, 2000). The inability of prototype theory to account for variability in speech is a problem because information about talkers is considered to be part of the representation (e.g., Bradlow, Nygaard, & Pisoni, 1999). Because both exemplar and prototype theory make almost identical mathematical predictions of categorization performance, Kuhl (1993) has suggested that the human perceptual system may utilize both systems, depending on

which would be more effective for determining category membership. Exemplar, prototype, or a combination of the two theories all are consistent with the PME because Kuhl makes no claim as to which approach is more appropriate and because all three notions indicate that the formation of representations is dependent upon experience.

Kuhl (1993) has proposed an explanation of the role experience has in language representation in her theory of language development, the native language magnet (NLM) theory. According to NLM theory, representations of phonemes develop early in infancy as a result of passive exposure to the native language. With continuous exposure to the native language, infants develop prototypical representations (not necessarily a single prototype) of phonemes specific to the language. The prototypes of a phonetic category act as perceptual magnets, attracting nearby stimuli. The result is a distortion of perceptual space, with stimuli close to the prototype being more similar than stimuli surrounding non-prototypical representations. As the prototype-centered categories develop, infants become more sensitive to the sounds of their native language and lose the ability to detect sounds from foreign languages. The main piece of evidence for NLM theory is finding that prototypes are less discriminable from neighboring stimuli than non-prototypes, i.e., the PME.

Several studies have obtained evidence of a PME (e.g., Aaltonen et al., 1997; Iverson & Kuhl, 1995, 2000; Kuhl, 1991; Sussman & Lauckner-Morano, 1995), and, consistent with NLM theory, the effect has been shown to rely upon experience with one's native language (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Polka & Werker, 1994). For example, Kuhl et al. (1992) tested six-month-old American and Swedish infants with the prototypic American English /i/ vowel and a prototypic Swedish /y/ vowel. They

found that both groups of infants showed worse discrimination around their native vowel and better discrimination around the foreign vowel. Kuhl et al. argued that both groups of infants demonstrated a magnet effect to their native vowel. Similar results have been found with American and German infants (Polka & Werker, 1994). Aaltonen et al. also demonstrated that the PME is dependent on linguistic experience by showing that native Finnish-speaking adults exhibited poorer discrimination performance among /i/ stimuli with low F2 values. Studies with American English speaking adults have indicated that stimuli from the same category with high F2 values are more difficult to discriminate (Iverson & Kuhl, 1995; Kuhl, 1991; Sussman & Lauckner-Morano, 1995). These findings have been suggested to indicate that the magnet effect is affected by early exposure to language and is native-language specific.

The PME also has been shown to generalize to consonants (Davis & Kuhl, 1994; Iverson & Kuhl, 1996). For example, Davis and Kuhl (1994) found significantly worse discrimination around empirically defined prototypes of the consonants /k/ and /g/ than around non-prototypes of each category. The same pattern of discrimination performance was found by Iverson and Kuhl (1996) with the consonants /r/ and /l/. Additional evidence for the magnet effect is provided by these studies because the effect is claimed to occur with speech sounds in general.

The PME has been further demonstrated with various psychophysical methods. For example, Iverson and Kuhl (1995) applied Signal Detection Theory analyses (e.g., see Macmillan & Creelman, 1991) to a single diagonal vector in acoustic space from Kuhl's (1991) stimuli. The vector consisted of 13 /i/ tokens, including the P and NP, along with variants to the left and right of each standard. Discrimination performance was measured

using d' , a theoretically bias-free measure of sensitivity. The d' statistic is calculated by subtracting the z-score transformation of the false alarm rate from the z-score transformation of the hit rate. False alarm rate is obtained by “different” responses when the stimuli are the same. Hit rate is obtained from “different” responses when the stimuli are actually different. Iverson and Kuhl also added an identification task to ensure that the vowels were falling within the same phonemic category. Their results were similar to those of Kuhl in that certain stimuli were consistently rated as better in perceived goodness than others, and discrimination from the P stimulus was significantly worse than discrimination of tokens that were psychoacoustically equidistant (along the mel scale) from the NP.

Iverson and Kuhl (1995) provided further support for the magnet effect using multidimensional scaling (MDS). MDS is a technique used for determining stimulus attributes that distinguish between members of a class of stimuli by providing a map of perceptual distances between items. The best-fitting map of perceptual distances between items was a three-dimensional matrix derived from participants' reaction times in a discrimination task with the 13 vowel stimuli. Shorter reaction times were an indication of perceptual similarity, and longer reaction times, an indication of dissimilarity. The MDS scaling solution revealed that stimuli around the P were perceptually closer, i.e. had shorter discrimination reaction times, than the stimuli reflecting the corresponding psychoacoustic distances (in mels) around the NP. This finding was interpreted to be consistent with the magnet effect in that poorer discrimination around the P was argued to reflect perceptual similarity of stimuli close to the prototype, and better discrimination around the NP to reflect salient differences in stimuli far from the P.

In addition to the behavioral evidence, support for the PME also has been provided by physiological studies. For example, Aaltonen et al. (1997) evaluated potential magnet effects using recordings of mismatch negativity (MMN), a preattentive ERP measure that is typically argued to reveal detection of a change from a standard sound in echoic memory (Ritter, Deacon, Gomes, Javitt, & Vaughan, 1995). The MMN consists of a negative peak that approaches 100-200 ms post-stimulus; the amplitude of the negative peak increases with greater deviations from the standard (Näätänen, 1992). Echoic memory, or the auditory sensory register, is the brief storage system for input hitting the auditory receptor cells. Echoic memory has been argued to consist of a fairly accurate acoustic trace with rapid decay occurring within approximately 2000 ms (Darwin, Turvey, & Crowder, 1972). The event-related potential (ERP) is a technique used to record small, brief fluctuations of electrical activity in the brain in response to a stimulus (Hillyard & Kutas, 1983). Because the subject does not have to actively attend to the stimuli for the measure to detect a change, the MMN allows a good measure of perceptual processes, such as discrimination.

Aaltonen et al. (1997) obtained MMN recordings from subjects performing a discrimination task in which the standard (P or NP) was either paired with itself or one of its surrounding stimuli. They found that MMN measures were lower for discriminations of stimuli from P than from corresponding NP discriminations, indicating that participants were less effective at detecting a difference between stimuli around the P. This finding was interpreted as providing more support for the PME. This effect was only found for subjects who were classified as “good” categorizers, that is, those Finnish speaking adults who were consistent in terms of rating a particular stimulus the same way

across trials, but not for the “poor” or inconsistent categorizers. Aaltonen et al. suggested that the difference in subjects may have been a result of the poor categorizers’ inability to accurately label and rate unnatural vowel stimuli. Their findings suggest that there may be important individual differences in the ability to identify and rate synthetic vowels, and only those who can do so reliably will demonstrate a PME.

Physiological support for the magnet effect also has been provided by the mapping of a neural model that fits well with the psychophysical data (Guenther & Gjaja, 1996). Using several computer simulations, Guenther and Gjaja derived a model that suggests the effect arises from experience-specific variability in the distribution of cell firing preferences. The model claims that the PME depends upon experience, but does not necessarily presume that the effect is specific to linguistic experience. The model makes the prediction that the magnet effect may occur with nonspeech stimuli. To test this prediction, Guenther, Husain, Cohen and Shinn-Cunningham (1999) trained participants to identify white noise stimuli in certain frequency ranges as belonging to the same perceptual category. They found that sensitivity among the stimuli decreased within the category after the categorization training. This finding is consistent with the PME for vowels in that sensitivity is also poorer around good representatives of a nonspeech category, albeit an artificial one.

Criticisms of the Perceptual Magnet Effect

Despite the behavioral and physiological support for the PME, it has been the target of several criticisms. One criticism of the magnet effect is that the results of studies reporting the effect cannot always be replicated. Using a paradigm similar to Iverson and

Kuhl (1995), several researchers have failed to find a magnet effect with the vowel /i/ (e.g., Lively & Pisoni, 1997). The physiological evidence for the magnet effect also has failed to be replicated. Rather than finding lower MMNs around the P (as in Aaltonen et al., 1997) the opposite pattern has been found (see Sharma & Dorman, 1998). The difficulty in replicating the magnet effect suggests that the effect may not exist as an independent perceptual phenomenon.

Another claim against the PME is that it may not be generalizable to other vowel categories. For example, Sussman and Gekas (1997) found no differences in discrimination performance between the P trials and the NP trials with the vowel /I/. Thyer, Hickson, and Dodd (2000) likewise failed to obtain evidence of a PME with any of five Australian English vowels. They found no differences in discrimination among the three vowels: /æ/, /a/, and /ɜ/, better discrimination around the NP than around the P with the vowel /ɔ/, and only partial support for poorer performance around the P with the vowel /i/. Thyer et al. argued that the PME may not be a reliable phenomenon.

The biggest criticism of the PME is that the stimuli used may reflect more than one phonemic category. For example, in Iverson & Kuhl's (1995) study Kuhl's (1991) NP was only identified to be within the /i/ category 55% of the time. Sussman and Lauckner-Morano (1995) found that Kuhl's NP was only identified to be within the /i/ category 8% of the time. Other researchers have found that Kuhl's NP is sometimes labeled to be within the /e/ category more often than it is labeled to be within the /i/ category (e.g., Lively & Pisoni, 1997). As a result, better discrimination around the NP rather than the P stimulus could be argued to be due to between-category differences, i.e., inter-categorical

perceptual tendencies (Lively & Pisoni, 1997; Lotto et al., 1998; Sussman & Lauckner-Morano, 1995).

Several researchers have claimed that the results of experiments reporting a PME can be explained in the framework of inter-categorical perception (e.g., Lively & Pisoni, 1997; Lotto et al., 1998; Sussman & Lauckner-Morano, 1995). The inter-categorical perception explanation claims that poorer discrimination around the P is a result of decreased sensitivity to within-category differences, and better discrimination around the NP is a result of either the NP or its surrounding stimuli being at, or across, the phonemic boundary.

One source of evidence for the claim that the PME is essentially inter-categorical perception is that the magnet effect has been shown to be influenced by context (e.g., Lotto et al., 1998). Context can affect perception of vowel stimuli to such an extent that it can change the perceived category (Thompson & Hollien, 1970). Contextual influences can be problematic for the claim of a PME because the perceived category of the stimuli could vary from /i/ to /e/ in specific contexts, causing differences in discrimination that may be due to between-category judgments.

To examine contextual influences on the PME, Lotto et al. (1998) replicated Iverson & Kuhl (1995) using pairs of stimuli in the identification, goodness rating and discrimination tasks. Previous research on the PME had presented isolated stimuli in the identification/goodness rating task, but pairs of stimuli in discrimination tasks (e.g., Iverson & Kuhl, 1995; Kuhl, 1991). Lotto et al. suggested that contextual influences from using an additional stimulus within a trial may have been present in the discrimination tasks. As a result, the difference in the number of stimuli in a trial between the tasks may

have confounded the results of experiments reporting a PME. Lotto et al. presented stimulus pairs in each task that consisted of the standard (P or NP) paired with itself or with one of its surrounding stimuli. The results confirmed that context did play an important role in identification and goodness ratings. For example, certain stimuli were reliably perceived as /i/ when presented with the NP but as /e/ when presented with the P. A subsequent discrimination task indicated a PME in that discrimination was significantly worse around the P than the NP. However, when controlling for any inter-categorical discriminations by comparing observed differences with predicted differences in generalization scores, their results revealed the opposite of what the PME predicts, i.e. better discrimination around the P than the NP. Lotto et al. concluded that differences in discrimination performance in studies reporting a PME may be due to between-category judgments, thus reflecting inter-categorical perception rather than a PME.

Further support for the inter-categorical perception explanation comes from studies that have found individual differences in the PME task (e.g., Lively & Pisoni, 1997). In the work of Kuhl and her colleagues (Iverson & Kuhl, 1995, 2000; Kuhl, 1991; Kuhl et al., 1992) the P and NP are obtained from average ratings across listeners. However, several studies have found that when the P and NP are individually normed, there is some variability in goodness ratings and in perceived categories. The logic behind norming P's and NP's is that individuals have different linguistic experience, so the stimuli should be selected to account for this by using each individual's best and worst ratings within the category. Studies with individually selected P's and NP's have failed to find a PME (e.g., Frieda, Walley, Flege, & Sloane, 1999; Lively, 1993; Lively & Pisoni, 1997). By demonstrating individual differences in identification and goodness ratings, the argument

can be made that the standard set of stimuli used by Kuhl (1991) may be bordering or crossing over to other phonemic categories for some listeners, introducing inter-categorical comparisons. This alternative interpretation of the PME is even consistent with more long-standing findings of phoneme perception based on category boundaries, such as that observed in studies of categorical perception, rather than prototypes or graded within-category structure.

Categorical perception refers to perceiving items from a physical continuum as belonging to discrete categories (e.g., see Studdert-Kennedy, Liberman, Harris, & Cooper, 1970). In a classic categorical perception paradigm, four criteria are needed to demonstrate that phonemic categories are perceived in an all-or-none fashion (Studdert-Kennedy et al., 1970). First, a labeling task is completed in which subjects are required to label a stimulus with its corresponding phonemic category, e.g., /ba/. The labeling task should yield a psychometric function with a steep slope at a category boundary, indicating that each stimulus is typically clearly identified to be a member of a particular phoneme category, not as an ambiguous exemplar. Second, equidistant pairs of stimuli are presented in a discrimination task in which subjects are required to judge whether the two stimuli are the same or different. The discrimination task should result in troughs of chance discrimination for stimulus pairs within a single category. Peaks of discrimination for stimuli straddling the category boundary should also be evident. The peaks in performance indicate that differences among stimuli at category boundaries are more salient than those within the category. Finally, performance along one task (labeling or discrimination) should be predictable from the other task.

Though several researchers have claimed that the PME may be a demonstration of inter-categorical perception (e.g., Lively & Pisoni, 1997; Lotto et al., 1998), there are several problems with this claim. For example, studies with individually normed P's and NP's (Frieda et al., 1999; Lively, 1993; Lively & Pisoni, 1997) do not offer compelling evidence against the magnet effect because there are problems associated with utilizing normed standards. One problem with the method of normed prototypes is the assumption of a normally distributed category around an individual's prototype. The stimuli created to surround an individual's prototype may not be representative of one phonemic category. If there are individual differences in selection of prototypes, it also is likely that there are individual differences in the distribution of phonemic representations. An additional problem with this method is that individualized prototypes may not be reliable from the first goodness rating task to the second identical task. For example, in Lively and Pisoni's (1997) study the individually selected prototypes shifted a mean of 80 mels from the first to the second testing in the same individual listeners. Given the variability in the selection of prototypes, it is difficult to determine if discrimination performance was based on comparisons with an appropriate standard.

Another problem associated with the inter-categorical perception explanation of the PME is that, categorical perception of vowels is difficult to demonstrate (e.g., Fry, Abramson, Eimas, & Liberman, 1962). Certain manipulations of stimuli not required to elicit categorical perception of consonants, such as degrading the signal (see Lane, 1965) and adding the context of another vowel (e.g., Fujisaki & Kawashima, 1969), have been found to result in small peaks in discrimination with vowels. But, there is still a lack of discrete perceptual category boundaries with vowels. Rather than categorical perception,

vowels typically reveal a continuous labeling function. This is problematic because the evaluation of inter-categorical effects may be complicated somewhat by the reliance on stimuli that do not indicate a sharp perceptual boundary between phonemic categories. Even though the lack of a distinct boundary might make an analysis of inter-categorical effects a little more difficult, failure to find categorical perception would not eliminate inter-categorical concerns. It would still be plausible that some stimulus comparisons in discrimination may reflect two different categories (or even that both stimuli may be out of the target category). Such comparisons would be especially likely around the NP, where stimuli are mostly poor representatives, thereby still resulting in slightly enhanced discrimination performance in NP, relative to P, context.

A final, fundamental problem with the inter-categorical perception interpretation of the PME is that it has been offered without directly testing it. In other words, no one has yet conducted a categorical perception task on the vowel stimuli used in PME experiments to assess the extent of any potential inter-categorical interference.

The Perceptual Magnet Effect and Musical Stimuli

To further explore the PME and its counter-arguments, it would be beneficial to determine if the effect occurs with nonspeech stimuli. Several implications could be made if a magnet effect can generalize beyond phonemes. First, it could be suggested that the PME is not specific to speech, consistent with Guenther and Gjaja's (1996) neural model which predicts that the magnet effect is not exclusive to speech. Second, it would be informative in terms of how we categorize and represent nonspeech sounds. Finding a magnet effect with stimuli other than vowels and consonants also would provide further

support for the effect and its generalizability. Assuming the speech PME evidence is valid, the failure of a magnet effect to be demonstrated with nonspeech stimuli would suggest that different mechanisms are involved in how we categorize speech and nonspeech sounds. It then would be reasonable to assert that the PME could reflect the operation of a phonetic module, a biologically specialized mechanism for the processing of speech (see Eimas, 1963). Musical sounds represent a class of sounds with which most people have had varying experience. Thus, musical stimuli would make a good nonspeech comparison for the speech results. Musical results could be capable of addressing potential differences in categorization between speech and nonspeech sounds.

A few studies have attempted to find a PME with musical stimuli. The first was conducted by Acker, Pastore and Hall (1995). Their stimuli were variants of C-major triads. The different variants of the chord were created by altering the frequencies of the E and G chroma (i.e. mistuning the chord). Rather than using the mel scale, they created their stimuli by varying the frequencies of the E and G chroma in 2 Hz intervals. Acker et al. presented the stimuli to a group of trained musicians who were required to rate the perceived goodness of pairs of stimuli. Consistent with Kuhl (1991), their results showed that participants consistently rated certain chords as better than others. The authors suggested that musical stimuli may be stored and represented in terms of prototypes or best exemplars.

Acker et al. (1995) then administered a discrimination task. Participants were presented with an individualized P, as obtained from their average highest rated chord, along with its surrounding stimuli. Participants also received a standard NP and its surrounding stimuli. The results of the discrimination task opposed predictions from the

PME, i.e., discrimination was better around the P and worse around the NP for corresponding differences in frequency (Hz). The authors concluded that the P acted as a perceptual anchor rather than a magnet. A perceptual anchor is a stimulus that is used as a reference point for surrounding stimuli. In a continuum of stimuli, an anchor is typically an end-point stimulus that is especially salient. Because the anchor is more distinct than the other stimuli, listeners may use the anchor as a standard to judge other stimuli. The result is better discrimination around the anchor, or P, as compared to the NP. Acker et al. concluded that the PME may not be generalizable to musical stimuli.

One potential problem with Acker et al. (1995) study was that their stimuli were not analogous to the stimuli that have been used in the speech work (e.g., Aaltonen et al., 1997; Iverson & Kuhl, 1995, 2000; Kuhl, 1991; Sussman & Lauckner-Morano, 1995). Acker et al. had equal changes in stimuli in physical terms (Hz) rather than equal changes in perceptual (mel/pitch) terms. Another problem with the Acker et al. design is that each participant received a customized P that was determined from their individual goodness ratings. Research on the PME using speech stimuli (Lively & Pisoni, 1997) has indicated that individually selected prototypes may not be stable from one test of goodness ratings to the second. Using unreliable normed stimuli may result in an assessment of the PME based on inappropriate standards. Also, Acker et al. did not address the distribution of stimuli surrounding the normed prototypes in their study, i.e., no mean data were reported indicating the nature of the distribution of within-category responses for each participant.

It has been suggested that the lack of a PME in Acker et al. (1995) study could be due to the fact that they used musicians as participants (Barrett, 1998). Musicians are well trained at discriminating in-tune from out-of-tune chords, which could result in better

discrimination around the P. Barrett (1998) adopted a similar procedure to that of Acker et al. to compare musicians with non-musicians. Her stimuli consisted of a continuum of chords that went from C-major to C-minor separated by 1.5 Hz intervals. Rather than obtaining perceived goodness ratings on a Likert scale, Barrett presented the stimuli to a group of musicians and had them identify the stimuli as C-major or C-minor. From this task she obtained individualized NP's for the musicians as those stimuli 5 Hz from their 50% crossover point, i.e. the point at which musicians classified a chord as major or minor with equal probability. Barrett's P stimulus was an equally tempered C-major chord, and was kept standard for all participants.

Barrett (1998) then gave a discrimination task to a group of musicians and non-musicians. The musicians received their individualized NP's, but the non-musicians all received the same NP that was an average of those identified by musicians. The results revealed that the non-musicians demonstrated poorer discrimination around the P than the NP, indicating a PME. There were no significant differences in discrimination around the P and NP for the musicians. However, when dividing the musicians into two groups, Barrett found that 6 out of 10 participants showed significantly better discrimination around the P than NP, which was interpreted as an anchor effect. The remaining 4 musicians demonstrated significantly poorer discrimination around the P than the NP, which was suggested to indicate a PME. Examination of the musical background of the two groups of musicians indicated that the 4 listeners who demonstrated a magnet effect had fewer years of musical training than the 6 listeners who demonstrated an anchor effect. Barrett concluded that the PME can occur with musical stimuli, but only with those participants who do not have a substantial amount of musical experience. Barrett

suggested that listeners with an extensive musical background are trained to devote more attention to well-tuned stimuli, resulting in better discrimination around prototypical musical chords. In contrast, listeners with little to no background in music are not adept at detecting well-tuned musical chords, leading to poorer discrimination around prototypes. This interpretation is a problem for standard treatment of a PME, given that it is argued to be the result of experience in the standard interpretation of the speech work (Kuhl, 1993).

Schellenberg (2002) also attempted to replicate Acker et al. (1995) to determine if a PME can occur with musical stimuli in musicians and non-musicians. The prototype was a tuned chord, and variants were created by mistuning the chord by varying amounts of cents. A cent corresponds to a relatively small frequency ratio. In music, differences in pitch are given in units called semitones. There are 12 semitones, representing equal log steps, within one octave (a doubling of frequency). Each semitone can be further divided into 100 smaller equal log intervals called cents (Sundberg, 1991). Musicians and non-musicians completed discrimination tasks with sequential and simultaneous musical intervals. The results did not reveal any significant differences in discrimination performance between the P and NP conditions, i.e., no PME or anchor effect, in either group of participants for either task. Schellenberg suggested that the magnet effect may be specific to speech. He also proposed that the magnet effect could be the result of design flaws in the speech research because of potential between-category comparisons around the NP and within-category comparisons around the P. Thus, the PME that has been reported (Iverson & Kuhl, 1995, 2000; Kuhl, 1991) may actually be a reflection of inter-categorical comparisons.

Taken together, this cohort of musical perceptual magnet paradigms does not offer a clear answer as to whether the effect can be obtained with musical stimuli. One fundamental reason for the discrepancy among the musical studies, and between the music and speech experiments, could be that musical chord stimuli may not be the most appropriate analogue to vowels. There are substantial physical differences between steady-state vowels and musical chords. Chords in the musical studies are generally made by mixing a few sinusoids, i.e., isolated frequencies, at the targeted chroma, whereas vowels are complex sounds made from a harmonic series. In vowels, the identifying information is conveyed by formants, not isolated frequencies (although sinusoidal versions of words based on the center frequencies of the formants can sometimes be sufficient for identification; (e.g., see Remez, Pardo, Piorkowski, & Rubin, 2001). This difference in stimuli between the speech and musical studies could have caused the different outcomes. Rather than making judgments of stimulus quality in the musical studies, participants could have been making judgments based on consonant or dissonant musical chords. In fact, some evidence has been obtained that suggests discrimination is better among consonant chords than dissonant ones (Iverson, Hall, & Kuhl, 1997).

A more analogous method of testing the PME with musical stimuli may be to use steady-state timbres, with clearly defined formant values that can be manipulated on a mel scale. For example, one timbre that has been proposed to particularly resemble the human voice is the violin. Askenfelt (1991) has suggested several similarities between violin tones and the human voice. First, both types of sounds reflect relatively distinct contributions from the sound source and filter. The source is the initial body put into vibration (larynx or bow string). The filter is the resonating body that the source energy is

being submitted to (vocal tract or violin body). There is an initial spectral envelope associated with the source which will be modified by the resonant properties of the filter. The result is several spectral peaks, or formants, in the resultant waveform. In both the voice and the violin the source of vibration and the filter are loosely coupled, i.e. the larynx or string can vibrate fairly independently of the vocal tract or violin body, and thus, can make relatively unique contributions to the resultant spectrum. Another similarity of the voice and violin is that the source spectrum of both consists of a -6 dB to a -12 dB spectral slope. Additionally, the voice and the violin are both fairly accurate in pitch control; both can continually manipulate fundamental frequency. Given the close relationship between the voice and the violin, a steady-state violin tone seems to be an appropriate analogue to the vowel stimuli for exploring the PME with musical stimuli.

Purpose of the Study

The purpose of the present study was to determine if a PME occurs in musical stimuli that are more analogous to vowels than the nonspeech stimuli that have previously been used. Specifically, a violin tone was used because of its several shared properties with the human voice (Askenfelt, 1991). Another reason for using musical timbres was the possibility that there may not be as many problems with crossing category boundaries by synthesizing a set and varying only formant center frequencies. This is because there are several other dimensions that contribute to the identification of timbre, such as attack-related characteristics (e.g., see Grey & Moorer, 1977). This study also attempted to evaluate any potential inter-categorical effects in a PME paradigm. Inter-categorical effects were in part examined in a regression analysis that compared the effects of

goodness and identification of the stimuli on discrimination performance. In addition, a categorical perception task was included to determine if the stimuli crossed or approached any other timbre categories. Conducting a categorical perception task in the context of a PME paradigm reflects a novel attempt to directly evaluate the nature of potential inter-category influences. The potential effects of context with the timbre stimuli also were examined.

Three experiments were conducted to explore the PME with timbre stimuli. Experiment 1 provided a determination of whether the timbre stimuli represented a broad category with differences in perceived goodness to permit an adequate assessment of a potential PME with musical stimuli. Experiment 1 also established the stimuli to be used for later experiments. To accomplish these goals, a set of 45 synthesized violin tones (the center tone and 44 surrounding variants) varying in the values of F1 and F2 in 30-mel intervals were presented for identification and goodness rating without context. Consistent with the speech (e.g., Iverson & Kuhl, 1995) and nonspeech (e.g., Acker et al., 1995) research on the magnet effect, it was hypothesized that the stimuli would be within the intended category when at least 120 mels from the center stimulus and that there would be salient difference in goodness among the stimuli.

The primary purpose of Experiment 2 was to determine if a PME occurs with the timbre stimuli. As part of this determination, the role, if any, of potential inter-categorical and context effects were examined. In Experiment 2, stimuli were restricted to a single diagonal vector of F1-F2 space consisting of 14 of the timbre tokens (including the P and the NP). Subjects performed the identification and goodness rating tasks with and without context, and then completed a subsequent discrimination task in the P and NP context.

Based on the assumption that the speech claims of a magnet effect are valid (e.g., Iverson & Kuhl, 1995), it was predicted that discrimination performance based on a single category would be better around the NP than the P (revealing a PME). Based on the finding that identification contributed more to discrimination than goodness in Lotto et al. (1998) study, it also was predicted that inter-categorical interference would be obtained in this experiment. Given that context has been shown to be very influential in speech (Lotto, Kluender, & Holt, 1998) and musical stimuli (Acker et al., 1995; Donnadieu, McAdams, & Winsberg, 1994), it also was expected to impact timbre stimuli.

A categorical perception task was conducted in Experiment 3 to examine the nature of any potential inter-categorical effects. The categorical perception task allowed a determination of whether multiple categories were represented in the stimulus set used to assess the PME in Experiment 2. Listeners labeled each of the 14 tones from the stimuli used in Experiment 2 and then completed a discrimination task on equidistant pairs of stimuli along the vector. Categorical perception of the timbres would indicate that multiple categories were being represented in the stimulus set. While continuous perception would not indicate distinct, separate categories, any continuous shifts in the labeling or discrimination functions of listeners could suggest that other categories were interfering in the stimulus set. Because categorical perception is difficult to demonstrate with steady-state stimuli (e.g., Fry et al., 1962) it was not expected to occur with the timbres.

CHAPTER 3

EXPERIMENT 1

The purpose of Experiment 1 was to determine whether the timbre stimuli could be synthesized to represent a broad category with qualitative differences in perceived goodness. A secondary purpose was to identify a subset of stimuli, including both a P and an NP stimulus, to be used in Experiment 2. Identification and goodness ratings were obtained for 45 synthesized violin tones presented individually. Consistent with previous findings for vowels that goodness ratings are graded around prototypes (e.g., see Grieser and Kuhl, 1989; also see Kuhl, 1991), it was hypothesized that listeners would consistently rate certain stimuli as better exemplars of the synthetic string timbre than other stimuli, and that the ratings would be higher around the tone that was synthesized with formant values and bandwidths corresponding to a naturally produced violin tone.

Method

Participants

In this and all subsequent experiments, permission to conduct research with human subjects was approved prior to experimentation by the UNLV Social and Behavioral Sciences Institutional Review Committee on June 10, 2004. Sixteen listeners with no reported hearing deficits from the University of Nevada, Las Vegas completed this experiment. Their ages ranged from 21 to 42 years (mean age = 26.75 years). Five

listeners were psychology graduate students, and 11 were personnel from our laboratory. Listeners from the laboratory knew little to no information about the experiment. Any pre-existing knowledge laboratory listeners did have of the research area in this and in subsequent experiments was not a concern because background knowledge of the PME should not affect low-level perceptual responses to the stimuli. An additional reason any pre-existing knowledge was not of concern is that laboratory participants' data looked much like that from the graduate students. Data from one participant, a graduate student, were excluded from all analyses because of an error in the setting of the filter. Therefore, the reported results are based on the remaining sample of fifteen participants.

Given that musical history has been shown to affect responses in previous tests for a musical PME with chords (see Acker et al., 1995; Barrett, 1998; and Schellenberg, 2002), the musical background of each participant was assessed in this experiment, as well as in all subsequently reported experiments. Each listener completed a musical history questionnaire that measured years and type of musical training (if any). Items were included in the survey that evaluated whether listeners were professionally taught and/or self-taught, the number and types of instruments they played, as well as their knowledge of music theory (for a copy of this survey, see Appendix II). Listener's musical experience ranged from 0 to 27 years (mean = 6.1 years).

Stimuli

Synthesis parameters for all stimuli were derived from acoustic measurements for a natural token that was produced using an actual violin. This initial violin tone was digitized at a sampling rate of 44.1 kHz from the series of bowed (with vibrato) tones within the McGill University Master Samples (MUMS) library (Opolko & Wapnick,

1987). The violin tone was taken from the pitch A₄ performed on an open string. The open string form of the tone was preferred because failing to depress the strings on the violin body removed vibrato. The vibrato technique, such as that produced by stopped tones in the series of recordings, would have compromised the desired steady-state nature of the tone by adding momentary frequency fluctuations across harmonics. As a result, the generalizability of the timbre stimuli to the speech research with steady-state vowels would be limited.

The tone was truncated to 400 ms from onset. The motivation for using a relatively short stimulus duration was to keep the stimuli analogous to the corresponding speech research, where vowel stimuli typically range from 400 to 500 ms. Another purpose for using a 400 ms duration was to reduce memory concerns in the discrimination tasks of Experiments 2 and 3, where two 400 ms stimuli were presented with a 250 ms interstimulus interval (ISI). The total time to complete stimuli on a trial in either discrimination task (1050 ms) did not exceed general estimates of the span of echoic memory, which are approximately 2000 ms (Darwin et al., 1972). By keeping trial length within the limits of echoic memory, it can be reasonably ensured that the sensory trace of one stimulus did not substantially decay before presentation of the second stimulus had ended (Macmillan, Braida, & Goldberg, 1987). The amplitude of the truncated tone was ramped off linearly over the final 10 ms. This linear decrease in amplitude imposed an admittedly artificial decay. However, the signal length was sufficient to preserve the full attack, which has been shown to be critical to instrument discrimination and identification (e.g., see Grey & Moorer, 1977).

To provide data for re-synthesis, center frequencies of formant values, along with corresponding bandwidths, were determined by analyzing the steady-state portion of the tone using Computerized Speech Research Environment (CSRE) software. Only the steady-state portion of the tone was analyzed to keep the stimuli analogous to the steady-state vowels previously used to study the PME. The average center frequencies of the first four formants were 2239, 3703, 6804 and 8699 Hz, respectively; their corresponding bandwidths were 386, 244, 306 and 436 Hz.

The formant values of all stimuli were derived according to their relation to the center frequencies of formants measured for this natural stimulus. Once average center frequencies of the formants and bandwidths were established, the values of F1 and F2 were converted to mel scale values using the formula: $m = (1000/\log 2)\log(1+f/1000)$ (Fant, 1973), where m is the value in mels, and f is the frequency value in Hz. From those values, the mel values of F1 and F2 for 44 stimuli surrounding the center stimulus (i.e., the natural tone's formant values) were calculated.

Similar to Kuhl (1991), the stimuli were distributed in psychoacoustic space in (seven) orbits. Figure 1 represents the distribution of the timbre stimuli on the mel scale and the corresponding F1 and F2 values in Hz. As can be seen in Figure 1, the orbits were separated by 30 mel intervals. That is, orbit 1 consisted of stimuli 30 mels away from the original tone, orbit 2 consisted of stimuli 60 mels away from the center, etc. The first four orbits consisted of eight stimuli each, reflecting (positive or negative) changes from the center stimulus in terms of F1, F2, or both formants.

The three last three orbits consisted of fewer stimuli than other orbits and represented more distant comparison stimuli. These orbits consisted of four stimuli each, one at the

end of each F1-F2 diagonal (see Figure 1). The more distant stimuli were created to permit a determination of how far stimuli may be from the center stimulus before they are classified as out-of-category. Previous speech work (e.g., Iverson & Kuhl, 1995) has indicated that vowel stimuli begin to reflect the /e/ category around a distance of 90 to 120 mels from the center, or prototypic region, of the intended /i/ stimulus set. It is possible that these results may not apply equally to nonspeech timbre stimuli. In fact, it was hoped that the timbres would instead constitute a more broadly defined category to enable a stronger evaluation of the PME (i.e., by reducing the likelihood of inter-categorical stimulus comparisons). For this reason, diagonal vectors of synthesized timbre stimuli were extended to 210 mels from the center stimulus.

To be consistent with the perceptual magnet literature for vowels, the center frequencies of other identified formants for the violin tone, e.g. F3 and F4, were held constant for all stimuli (Kuhl, 1991). For subsequent synthesis purposes, all F1 and F2 center frequencies of the 45 stimuli were converted back to Hz using an inverse of Fant's (1973) formula: $1000 * (\text{Invlog}(m/(1000/\log 2))) - 1$.

Each steady-state stimulus was generated using Praat (version 4.2) speech synthesis software (Boersma & Weenink, 1992). This speech synthesis program kept stimulus creation consistent with speech work by relying on Klatt (1980) parameters. Praat additionally permits the avoidance of articulatory parameters in synthesis that are potentially disadvantageous to musical synthesis. The stimuli were constructed using an algorithm for source-filter synthesis. The source was generated by specifying all harmonics of the A₄ (440 Hz) fundamental frequency throughout the signal assuming a -12dB/octave spectral slope (Sundberg, 1991). The filter for any given stimulus was

generated by specifying the center frequencies of the first four formants and their corresponding bandwidths. The center frequencies of F1 and F2 varied depending on the particular stimulus. See Figure 1 for the F1 and F2 values for each of the stimuli in Hz. The source and filter were combined to obtain the final, filtered spectral envelope and synthesized at a sampling rate of 44.1 kHz (16-bit).

To retain the general, time-varying characteristics of the original violin tone in terms of peak amplitude, a linear attack function was imposed to all 45 synthesized stimuli using Syntrillium's Cool Edit Pro (version 2.1). The function was modeled after the original violin tone from MUMS and consisted of a 312 ms attack plus a 78 ms steady-state. A 10 ms linear off-ramp also was imposed to avoid abrupt, audible changes in amplitude at stimulus offset. All stimuli were matched for root-mean square (RMS) amplitude to roughly equate loudness across tokens.

Throughout all experiments, presentation of stimuli was submitted to a low-pass Butterworth filter with a cut-off frequency of 10 kHz and a -24 dB/octave slope. This cut-off frequency was selected to be optimal both for the elimination of unwanted noise and for anti-aliasing purposes (i.e., to smooth the synthesized waveform on output so that it more closely approximates the original, analog signal). Stimuli were presented at a peak intensity of 80 dB[A] over Sennheiser HD 25 SP headphones in a single-walled sound attenuated chamber.

Procedure

First, a familiarization period was administered to give listeners exposure to the range of stimuli that were presented in the experimental task. Each of the 45 timbre stimuli was presented once in random order. Familiarization was followed by 4 blocks of 225

experimental trials each (900 trials total). In each block of trials listeners received 5 randomized repetitions of the 45 stimuli (20 repetitions total across all blocks). A single tone was presented on each trial. In this experiment, as well as in subsequent experiments, a 500 ms inter-trial interval (ITI) was imposed following each response to prevent listeners from automatically beginning the next trial upon pressing a response button.

The listeners' task was to rate on a numeric scale each tone with respect to its perceived goodness as an example of a synthesizer's violin sound (i.e., timbre). Listeners were instructed to indicate their rating by pressing any key from 1 to 7 on the keyboard. A rating of 1 indicated a very bad example of the violin timbre, and a rating of 7 indicated a very good example (as in Kuhl, 1991). Additionally, listeners were told that they might occasionally hear a tone that sounds like the synthesizer simulated an instrument other than a violin. If this occurred on a trial, the instructions were to press 0 on the keyboard to indicate that the timbre was from some "other" instrument category. This method of response collection allowed a single response to provide both goodness and identification information. The trials were self-paced, and rest breaks were provided after each block. The entire session lasted approximately 60 minutes. As in subsequent experiments, presentation of stimuli and collection of responses were handled by a PC running Kendall's (2001) Music Experiment Development System (MEDS-2001-A).

Results and Discussion

Mean percentages of within-category identification responses and goodness ratings were calculated for all 45 stimuli. Figure 2 summarizes these two measures for each

timbre stimulus. Similar to Iverson and Kuhl (1995), both the P and NP were selected from the diagonal vector in which F1 and F2 shared an inverse relationship (see the shaded vector in Figure 2). An additional reason for selecting this vector was that these stimuli had more salient differences in goodness (range = 2.6) than the diagonal vector in which F1 and F2 shared a positive linear relationship (range = 0.7). To allow for an equal number of comparisons as a function of distance from the P and NP, one token from the end of the vector was excluded from the stimulus set. The excluded stimulus was chosen from the end of the diagonal vector that constituted the highest F2 and lowest F1 center frequencies. The shaded portion of Figure 2 represents the remaining 14 stimuli. The 14 stimuli allowed six comparisons for each determined standard (P and NP) for the following analyses and for the discrimination task of Experiment 2. Each standard had three comparison stimuli immediately to the left and three comparison stimuli immediately to the right along the vector. Stimuli to the left of the P and NP were assigned a negative distance, and stimuli to the right, a positive distance. As can be seen in Figure 2, the comparison stimuli consisted of tokens ± 30 , ± 60 and ± 90 mels away from the P and NP. The resultant stimulus set also allowed for a P and NP that were identified to be within the violin category at or near ceiling levels. As can be seen in Figure 2, the P received 100 percent violin category identifications and the NP received 97 percent within-category identifications.

Identification

An examination of the mean percentages of within-category identification responses in Figure 2 reveals that, as hypothesized, the stimuli as a set were almost always identified to be members of the violin category. In fact, the P was always judged to be in

the violin category. The NP in this experiment also received substantially more within-category judgments than the NP in Iverson and Kuhl's (1995) study with vowel stimuli (97 percent v. 55 percent, respectively). Each stimulus falling to the right of the NP in this experiment likewise received a higher percentage of within-category judgments (greater than 68 percent) than in the corresponding vowel stimuli used by Iverson and Kuhl (1995; less than 50 percent).

A 2 x 7 repeated measures ANOVA was conducted to examine the effects of Condition (P and NP) and Distance (0, ± 30 , ± 60 , and ± 90) on the obtained percentages of within-category identification responses along the diagonal vector. Musical History (based on the total years of musical experience) was included in this analysis as a between-subjects factor.

The analysis indicated that listeners with an extensive musical background assigned lower within-category identification responses than listeners with a moderate, or no, amount of musical training. This tendency resulted in a significant effect of Musical History, $F(7,7) = 5.259$, $p < .05$. Planned orthogonal comparisons of means demonstrated that listeners with 18 or more years of musical experience (mean identification rate of all stimuli was 81.43 percent) gave lower within-category identifications to stimuli than listeners with 8 or fewer years (mean identification of all stimuli was 96.51 percent), $p < .05$. This finding could reflect that the timbre stimuli may not have represented a natural category for listeners who have had a great deal of experience with musical tones.

Consistent with the speech results reported by Iverson and Kuhl (1995), percentages of within-category identification responses were more frequent among stimuli in the P

condition than among stimuli in the NP condition (means = 98.67 and 90.33, respectively). This was indicated by a significant effect of Condition, $F(1, 14) = 13.512$, $p < .01$, as well as by planned orthogonal comparisons of mean response tendencies between stimuli constituting the P and NP conditions, $p < .01$.

Percentages of within-category identification responses also differed as a function of distance from the P or NP stimulus. This tendency was revealed by a significant effect of Distance, $F(6, 84) = 14.209$, $p < .01$. The pattern of within-category responses paralleled that of speech (e.g., Iverson & Kuhl, 1995). Specifically, within-category responses decreased symmetrically along the diagonal vector as stimuli approached the NP. Planned orthogonal comparisons indicated that the standard (i.e., the P or NP) and the three stimuli to the left along the vector received significantly more within-category responses than the three stimuli to the right, $p < .05$. Also, the first two stimuli to the right of the standard, +30 and +90, both received significantly more violin identification responses than +90, $p < .05$. The P and comparison stimuli to its left along the vector all received more frequent within-category judgments than stimuli approaching the NP. This was demonstrated by a Condition x Distance interaction, $F(6, 84) = 11.17$, $p < .01$. Post-hoc Tukey test additionally indicated that the stimuli +30, +60 and +90 mels away from the NP received significantly fewer within-category identifications than any other stimulus along the vector, $p < .05$.

Goodness Ratings

Calculations of mean goodness ratings were based solely upon rating responses that reflected within-category judgments (i.e., responses of 1 to 7). Any responses that failed

to meet this criterion were only included in the analysis of results for the identification task.

The stimulus with the average highest goodness rating (4.4) along the diagonal vector established the P. As can be seen in Figure 2, the P was not the center stimulus produced using measurements from the natural tone. The P in the present study was located more to the left along the vector with a lower F1 and higher F2 value than the center tone. The stimulus with the lowest goodness rating (3.1), and that was still identified to be within the violin category near ceiling levels, established the NP. As shown in Figure 2, the tone designated as NP consisted of higher F1 and lower F2 values than the center stimulus.

Once the P and NP were established, a 2 x 7 repeated measures ANOVA was conducted on the 14 stimuli along the vector. The 14 stimuli consisted of the P and six comparison stimuli and the NP along with a corresponding set of six comparison stimuli. The ANOVA examined the effects of Condition (P and NP) and Distance (0, ± 30 , ± 60 , and ± 90 mels) on goodness ratings of the 14 stimuli along the diagonal vector. Musical History (based on the total years of musical experience) was included in the analysis as a between-subjects factor. Since there was no significant effect of Musical History, $F < 1$, data were collapsed across that variable for further analyses.

Consistent with the speech research on the PME (e.g., Kuhl, 1991), the timbre stimuli differed in perceived goodness. This was reflected by a significant effect of Condition, $F(1, 14) = 36.517, p < .01$. The nature of the differences in goodness followed the same pattern as those obtained in speech; planned orthogonal comparisons revealed that stimuli around the P were rated higher in perceived goodness than stimuli around the NP, $p < .01$.

Also, consistent with the findings for vowels (e.g., Kuhl, 1991), the present results indicated that timbre stimuli increased in goodness as they approached the prototype of the category. This trend contributed to a significant effect of Distance, $F(6, 84) = 32.385$, $p < .01$. Planned orthogonal comparisons indicated that the distance of 0 mels, i.e., the P or NP, received higher goodness ratings than the three stimuli along the vector to the right, +30, +60, and +90, $p < .01$. The three stimuli to the left of the P or NP (-30, -60, and -90) received higher goodness ratings than the three stimuli to the right (+30, +60, and +90), $p < .01$. Also, stimuli at a distance of +30 mels from the P or NP received higher goodness ratings than stimuli +60 and +90 mels away, $p < .05$, and stimuli at +60 mels received higher goodness ratings than +90, $p < .05$.

More evidence that the within-category timbre stimuli were qualitatively graded, as has been claimed for vowels (e.g., Kuhl, 1991), was provided by a significant Condition x Distance interaction, $F(6, 84) = 23.505$, $p < .01$. Post-hoc Tukey tests demonstrated that the P received higher goodness ratings than the NP, $p < .01$. Furthermore, the NP received significantly lower ratings than any stimulus in the P condition, $p < .05$. Conversely, the P received significantly higher ratings than every stimulus in the NP condition ($p < .05$) except -60 and -90 ($p > .05$). Figure 2 provides a depiction of this response tendency.

Implications for an Evaluation of the PME

In order to assess a potential magnet effect, it is essential that stimuli fall within the same category. Almost all of the timbre stimuli were identified to be within the violin category over 90% of the time. Though the three stimuli to the right of the NP in this experiment seem to be approaching another timbre category, they still were reliably

identified to be within the violin category.

It also is critical to the magnet effect that there are differences in perceived goodness of the stimuli within the category. The results of the goodness rating analyses indicated that there were considerable mean differences in perceived quality between the P and NP (and between the stimuli that surround them). The considerable differences in goodness between the P and NP conditions, in combination with the high percentages of within-category responses, ensure that the present stimulus set should permit an appropriate evaluation of a potential magnet effect.

CHAPTER 4

EXPERIMENT 2

The purpose of Experiment 2 was to determine if a PME can be obtained with musical timbres utilizing a subset of stimuli from Experiment 1. As part of this determination, the experiment included an evaluation of the role of context in identification and goodness ratings of timbre stimuli. Similar to the approach used by Lotto et al. (1998), listeners completed identification and goodness ratings within the context provided by a standard timbre stimulus (P or NP). To allow a direct comparison of how context affects identification and goodness ratings, listeners also completed the tasks in the absence of context. Analogous to the speech work that has used a discrimination task to compare performance with a P, versus a NP, standard (e.g., Iverson & Kuhl, 1995; 1996; 2000; Kuhl, 1991; Lively & Pisoni, 1997; Lotto et al., 1998), a discrimination task for each standard and its variants also was administered to listeners. Consistent with speech work that supports the PME (e.g., Kuhl, 1991), as well as with the results of Experiment 1, it was hypothesized that there would be substantial differences between the goodness ratings of stimuli in the P and NP conditions. It also was predicted that discrimination would be significantly better around the NP than around the P, revealing a PME. Also, similar to the findings of Lotto et al. (1998) it was predicted that identification and goodness ratings would vary as a function of context.

Method

Participants

Twenty-five listeners from the University of Nevada, Las Vegas with no reported hearing deficits participated in this experiment. Their ages ranged from 18 to 42 years (mean age = 24.3 years). Eleven listeners were introductory psychology students who participated in partial fulfillment of course requirements. The remaining listeners consisted of seven psychology graduate students and seven laboratory personnel. Both the graduate students and the laboratory listeners had little to no knowledge of the experiment, and data of these two groups looked very similar. There was, however, a slight difference between these two groups of listeners and the psychology 101 students. Graduate students and laboratory personnel tended to use the entire goodness rating scale more often, which resulted in more widely distributed goodness ratings among stimuli. Similar to Experiment 1, all listeners were surveyed for musical training using the provided musical history questionnaire (see Appendix II). Listener's musical experience ranged from 0 to 18 years (mean = 6 years).

An a priori criterion was set for inclusion in the data analyses to ensure that the discrimination task could accurately assess each listener's prototype. The criterion was that each listener had to assign the P a higher goodness rating than the NP in the rating task without context, as well as to identify (at ceiling level) more than half of the stimuli in the P condition to be within the violin category. Nine listeners did not meet the criterion. Therefore, the results are reported from the remaining sample of sixteen participants. Of these remaining listeners, five were psychology 101 students, six were laboratory personnel and five were psychology graduate students.

Stimuli

Stimuli in Experiment 2 were selected from a single diagonal vector taken from the stimulus set generated in Experiment 1. Figure 3 represents the 14 stimuli, which consisted of the P and the NP, as defined by Experiment 1, plus the closest variant stimuli around P and NP. As can be seen in Figure 3, the first variant to the left or right was 30 mels away from the P (henceforth labeled – or + 30, respectively), the second variant was 60 mels away (i.e., labeled – or + 60), and the third variant was 90 mels away (i.e., labeled – or + 90). For example, in the P condition, stimuli consisted of the P and 3 variants to the left (stimuli -30P, -60P and -90P) and 3 to the right (stimuli +30P, +60P and +90P) along the vector.

Procedure

First, listeners received a familiarization period of 14 randomized trials consisting of one presentation of each stimulus. Tones in the familiarization phase were presented with a 1000 ms ISI. Listeners then completed the identification and goodness rating task both with and without context. Identification and rating responses were obtained in the manner previously described for Experiment 1. Each context condition consisted of four blocks of trials: two in which the P served as the standard stimulus for its 6 closest variant stimuli (3 to the left, and 3 to the right along the vector of stimuli), and two in which the NP served as the standard for its 6 variant stimuli.

In the task with context, there were 14 possible pairs of stimuli total across P and NP context. Table 1 provides a depiction of all possible pairs of stimuli that were used in the identification/goodness rating task and in the subsequent discrimination task. The pairs of stimuli were presented in an AX form where A was the standard stimulus and X was a

comparison stimulus that was either physically identical to the standard or was one of the 6 deviant stimuli. Across trials, the standard was paired with comparison stimuli 0, 30, 60, and 90 mels away in both directions along the selected vector. For example, a trial in which the standard was paired with a comparison stimulus 0 mels away consisted of P-P; a trial in which the standard was paired with a comparison stimulus 30 mels away consisted of P+30 or P-30. All pairs of stimuli were presented with a 250 ms ISI and trials were separated by a 500 ms ITI.

In any one block of trials, there were 140 randomized trials (i.e., 560 trials across blocks). Each block of trials consisted of 20 randomized repetitions of all pairs of stimuli in one order in either the P or NP context. The order of blocks of trials with respect to the standard stimulus (P or NP), as well as which comparison stimulus within a trial that was to be identified and rated (first or second), was counterbalanced across participants. One block of trials in the P and NP context consisted of the seven possible pairs of stimuli in AX order, and one block consisted of the seven pairs in the reverse, XA, order (see Table 1). In two of the blocks, one from the P context condition and one from the NP context condition, listeners identified and rated the first stimulus in the pair; they identified and rated the second stimulus in the remaining block of trials from each condition. Whether presented first or second, the comparison stimulus (see X column in Table 1) was always the target to be judged.

In the task without context there were 4 blocks of 140 randomized trials (560 trials total). Each block of trials consisted of 20 repetitions of each of the 7 possible stimuli (P or NP, plus the 6 corresponding variant stimuli). Only one stimulus was presented on each trial in this task. Both the order of blocks (P or NP stimulus set) and the order of

tasks (with or without context) were counterbalanced across participants.

Next, listeners completed the discrimination task. The listeners' task was to indicate by a key press on the keyboard whether the two tone stimuli were the same (as indicated by the 1 key) or different (as indicated by the 3 key). The stimuli were presented in two counterbalanced blocks of trials; one block of trials consisted of the P as the standard stimulus, and the other block of trials consisted of the NP as the standard stimulus. In one block, there were 280 randomized trials consisting of 20 repetitions of the 7 pairs of stimuli, in both possible orders (see Table 1). Pairs of stimuli, the number of repetitions per stimulus pair, ISI and ITI were the same as described for the identification and goodness rating task with context. The entire experimental session lasted approximately 2 hours. Multiple rest breaks were provided throughout the experiment, after each block and in between tasks.

Results and Discussion

Identification

Figures 4 and 5 represent the mean within-category identification responses and goodness ratings for all 14 stimuli in the tasks without context and with context, respectively. Mean identification responses were submitted to a mixed model repeated measures ANOVA with two between subjects factors [Order (8 possible) and Musical History (0-18 years)] and three within subjects factors [Condition (P and NP), Distance (0, ± 30 , ± 60 , and ± 90 mels) and Context (with context and without context)]. There was no significant effect of Order, $F < 1$, so data were collapsed across that variable for subsequent analyses.

Listener's amount of musical history affected identification responses to the timbre stimuli. Musicians with extensive training were less likely to rate the synthesized timbre stimuli as natural violin tones. This finding was supported by a significant effect of Musical History, $F(6, 9) = 5518.312, p < .01$. Planned orthogonal comparisons indicated that listeners with 14 or more years of musical experience gave significantly fewer within-category responses than listeners with 10 or less years of experience, $p < .01$. This finding indicates that the timbre stimuli may not have represented as natural a category to listeners with the most experience and training with music.

Although listeners with an extensive musical history assigned lower within-category identifications to the timbres, this was a small reduction and within-category identifications were still considerably high. Listeners with 14 or more years of musical training assigned the timbres as within the violin category 81.8 percent of the time as compared to 99.9 percent from listeners with 10 years or less of training. Particularly, the listeners with more musical training seemed to be more reluctant to assign within-category identifications to stimuli NP +60 and NP +90. Their mean within-category identifications for these stimuli were 44.4 and 33.8 percent, respectively. Listeners with 10 years or less of musical experience assigned these same stimuli 100 and 99.3 percent within-category identifications, respectively. Despite the reduction in within-category identifications from listeners with extensive musical experience, it was still clear that stimuli were almost always labeled as members of the intended category when averaged across all listeners, stimuli and context conditions (mean = 97.7 percent). This consistent identification performance is evident from the mean within-category identification percentages for all stimuli in the task with context and without context, which were 97.0

and 98.3, respectively. In order to examine a potential PME, it is essential that the stimuli fall within the same category. The percentages of within-category responses in this experiment were at ceiling levels and are substantially higher than have been typically reported in the speech literature. For example, the mean percentage of within-category identifications was 72.6 percent in Iverson and Kuhl's (1995) study as compared to 97.7 percent in this experiment.

Because all of the stimuli received a high percentage of within-category responses, there were no differences in identifications assigned to stimuli in the P and NP conditions [as indicated by the lack of a significant main effect of Condition, $F(1, 15) = 2.595, p > .05$]. Also, because of the high percentage of violin identifications, there were no significant differences in response tendencies as a function of a timbre's distance from the standard (the P or NP), $F(6, 90) = 2.075, p > .05$. Despite the high percentages of responses within the intended category, there were still some significant trends in categorization rates. There was a significant Condition x Distance interaction, $F(6, 90) = 2.292, p < .05$. Post-hoc Tukey tests revealed that the NP+90 stimulus received a significantly lower percentage of within-category judgments than all other stimuli except the stimulus NP+60, $p < .05$. This finding is similar to Experiment 1 in that the NP+90 stimulus received the lowest percentage of within-category responses. The NP+90 tone clearly was not perceived as approaching a category boundary, given that it was classified as within-category 91.09 percent of the time. However, the significant reduction in categorization rate for this stimulus relative to all others could be a potential problem for the discrimination analysis. Specifically, the lower categorization of this stimulus could introduce the possibility of inter-categorical effects in the discrimination task. Inter-

categorical effects would be particularly problematic for a discrimination analysis based solely on goodness factors. To address any potential inter-categorical concerns an evaluation of the respective contributions of identification and goodness was included in the discrimination analysis. A regression analysis examining factors relating to both identification and goodness (see *Discrimination* section in this experiment) was conducted to eliminate this concern.

Within-category identification responses of timbre did not significantly vary as a function of context. This interpretation was indicated by the failure to find a significant effect of Context, $F < 1$. The lack of any contextual effects on identification of the timbre stimuli conflicts with the results of Lotto et al. (1998) study with speech stimuli, which found apparent contextual influences on category identification. The lack of contextual influences in this experiment implies that the category of the timbre stimuli was not being judged relative to the additional stimulus presented on each trial, but rather that category identifications were based on the perceived category of the target tone.

Goodness Ratings

Mean goodness ratings from all stimuli that were labeled to be within the (violin) category are depicted in Figures 4 and 5. The mean goodness ratings were initially submitted to a mixed model, repeated measures ANOVA. The ANOVA examined the effect of two between-subjects factors [Order (8 possible orders) and Musical History (0-18 years)] and three within-subjects factors [Condition (P and NP), Distance (0, ± 30 , ± 60 , and ± 90) and Context (with and without)] on goodness ratings. The ANOVA did not reveal a significant effect of either Order, $F < 1$, or Musical History, $F(6, 9) = 2.204$, $p > .05$. Therefore, data are collapsed across these variables in all reported analyses.

Context did not play a considerable role in the goodness ratings of the timbre stimuli. The mean goodness ratings across all stimuli in the task with context and without context were 4.13 and 4.02, respectively. This difference was not significant (i.e., there was no significant effect of Context from the ANOVA, $F(1, 15) = .672, p > .05$). This null finding is in contrast to Lotto et al. (1998) study that did find significant contextual influences on the goodness ratings of steady-state vowel stimuli. This different result could be due to the fact that the stimuli used in speech PME paradigms have typically represented more than one category, and thus have been more susceptible to the influence of context. The presence of an alternative category in the speech cases may have resulted in heightened perceptual contrast, making goodness ratings among stimuli bordering between categories more salient. Conversely, because the stimuli used in this study were consistently identified to be within one category, the goodness ratings among stimuli were closer together than in speech and less likely to be affected by context.

As in Experiment 1, the stimuli retained for Experiment 2 differed in goodness as members of the timbre category. The mean goodness ratings for all stimuli in the P and NP conditions were 4.72 and 3.43, respectively. This difference contributed to a significant effect of Condition, $F(1, 15) = 22.716, p < .01$. As can be seen in Figures 4 and 5, stimuli in the P condition received higher goodness ratings than stimuli in the NP condition. Planned orthogonal comparisons indicated that this trend was significant, $p < .01$. This finding ensures that the stimulus set is capable of evaluating a potential magnet effect by having significant differences in perceived quality across the P and NP conditions. This finding also is consistent with the speech research with analogous vowel stimuli (e.g., Kuhl, 1991) that has similarly found that stimuli in the P condition receive

higher goodness ratings than stimuli in NP condition.

As expected from corresponding speech work (e.g., Iverson & Kuhl, 1995), the goodness ratings for timbres varied as a function of distance from the standard (i.e., the P or NP). This was demonstrated by a significant effect of Distance, $F(6, 90) = 7.951, p < .01$. Orthogonal comparisons demonstrated that goodness ratings of stimuli +60 and +90 mels away from the standard received significantly lower ratings than all other stimuli, $p < .05$. No other comparisons were significant. See Figures 4 and 5 for the mean ratings of the +60 and +90 stimuli with reference to the other timbres in the task without context and with context. The main effect of Distance is consistent with a corresponding effect from Experiment 1. However, the nature of the effect is not as symmetrical. In Experiment 2 the only significant differences in goodness ratings as a function of distance were those of +60 and +90 mels away from the standard.

The timbre stimulus set also was found to be qualitatively graded. Although there was not a significant Condition x Distance interaction, $F(6, 90) = .828, p > .05$, post-hoc Tukey tests did reveal several important significant differences between stimuli in terms of mean goodness ratings. Specifically, the P stimulus was rated significantly higher than any stimulus in the NP condition, $p < .05$, and the NP was rated significantly lower than any stimulus in the P condition, $p < .05$. This finding provides evidence that the stimulus set used to examine the magnet effect consisted of distinct prototypical and nonprototypical regions. The substantial differences in perceived quality among the stimuli, in addition to the high within-category identification rates, allows it to be reasonably argued that the timbre stimuli can be used to provide an appropriate assessment of the magnet effect.

Discrimination

The discrimination data was analyzed according to the framework of signal detection theory. Previous speech research on the PME has used d' to evaluate discrimination (e.g., Iverson & Kuhl, 1995; Lotto et al., 1998) on a disproportionate amount of same and different trials, even though d' is typically calculated from different sets of observations (i.e., an equivalent number of same and different trials). Having nonequivalent same and different trials may be problematic because the data from the same trials (and thus, false alarm rates) in the speech studies may be used repeatedly across stimulus comparisons, resulting in measures of d' that are correlated to some degree (i.e., not independent).

Another problem with using d' here is that the AX (same/different) paradigm is a difficult task for listeners. According to Macmillan and Creelman (1991), the demands of the task may lead to two different response strategies. The first is the independent observations strategy. In this strategy each stimulus on a discrimination trial is categorized independently. Once the observer has made an independent determination of each stimulus, a decision is made regarding whether the two stimuli are the same or different. The second strategy is the differencing strategy. When an observer adopts this strategy, the two stimuli are compared, and the perceived acoustic differences between them are subtracted out. The listener has a decision criterion that must be exceeded to respond “different”; otherwise, the stimuli will be deemed sufficiently similar to respond that they were physically identical. The two different response strategies may create two distinct decision spaces, and thus two different calculations of d' in a same/different paradigm. It is difficult to determine which calculation of d' would be correct because the preferred strategy of the observer is unknown, as is the consistency of that strategy.

One way to potentially avoid these problems of strategy is to rely upon $p(c)_{\max}$ as a measure of sensitivity instead of d' . The $p(c)_{\max}$ statistic is consistent across both response strategy situations. It is a true sensitivity measure that does not depend on the particular strategy employed by the listener in the same-different task (Macmillan & Creelman, 1991). Analysis of the discrimination data in this experiment were conducted using $p(c)_{\max}$ scores. $P(c)_{\max}$ is a sensitivity measure related to d' that adjusts for guessing; it represents the percentage of correct responses that would be obtained if the listener was unbiased in his/her responding (Macmillan & Creelman, 1991). It is a more appropriate statistic to use for the present analyses than d' because the proportion of same (P-P or NP-NP) trials is different from the proportion of different trials (e.g., P-30, P+90).

Figure 6 represents the mean $p(c)_{\max}$ scores for each stimulus pair, which consisted of one comparison tone and its corresponding standard (either P or NP). To obtain the $p(c)_{\max}$ scores, d' scores were calculated for each tone pairing across subjects by comparing the percentage of hits to the percentage of false alarms. Hit rates were obtained from “different” responses when the stimuli were different (e.g., P-3). False alarm rates were obtained from “different” responses when the stimuli were the same (e.g., P-P). The d' scores were calculated using the formula: $d' = z(H) - z(F)$ where z is the z-score transformation, H refers to the percentage of hits, and F refers to the percentage of false alarms. The d' scores were then used to calculate $p(c)_{\max}$ scores, which were obtained from the formula: $P(c)_{\max} = [(z(H) - z(F)) / 2]$ (for further details on this procedure, see Macmillan & Creelman, 1991). The $p(c)_{\max}$ values were submitted to a mixed model ANOVA with Order (8 possible) and Musical History (0-18 years) as between-subjects factors, and Condition (P and NP) and Distance (± 30 , ± 60 , and ± 90) as

within-subjects factors. There was no effect of Order, $F(7, 8) = 1.385, p > .05$ or of Musical History, $F(6, 9) = 1.286, p > .05$. Therefore, data were collapsed across these two factors for further analyses.

The analyses revealed a tendency for discrimination performance to improve as psychoacoustic distance from the standard increased. This was demonstrated by a significant effect of Distance, $F(5, 40) = 16.127, p < .01$. Planned orthogonal comparisons revealed that $p(c)_{\max}$ scores of the stimulus pairs consisting of the standard and -30 comparison stimulus (P-30 and NP-30) were significantly lower than $p(c)_{\max}$ scores of all other stimulus pairs except the standard and the +30 comparison stimulus, $p < .05$. Also, as can be seen in Figure 6, the standard and +30 stimulus pair received lower $p(c)_{\max}$ scores than the standard paired with stimuli -90, +60 and +90, $p < .05$. The standard and the ± 60 pairs received lower $p(c)_{\max}$ scores than the standard and +90 pairs, $p < .05$. Also, the standard and -60 pairs received lower scores than the -90 comparisons, $p < .05$. These findings taken together imply that as physical distance between the two stimuli in a pair increased, perceptual differences increased as well.

Although there were significant differences in stimulus pairs as psychoacoustic distance between the comparison tone and the standard increased, there were no differences in discrimination performance between the P and NP conditions, mean $p(c)_{\max}$ scores were 0.656 and 0.684, respectively. The small mean difference between the two conditions resulted in a null effect of Condition, $F(1, 8) = 1.943, p > .05$. In contrast to the hypothesis, which was based on claims for a PME with vowels (e.g., Iverson & Kuhl, 1995; Kuhl, 1991), this finding suggests that a PME may not occur with timbre stimuli.

Across-Task Evaluations of the Magnet Effect

Some additional analyses of the discrimination data have been argued to be necessary to control for any inter-categorical confounds. For example, Lotto et al. (1998) attempted to control for any confounds resulting from inter-categorical discriminations in a speech paradigm by taking into account phonemic identity. Lotto et al. compared observed differences with predicted differences in generalization scores based on phonemic identity. Generalization scores are calculated from the percentage of misses on different trials, and have previously been used by Kuhl (1991) to demonstrate a PME. To demonstrate a magnet effect, generalization errors should be committed more often in the P condition than in the NP condition. Guenther (2000) has suggested that the approach of Lotto et al. (1998) may be confounded because it is based on phonemic identity alone. Guenther claims that their findings may have been invalid because they failed to take spectral differences (distance from P) into account. Generalization differences between the P and NP conditions can only be appropriately assessed if based on phonemic identity and spectral differences. Analyzing the PME data with generalization scores also has been criticized because it may not reliably estimate sensitivity and response bias (Iverson & Kuhl, 1995).

To correct for the potential confound of basing analysis of discrimination data solely on phonemic identity, Lotto (2000) provided a different analysis to subtract out the effects of context from discrimination. In this analysis, step-wise regression was conducted with 4 factors: spectral distance (distance from the standard in mels), PI (probability that both stimuli in the pair will be identified as the same), goodness ratings for the standard, and average goodness ratings for pairs. To demonstrate a PME, either of

the last two factors should account for a significant amount of variance in the model. To the contrary, Lotto found that only the first two factors contributed significantly to the model for vowels. As a result, it was suggested that the data from Lotto et al. (1998) do not indicate a PME when corrected for context effects.

Based on the suggestion that a step-wise regression is more appropriate than using generalization scores to examine the effects of context (Lotto, 2000), a corresponding regression analysis was included in this study. The same four factors used by Lotto (2000) were tested in various combinations and fitted into a step-wise regression equation in an order based upon how much variance each contributed to the dependent variable, $p(c)_{\max}$. Criterion for entry into the model was $p < .05$, and criterion for removal was $p < .10$. In the resultant model spectral distance was entered first, $R^2 = .226$, $F(1, 190) = 55.563$, $p < .01$. The second factor to enter the model was PI, $R^2 = .269$, $F(2, 189) = 34.832$, $p < .01$. The two goodness factors, goodness rating of the standard and average goodness of the pair, did not contribute significantly to the model, $p > .05$.

The regression analysis demonstrates that category goodness did not make a contribution to discrimination performance beyond that of the factors related to identification and categorization. The results of the regression analysis for instrument timbre are consistent with the conclusions of Lotto et al. (1998) for vowels in that factors relating to goodness also did not contribute significantly to the model in speech.

Despite the use of a stimulus set with clear differences in perceived goodness and with minimal influences from other timbre categories, there was no evidence of a PME. There were no differences in discrimination between the P and NP conditions. Further evidence that no magnet effect occurred with the timbre stimuli was provided by the

results of the regression analysis. The analysis indicated that category identification, rather than goodness, was driving discrimination performance. Despite the fact that stimuli in this experiment were reliably identified within a single category, there was one stimulus, NP+90, that was identified as a violin timbre significantly less often than most other timbres. The identification rate of this stimulus, along with the regression analysis, suggests that there may have been inter-categorical interference in the discrimination task. To further determine the nature and extent of any potential inter-categorical effects, Experiment 3 was conducted.

CHAPTER 5

EXPERIMENT 3

Experiment 3 was conducted to determine if categorical perception occurs with the timbre stimuli used in Experiment 2. Previous claims of a magnet effect (e.g., Kuhl, 1991) have been complicated by the possibility that discrimination may be based on inter-categorical comparisons. It is problematic to claims of a magnet effect that discrimination performance may be based on stimuli representing more than one category because enhanced discrimination around the NP could be a result of the NP or its comparison stimuli bordering between two different categories. Discrimination may be enhanced because the perceptual differences between stimuli representing two separate categories are typically more salient than differences between stimuli belonging to the same category. Better discrimination performance around the NP has been claimed to be a result of the NP's distance from the category P, rather than inter-categorical perception, in the PME literature (e.g., Kuhl, 1991).

An assessment of categorical perception was included as a check to help determine whether discrimination performance in Experiment 2 was based on stimuli representing more than one category. If categorical perception is obtained with the stimuli, it would clearly indicate multiple categories existing for listeners. This finding also would suggest that inter-categorical perception may have been influenced discrimination performance in Experiment 2. It is unlikely that the timbre stimuli used in Experiment 2 were bordering

between two separate categories, given the high within-category identification rates. A categorical perception task will additionally help to determine if the stimuli were within the intended category by revealing whether multiple categories were present in the stimulus set.

Because categorical perception has not reliably been demonstrated with steady-state vowel stimuli (e.g., Fry et al., 1962), it was hypothesized that categorical perception also would not occur with analogous timbre stimuli. One reason that categorical perception might not be found with the timbre stimuli is that the stimuli may have crossed a category that does not have a distinct boundary. This finding would be indicative of gradual, continuous perception and would be consistent with the speech research that has found continuous perception of vowels (see Fry et al., 1962). Though this finding would suggest that there are not discrete multiple categories, continuous shifts in the labeling and discrimination functions of listeners still could indicate the inter-categorical relationships between stimuli that could have contributed to discrimination performance in Experiment 2.

Another reason not to expect categorical perception with the timbres may be that the stimuli never cross over into additional categories. This finding would be indicated by flat labeling and discrimination functions that do not change or change very little across the stimulus continuum. This possibility seems quite likely given the results of Experiments 1 and 2, which indicated that listeners were consistently rating the stimuli as a single category. Given the reliability of the within-category identification responses in Experiments 1 and 2, combined with the fact that the same stimuli were used in

Experiment 3, there was no reason to suspect that the same pattern of identification responses would not be obtained.

Method

Participants

Ten listeners from the University of Nevada, Las Vegas completed this experiment. Their ages ranged from 20 to 31 years (mean age = 25.7 years). Six listeners were students in an introductory psychology course who received credit for their participation toward partial fulfillment of course requirements. The four remaining listeners were personnel in our laboratory who had very little knowledge about the experiment. Data from the two groups of listeners were only different with regards to discrimination performance. Laboratory personnel did slightly better on the discrimination task, i.e., they reported hearing a difference between the tones more often, than the introductory psychology listeners. As in previous experiments, all listeners were self-screened for hearing impairments and surveyed for musical training using the provided musical history questionnaire (see Appendix II). The listeners' musical experience ranged from 0 to 14 years (mean = 5.7 years).

Stimuli

The stimuli were the 14 tones that were used in Experiment 2.

Procedure

Listeners first received a familiarization period in which all 14 stimuli were presented once in random order and separated by an ISI of 1000 ms. Then, to determine if the timbre stimuli were labeled as discrete categories, as would be predicted by categorical

perception, listeners completed a labeling task. They received two blocks of 140 trials each (280 trials total). Each block consisted of 10 randomized repetitions of the 14 timbre stimuli (i.e., 20 repetitions across blocks). Trials were separated by a 500 ms ITI. Listeners were required to label the instrument for the tone that they heard on each trial by pressing *1* on the keyboard if they perceived the tone to reflect a violin timbre and *3* if they perceived some other timbre.

A discrimination task also was subsequently completed by the listeners. The discrimination task was included because categorical perception typically has been empirically defined in part by performance on a discrimination task (e.g., see Studdert-Kennedy et al., 1970). When perception is categorical, peaks in discrimination performance result for pairs of stimuli that straddle the category boundary. For an individual listener, the peaks in discrimination are predictable from the category boundary obtained in the corresponding labeling function. Troughs of discrimination performance for pairs of stimuli that reflect the same category also are observed in categorical perception.

In the discrimination task listeners received a total of 40 randomized repetitions of every possible adjacent pair of stimuli along the vector/continuum (i.e., there was a fixed step size of 30 mels). For each pair of stimuli, 20 repetitions occurred with the stimuli presented in one order, and 20 occurred in the alternative order. Thus, there were a total of 520 trials in this task. These trials were separated into two blocks of 260 trials, with each block consisting of 10 randomized repetitions of all possible pairs of stimuli. Pairs of stimuli were separated by a 250 ms ISI and trials were separated by a 500 ms ITI.

Listeners were instructed to press the *1* key to indicate that the two tones on a trial

sounded identical (i.e., as if the same tone was presented twice on the trial). Listeners were instructed to instead press the 3 key to indicate when the two tones on a trial sounded different. The experimental session lasted approximately 1 hour. All other aspects of procedures, including filter settings, software-based delivery of stimuli and collection of responses was as described for Experiment 1.

Results and Discussion

Figure 7 represents the stimuli used in this Experiment and the corresponding numbers that were assigned to them for analyses purposes. Each stimulus was assigned a number between 1 and 14 depending upon its place along the continuum. As can be seen in the figure, the timbre with the lowest F1 and highest F2 value was assigned as stimulus 1. Numbers assigned to stimuli increased consecutively as the timbres increased in F1 and decreased in F2.

Labeling

Category boundary location was calculated for each listener as the point at which each stimulus was labeled as a violin 50% of the time. Linear interpolation was used to find the category boundary for each listener. The interpolated boundary yielded a predicted point of peak performance for the discrimination task.

The labeling and discrimination functions from each listener are provided in Figure 8. The broken lines in the figure correspond to the percentage that listeners labeled each stimulus as within the violin category. As can be seen in the figure, each listener's predicted boundary varied considerably. The labeling boundaries ranged from 4.4 along the stimulus continuum (subject 2) to 12.6 along the continuum (subject 3). It was

impossible to determine a predicted boundary for two subjects, 5 and 10, because they did not reliably label any of the stimuli to be within the violin category at least 50% of the time. Therefore, data from subjects 5 and 10 were discarded from further analysis.

As can be seen in Figure 8, the labeling functions of listeners were not uniformly within the violin category. This finding was confirmed by a repeated-measures ANOVA that examined the percentage of violin responses at each stimulus (stimuli 1-14). Musical History (based on total years of musical experience) included in the analysis as a between-subjects factor was not significant, $F < 1$. The analysis indicated a significant effect of Stimulus, $F(13, 104) = 3.925, p < .01$. Orthogonal comparisons revealed that stimuli 4, 5, 6, 7 and 8 along the continuum all received higher within-category identification responses than stimuli 9, 10, 11, 12, 13 and 14, $p < .05$. Although labeling responses to stimuli 1, 2 and 3 were not significantly different from any other stimuli, they did receive higher within-category identifications than stimuli 9, 10, 11, 12, 13 and 14 (mean across stimuli 1-3 was 65.4 percent and across stimuli 9-14 was 32.5 percent). This finding indicates that the last few stimuli to the right of the vector were not as likely to be judged within the violin category. This also suggests that other categories may have been interfering in the labeling task.

In addition to the ANOVA, several examples in Figure 8 demonstrate that the timbres were not consistently labeled as a single category. Most labeling functions reflect gradual slopes (e.g., see data from subjects 1 and 3) moving from high violin category responses to very low within-category responses. For some listeners, a strong shift in labeling (as would be predicted by categorical perception) was observed. For example, the labeling function from subject 4 remains fairly flat until the middle of the stimulus continuum

where it then has a sharply decreasing trend. This sharp decrease in labeling as a violin indicates that stimuli may have been crossing into another timbre category for this listener. An additional example of a steep slope in labeling can be seen in the function from subject 2. For this subject, there was a sharp increase in within-category responses in the middle of the continuum. The nature of the labeling function of subject 2 indicates that only stimuli in the middle of the continuum were perceived as within the violin category.

Discrimination

The percentage of “different” responses for each stimulus pair was calculated to assess potential peaks in individual discrimination performance. The solid lines in Figure 8 represent these percentages for each listener. As can be seen in the figure, overall discrimination performance was quite poor. All subjects, except for subject 2, failed to label the stimuli as different at least 50% of the time (averaged across all stimuli). Discrimination performance for three listeners, (7, 8 and 9), was so poor that they only reported hearing an adjacent pair of stimuli as different 3 percent or less of the time. Because these three listeners were incapable of detecting a difference between any of the stimulus pairs, their data were excluded from further analysis. Data from these three listeners, in addition to the two listeners that were discarded because of labeling performance, necessarily restricted subsequent statistical analyses to the remaining sample of five listeners (1, 2, 3, 4 and 6).

Although discrimination performance was generally poor, a few listeners did demonstrate small peaks in discrimination. For example, subjects 1 and 6 indicated small peaks in the middle of the stimulus continuum. In addition, subject 2 demonstrated two

peaks in discrimination, one at the beginning and one at the end of the continuum. To determine if the discrimination trends noted with subjects 1, 6 and 2 were sufficient to produce a significant average tendency across the five listeners, a repeated-measures ANOVA was conducted examining Stimulus Pair (13 pairs) on the percentage of different responses. Musical History (based on years of musical experience) was included in the analysis as a between-subjects factor. Since there was no effect of Musical History, $F < 1$, data were collapsed across this variable for further analysis. There were no significant differences in discrimination among the stimulus pairs (i.e., no significant peaks in discrimination). This finding was indicated by the failure to find a significant effect of Stimulus Pair, $F(12, 48) = 1.691, p > .05$.

Though there were no significant differences in discrimination performance as a function of the stimulus pair, a few listeners did exhibit enhanced discrimination at certain points along the continuum. However, none of the peaks appear to fall at the predicted labeling boundary (e.g., see subjects 4 and 6 in Figure 8). Peaks in discrimination performance at labeling boundaries are a necessary component to demonstrate categorical perception. To further determine if each listener's labeling boundary predicted any peaks in discrimination performance, a one-way repeated measured ANOVA was conducted to examine the effect of distance from the predicted boundary (0, -30 mels, -60 mels, -90 mels, and +30 mels) on the percentage of "different" responses. The labeling boundaries of each listener did not predict any peaks in discrimination performance, as indicated by the lack of a significant effect of Distance, $F(4, 16) = 1.77, p = .184$. This finding confirms the hypothesis that no categorical perception would occur with the timbres.

Although there were gradual and sharp shifts in labeling, there were no sharp peaks in discrimination performance. In addition, the potential category boundaries from labeling did not predict the location of heightened discrimination performance. These trends in the labeling and discrimination functions indicate that categorical perception did not occur with the timbre stimuli. This finding implies that the timbre stimuli were not perceived as discrete categories, but rather as samples from a gradually/continuously perceived continuum of stimuli consisting of fuzzy boundaries. This finding also suggests that, despite the high within-category identification rates obtained in Experiments 1 and 2, discrimination performance in Experiment 2 was still influenced by inter-categorical tendencies. The nature of the labeling functions obtained in this Experiment suggests inter-categorical effects influenced perceptions of the present stimulus set.

CHAPTER 6

GENERAL DISCUSSION

Three experiments were conducted to address the two major issues in the PME literature. One problem with interpreting the results of speech research on the PME has been the possibility of inter-categorical comparisons. The present investigation attempted to control for inter-categorical interference by utilizing a single, broad category that consisted of substantial differences in quality. Another difficulty with claims of a PME is that it has not been found to generalize to musical stimuli (e.g., Acker et al., 1995; Schellenberg, 2002). The results of the musical versions of the PME are difficult to compare to the corresponding speech literature, however, because all musical versions of the magnet effect reported to date (see Acker et al., 1995; Barrett, 1998; Schellenberg, 2002) have used stimuli that are not analogous to the speech stimuli that have been used to examine the effect. The present investigation attempted to correct this problem by using musical, timbre stimuli that are more analogous to vowels than the previously used musical chord stimuli.

Experiments 1 and 2 established that the timbre stimuli used to assess the magnet effect represented a broad category. In both experiments, the violin tones were identified to be within the violin category at ceiling levels. Experiment 3 provided additional evidence that the timbre stimuli used to examine the PME represented a broad category. In Experiment 3, stimuli were generally perceived to be within the violin category for

most listeners. In addition, a categorical perception task provided evidence that even if an additional category may have been occasionally present, it was not discretely perceived relative to the violin category. Rather, the data revealed gradual, continuous perception.

In addition to the evidence that the timbre stimuli represented a broad category, the present findings suggest that, though inter-categorical effects were obtained, they were minimized as much as possible. Inter-categorical effects were obtained in Experiments 2 and 3. In Experiment 2, a regression analysis indicated that discrimination was based on categorization tendencies rather than category goodness. In Experiment 3, the labeling functions of listeners indicated that some timbres, typically those at the right-most end along the vector, were more often identified as a category other than the violin. Although these inter-categorical effects existed, the high within-category identification results of Experiments 1 and 2 and of the corresponding stimuli in Experiment 3 suggest that they were reduced. Having a broadly defined category that minimized the potential contributions from inter-categorical comparisons should have allowed the present study a cleaner evaluation of the magnet effect than the speech cases (e.g., Iverson & Kuhl, 1995).

Just having a single, broad category with minimal inter-categorical interference is not sufficient for evaluating a PME. The stimuli also should consist of differences in perceived goodness. Having salient differences in goodness is important in examining a magnet effect because the effect is claimed to be a result of the category being organized in terms of category goodness. Significant differences in perceived goodness were found in Experiments 1 and 2. Goodness ratings of listeners in both experiments indicated that

the P and its surrounding stimuli were rated significantly higher than the NP and its surrounding stimuli.

Even with a stimulus set that was designed to get around the problems in the speech and music PME paradigms, no evidence of a PME was obtained with timbre. In order to demonstrate a PME, discrimination should be significantly worse around the P than around the NP. However, analysis of $p(c)_{\max}$ scores in Experiment 2 did not indicate any significant differences between the P and NP conditions. A step-wise regression also was conducted on two factors relating to goodness and two relating to identification. The analysis indicated that goodness did not contribute to discrimination performance beyond the role of identification tendencies. This finding indicates that categorization tendencies, rather than differences between stimuli in terms of relative goodness, were driving discrimination performance.

Comparisons with Speech and Musical Research

In comparison with the typical speech demonstrations of the PME (e.g., Iverson & Kuhl, 1995), the present study defined a more broad stimulus category. Stimuli in the present investigation were identified to be within the intended category at substantially higher rates and at a farther psychoacoustic distance from the center tone than in Iverson and Kuhl. Stimuli in Iverson and Kuhl's study were only 120 mels away from the center stimulus when within-category identification responses dropped below 60 percent. In the present investigation comparable within-category rates were obtained for stimuli as far as 210 mels from the center stimulus. Having high within-category identifications at a farther distance from the center in this study suggests that the timbre stimulus set was less

likely to be approaching a category boundary than the speech stimuli that have typically been used to examine the PME.

Another difference between the present study and the speech research on the PME is the nature of the goodness ratings. Goodness ratings in this study were lower overall and closer together than in Iverson and Kuhl (1995). Though, the goodness ratings in this study were significantly different between the prototype and nonprototype conditions. The lower goodness ratings in this study with musical timbres may have occurred because the stimuli were more representative of a single category than the speech stimuli that have been used to assess the PME. The substantial differences in goodness in the speech studies may be a result of stimuli bordering between categories. A stimulus on the border of two separate categories in the speech case would most likely be assigned a much lower goodness rating than one more representative of the category. Another reason that the timbre stimuli received lower goodness ratings than the speech stimuli may be that the timbres represented a more artificial category that does not relate well to natural violin tones.

An additional difference between the present investigation and the speech paradigms of the PME is the nature of context effects. Contextual influences in the speech research have been central to the claim that the PME may be a reflection of inter-categorical comparisons because context has been found to affect the perceived category of speech stimuli (e.g., Lotto et al., 1998). Contextual influences on identification of vowel stimuli may occur as a consequence of crossing into additional categories in the speech research. Context effects may be accommodated by the presence of additional categories because of a contrast effect. For example, poor representatives of a category that approach a

boundary may sound like exemplars from an unintended category when presented with a good exemplar (see Lotto et al., 1998). Conversely, poor exemplars that are close to a category boundary may be perceived as within the intended category when paired with another exemplar that reflects an alternative category. Similar shifts in identification responses have been found in vowel contrast studies (e.g., Thompson & Hollien, 1970).

No effects of context were found in the present investigation. In Experiment 2 listeners completed the identification and goodness rating tasks with and without the context of an additional stimulus. There were no significant differences in the responses between the context and no context tasks. This finding is likely due to the fact that the stimuli were representative of a single category. It may be possible to obtain context effects with stimuli representing only one category, but such effects are not likely to be strong enough to elicit shifts in the perceived category of the stimuli. Reliance on a single category may reduce the likelihood of context effects because of smaller perceptual differences between context and stimulus.

The present results also differ from previous musical evaluations of the PME (Acker et al., 1995; Barrett, 1998). Researchers utilizing musical paradigms of the PME have found varying outcomes. It has been suggested that musicians demonstrate an anchor effect (see Acker et al., 1995; Barrett, 1998), and it has been claimed that non-musicians will demonstrate a magnet effect (see Barrett, 1998). This latter claim directly contradicts the fundamental claim of the PME for speech (i.e., that the PME is dependent upon experience with the sounds). Another study has found that neither an anchor nor a magnet effect occurs in musicians and non-musicians (Schellenberg, 2002). The results of the musical versions of the magnet effect have been difficult to interpret, however, because

of the use of stimuli that are not analogous to the vowel stimuli originally used to demonstrate the effect (Kuhl, 1991). Using more analogous musical stimuli, the present investigation still did not find a PME or an anchor effect in either musicians or non-musicians. The present results are most consistent with the findings of Schellenberg (2002) in that they suggest that that no magnet effect occurs with musical stimuli.

Contrary to the claim that musical history influences the presence of a magnet effect with musical stimuli (as claimed by Barrett, 1998), listeners' musical background had no influence on discrimination performance. Musical history only influenced identification responses to the stimuli in Experiments 1 and 2. Listeners with more years and more extensive experience playing one or more instruments were more likely to designate certain stimuli as out of the violin category. This finding may reflect the reluctance of those with an extensive musical background to determine that a synthesizer is approximating a natural instrument category. This would be expected given the depth of experience with natural instruments these listeners have. This finding does not threaten the validity of the evaluation of the magnet effect in this study, however, because within-category identification responses were still at ceiling levels when averaged across listeners.

Theoretical Implications

Several interpretations may be posited to explain why a PME was not observed with the musical timbre stimuli in the current investigation. One possible explanation could be that the PME only occurs with speech stimuli. This account suggests that there may be different mechanisms involved in how speech and nonspeech (musical) sounds are

represented in memory. For example, the claim could be made that speech sounds are organized around prototypes that attract surrounding exemplars and warp perceptual space, but that nonspeech sounds are organized in terms of category boundaries. This idea is central to Kuhl's (1993) NLM theory. According to Kuhl's theory, category prototypes become fine-tuned with exposure to one's native language. As a result, we become more sensitive to the sounds of our native language and lose the ability to detect sounds from foreign languages early in infancy. The role of the prototype then becomes to insure proper categorization and, to some extent, to disregard finer acoustic differences between stimuli. However, given that the role of experience is fundamental to Kuhl's theory, the NLM theory could be argued to suggest that a magnet effect could be observed with musicians given enough exposure to a particular instrument's timbre.

To posit that the PME may only occur with speech stimuli also suggests that the PME may reflect the operation of a speech-specific module in the brain (see Eimas, 1963). If one assumes that the speech claims of a PME are valid (e.g., Iverson & Kuhl, 1995) and the musical claim of a PME with non-musicians (see Barrett, 1998) is not valid, then the failure to obtain the effect with timbres would make it plausible that such a specialized mechanism exists. Further support for this position has been provided by the failure to demonstrate the PME in non-human species, which would not be posited to possess a biologically specialized mechanism for processing human speech (see Kuhl, 1991).

There is, however, a fundamental reason why the claim that the PME only occurs with speech stimuli should not be regarded as the most likely explanation for the present findings. In order to demonstrate a magnet effect, stimuli should fall within one category, or, alternatively, there should at least be a tendency within the category to differentiate

the PME from inter-categorical perception. The current timbre stimuli were more frequently labeled to reflect a single category than in the corresponding speech research (e.g., Iverson & Kuhl, 1995). In fact, the timbres appeared to be almost solely representative of one category. In contrast, the stimulus sets used thus far in the speech research have been argued to contain several tokens that have been frequently identified by most listeners as out of the intended vowel category (see Sussman & Lauckner-Morano, 1995). The speech examples therefore could be argued not to have met the conditions necessary to demonstrate the PME. In addition, despite the fact that the stimuli reflected a single category in this study, there were still significant changes in goodness that are necessary to evaluate a potential PME. The fact that the present stimulus set was more so prepared to appropriately examine a PME (with high within-category identification rates and qualitative differences among stimuli) than the speech cases makes the speech-specific argument a questionable interpretation.

An alternative interpretation of the present results may be that a magnet effect was not observed because of a reliance on unnatural stimuli. The argument could be made that the synthesized violin tones did not represent a natural category for listeners. It could then be suggested that the present study may have been incapable of assessing a potential magnet effect because listeners were not comparing the tones to a memory of a natural violin. This would presume that listeners did not have a representation in memory for a category that relates to the stimulus set. It has been implied in the speech literature as well that some listeners may not demonstrate a magnet effect because it is too difficult to rate and identify a synthetic category (see Aaltonen et al., 1997). This explanation could account for why the goodness ratings of the timbre stimuli were fairly moderate as

compared to those reported in the speech studies. The average rating for the prototype in the present study was 4.8 on a 7-point scale, as compared to an average rating of 6.1 for the vowel stimuli presented in Iverson and Kuhl (1995).

This alternative explanation also is not very likely because of the nature of the instructions that were given to listeners. In Experiments 1 and 2, listeners were not asked to compare the tones with a natural violin. Rather, they were informed that the tones all were produced by an analogue synthesizer, and were intended to approximate only a few general characteristics of natural violin tones. Listeners also were told that, as a result, the tones would sound impoverished and artificial compared to the tones produced by an actual violin, and that their task would be to rate each tone with respect to its perceived goodness as an example of a synthesizer's violin sound (i.e., timbre).

While it is acknowledged that the stimuli used in the present study probably do not constitute as natural a category as actual violin productions, reliance on these stimuli does not likely pose a threat to the interpretation of results because of the ease and frequency with which listeners classified the stimuli as members of the violin category. Within-category identification responses were above 68% in Experiment 1 and above 91% in Experiment 2. Additionally, no listener indicated during debriefing that they had difficulty labeling the tones as examples of the violin category. If the PME is based on experience, failing to demonstrate the effect could mean that there is no representation of the category. Yet, no listener had a problem classifying the stimuli as an example of a violin tone. The fact that listeners recognized the stimuli as string tones could suggest that they had a concept or representation of a synthesized string tone, or, alternatively,

that listeners were easily able to relate the artificial stimuli to a different representation, such as that of a natural violin tone.

Another potential explanation of why no PME occurred with timbre stimuli is that the current study did not limit the investigation to listeners who have specific experience with violin tones. Although overall musical history did not affect goodness ratings or discrimination performance of the timbre stimuli in this study, an extensive history with the violin, or string instruments in general, could potentially yield a different outcome. Given that the magnet effect is claimed to be driven by experience with sounds, a more accurate assessment of the effect might be obtained by relying upon listeners who have specific experience playing the violin.

Previous musical explorations of the PME have indicated mixed results in terms of the affect of musical experience, at least for studies relying on musical chord stimuli. It has been suggested that listeners with musical experience will demonstrate an anchor effect (see Acker et al., 1995), and that listeners with little to no experience will show a magnet effect (see Barrett, 1998). It also has been suggested that experience makes no difference, and that neither an anchor nor a magnet effect occurs with musical stimuli (Schellenberg, 2002). None of the previous musical versions of the PME have assessed specific experience with an instrument, only general musical background. Further examination into specific musical experience is beyond the scope of this paper, but is important to explore in further research. An experience-specific evaluation of the PME will allow a determination of whether the effect will occur in listeners with extensive experience with the instrument timbres. Based on the present findings, there is no reason to suspect that listeners with specific experience playing the violin would yield different

discrimination results than those obtained in this study. Discrimination data from one listener that was a violinist in Experiment 2 followed the same pattern as all other listeners. Though a PME may not be demonstrated with experienced listeners, the identification, goodness and discrimination tasks will provide useful information as to the nature of representation in listeners with and without extensive experience with the sounds.

Yet another interpretation of the finding that a PME does not occur with timbre could be that the methods used to explore the effect in this study were different than those in the speech work. In the typical speech examinations of the PME, discrimination performance is measured using d' . In the present investigation, $p(c)_{\max}$ was used to measure discrimination performance. This argument is problematic, however, because d' may not be the most appropriate means of measuring performance in the speech work. In the speech versions of the PME (e.g., Iverson & Kuhl, 1995), d' has been calculated from an unequal number of same and different trials. This is problematic because it creates a scenario in which false alarm rates are continuously reused. The reusing of false alarm rates violates one of the assumptions of d' . The $p(c)_{\max}$ statistic was used in this study to resolve this problem, and still provides a true measure of sensitivity like d' .

The most probable explanation of the current null finding for a PME with timbre stimuli is that if the PME exists, it is a small effect that, at best, is too difficult to disentangle from categorization issues. This assertion is supported by the finding that despite the high within-category identification responses (even at 210 mels from the center tone) and significant differences in perceived goodness among the timbres in Experiments 1 and 2, no significant differences in discrimination performance were

obtained between the P and NP conditions. Additional evidence against the magnet effect was provided by a regression analysis, which indicated that obtained differences in goodness ratings did not significantly contribute to discrimination performance.

The mixed results in the speech literature, i.e., of finding a PME with vowels (e.g., Iverson & Kuhl, 1995; Kuhl, 1991) and not finding a PME with vowels when controlling for inter-categorical comparisons (e.g., Lotto et al., 1995), may be a result of the difficulty in establishing a broad category in speech that does not contain inter-categorical comparisons. The current investigation suggests that it is at least possible that the PME will not be obtained under properly controlled stimulus conditions. The stimuli in this investigation were more properly controlled than speech examinations (e.g., Iverson & Kuhl, 1995) because the timbres represented a broadly defined category combined with stimuli that are reliably identified to be within the intended category. In addition, the stimuli consisted of steady-state violin tones that were synthesized more analogously to vowels than the musical stimuli that have previously been used to examine the effect (see Acker et al., 1995). The reliance on more analogous conditions, though more artificial, should have allowed this study to more closely approximate the approach that listeners would use in responding to vowels.

To summarize, the present investigation suggests that the PME disappears when examined in an appropriate manner. The present findings also indicate that speech and nonspeech (at least musical) sounds may not be centered around category prototypes that serve as magnets to surrounding exemplars. This is not to suggest that category goodness plays no role in perception, but rather that speech and nonspeech perception is not likely based solely on perceived goodness. Rather than being prototype based, speech and

nonspeech perception may rely heavily on category boundaries in identification and discrimination. This claim is consistent with early views of speech perception, which suggest that boundaries between categories are the basis for phoneme perception (see Repp, 1984 for a review).

Future Directions

The present investigation offers a novel approach to exploring the PME by applying musical timbre stimuli to the established speech methods. However, given that the PME is claimed to be a result of extensive listening experience and the resulting formation of a representation for a sound category in memory, a task that assesses the effect with listeners who have specialized training with the stimulus set is needed. The current investigation lacked a subject pool that has specific experience with the violin tones, and thus momentarily leaves this important question unanswered. As mentioned in the preceding section, no reported musical study to date has examined the influence of experience with either a specific instrument or a class of instruments on discrimination performance. This issue is currently being examined in our laboratory. Even if a PME cannot be obtained with experienced listeners, an investigation of specific experience with musical instruments could still reveal important processing and representation differences as a function of experience.

An additional issue that calls for further investigation is the question of whether a PME can be demonstrated with speech stimuli that are reliably identified to be within a single phonemic category. One reason that such a stimulus set has not been created to date may be that the problem of inter-categorical effects is unavoidable. This assertion is

supported in the present results, which suggest that even with considerably high rates of identifying stimuli to reflect a single category (see Experiments 1 and 2), inter-categorical comparisons still appear to influence performance (see Experiments 2 and 3). Future research will need to be conducted that establishes a speech stimulus set that consists of differences in perceived goodness, high within-category identification rates and minimal inter-categorical interference with which to examine a potential PME. Even if such a stimulus set can be created in speech, evidence from the present study suggests that the outlook may not be positive for the PME.

APPENDIX I

TABLES AND FIGURES

Table 1

Stimulus Pairs Used in Identification and Goodness Rating Task with Context and in AX Discrimination Task in Experiment 2

	P Condition		NP Condition	
	A X	X A	A X	X A
1	P -90	-90 P	NP -90	-90 NP
2	P -60	-60 P	NP -60	-60 NP
3	P -30	-30 P	NP -30	-30 NP
4	P P	P P	NP NP	NP NP
5	P +30	+30 P	NP +30	+30 NP
6	P +60	+60 P	NP +60	+60 NP
7	P +90	+90 P	NP +90	+90 NP

Note. Negative deviants represent stimuli to the left of the standard and positive represent the neighboring stimuli to the right.

Figure 1. Values of Stimuli (in Hz) That Were Used in Experiment 1.

Figure 2. Mean Goodness Ratings and Percentage of Within-category Identifications of All Stimuli Used in Experiment 1.

Figure 3. Diagonal Vector of Stimuli Used in Experiment 2.

Figure 4. Mean Goodness Ratings and Within-category Identification Responses for Stimuli in Task With No Context in Experiment 2.

Figure 5. Mean Goodness Ratings and Within-Category Identification Responses for Stimuli in Task With Context in Experiment 2.

Figure 6. $P(c)_{\max}$ Scores for Each Stimulus and its Corresponding Standard from Experiment 2.

Figure 7. Stimuli Used in Experiment 3.

Figure 8a. Experiment 3 Labeling and Discrimination Functions (Subjects 1 – 6).

Figure 8b. Experiment 3 Labeling and Discrimination Functions (Subjects 7 – 10).

APPENDIX II

MUSICAL HISTORY QUESTIONNAIRE

Experiment #: _____
 Participant # _____
 Experimenter Initials: _____

1.

What instruments (including vocals) have you played? (Principal instrument followed by any others)	How many years have you played this instrument?	Approximate period (by dates) that you have played this instrument:

2. Did you _____ take lessons or
 _____ teach yourself?

3. Do you possess any knowledge of music theory, or do you play purely "by ear?"
 _____ know theory
 _____ by ear only

4. If you know theory, check all that you have studied:
 _____ scale patterns _____ formal theory instruction
 _____ chord progressions _____ formal ear training

5. Are you currently taking lessons?
 _____ yes
 _____ no

6. Where have you received instruction
 _____ high school _____ private instruction
 _____ college _____ other (specify _____)
7. Please estimate the number of hours you spend playing, practicing, and performing on average per week: _____ hours.
8. Are you or have you ever been a performing musician?
 _____ yes, currently _____ yes, previously (dates: _____)
 _____ no
9. If yes, under what circumstances?
 _____ parties _____ classical recitals/concerts
 _____ night clubs _____ popular music group
 _____ church _____ other (specify _____)
10. Please estimate the frequency of performances: _____ times a month.
11. Are/were you paid for performing?
 _____ yes, currently _____ yes, previously (dates: _____)
 _____ no
12. Are/were you making your living as a musician?
 _____ yes, currently _____ yes, previously (dates: _____)
 _____ no

REFERENCES

- Aaltonen, O., Eerola, O., Hellström, A., Uusipaikka, E., & Lang, A. H. (1997). Perceptual magnet effect in the light of behavioral and psychophysiological data. *Journal of the Acoustical Society of America*, 101, 1090-1105.
- Acker, B. E., Pastore, R. E., & Hall, M. D. (1995). Within-category discrimination of musical chords: Perceptual magnet or anchor? *Perception & Psychophysics*, 57, 863-874.
- Askenfelt, A. (1991). Voices and strings: Close cousins or not? In J. Sundberg, L. Nord, & R. Carlson (Eds.), *Music, language, speech, and brain* (pp. 243-256). Houndmills, England: Macmillan Press.
- Barrett, S. (1998). *The perceptual magnet effect is not specific to speech prototypes: new evidence from music categories*. Unpublished manuscript, University of Cambridge.
- Boersma, P., & Weenink, D. (1992). Praat: Doing phonetics by computer (Version 4.2) [Computer software and manual]. Retrieved from www.praat.org
- Bradlow, A. R., Nygaard, L. C., & Pisoni, D. B. (1999). Effects of talker, rate, and amplitude variation on recognition memory for spoken words. *Perception & Psychophysics*, 2, 206-219.
- Darwin, C. J., Turvey, M. T., & Crowder, R. C. (1972). An auditory analogue of the Sperling partial report procedure: Evidence for brief auditory storage. *Cognitive Psychology*, 3, 255-267.

- Davis, K., & Kuhl, P. K. (1994). Tests of the perceptual magnet effect for American English /k/ and /g/. *Journal of the Acoustical Society of America*, 95, 2976.
- Donnadieu, S., McAdams, S., & Winsberg, S. (1994). Context effects in timbre space. *Proceedings of the 4th International Conference of Music Perception and Cognition, Montreal*.
- Eimas, P. D. (1963). The relation between identification and discrimination along speech and nonspeech continua. *Language and Speech*, 6, 206-217.
- Fant, G. (1973). *Speech sounds and features*. Cambridge, MA: MIT Press.
- Frieda, E. M., Walley, A. C., Flege, J. E., & Sloane, M. E. (1999). Adults' perception of native and nonnative vowels: Implications for the perceptual magnet effect. *Perception & Psychophysics*, 61, 561-577.
- Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M. (1962). The identification and discrimination of synthetic vowels. *Language and Speech*, 5, 171-189.
- Fujisaki, H., & Kawashima, T. (1969). On the modes and mechanisms of speech perception. *Annual Report of the Engineering Research Institute*, 28, 67-73.
- Grey, J. M., & Moorer, J. A. (1977). Perceptual evaluations of synthesized musical instrument tones. *Journal of the Acoustical Society of America*, 62, 454-462.
- Grieser, D., & Kuhl, P. K. (1989). Categorization of speech by infants: support for speech-sound prototypes. *Developmental Psychology*, 25, 577-588.
- Guenther, F. H. (2000). An analytical error invalidates the "depolarization" of the perceptual magnet effect. *Journal of the Acoustical Society of America*, 107, 3576-3577.
- Guenther, F. H., & Gjaja, M. N. (1996). The perceptual magnet effect as an emergent

- property of neural map formation. *Journal of the Acoustical Society of America*, 100, 1111-1121.
- Guenther, F. H., Husain, F. T., Cohen, M. A., & Shinn-Cunningham, B. G. (1999). Effects of categorization and discrimination training on auditory perceptual space. *Journal of the Acoustical Society of America*, 106, 2900-2912.
- Handel, S. (1989). *Listening*. Cambridge, MA: MIT Press.
- Hillyard, S. A., & Kutas, M. (1983). Electrophysiology of cognitive processing. *Annual Review of Psychology*, 34, 33-61.
- Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review*, 93, 411-428.
- Iverson, P., Hall, M. D., & Kuhl, P. K. (1997). *Context and within-category discrimination of musical chords*. Unpublished manuscript, University of Washington.
- Iverson, P., & Kuhl, P. K. (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *Journal of the Acoustical Society of America*, 97, 553-562.
- Iverson, P., & Kuhl, P. K. (1996). Influences of phonetic identification and category goodness on American listeners' perception of /r/ and /l/. *Journal of the Acoustical Society of America*, 99, 1130-1140.
- Iverson, P., & Kuhl, P. K. (2000). Perceptual magnet and phoneme boundary effects in speech perception: Do they arise from a common mechanism? *Perception & Psychophysics*, 62, 874-886.

- Kendall, R. A. (2001). Music Experiment Development System (Version 2001-A)
[Computer software and manual]. Retrieved from
www.ethnomusic.ucla.edu/systematic/Faculty/Kendall/meds.htm
- Klatt, D. H. (1980). Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America*, 67, 971-995.
- Kuhl, P. K. (1991). Human adults and human infants show a “perceptual magnet effect” for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, 50, 93-107.
- Kuhl, P. K. (1993). Innate predispositions and the effects of experience in speech perception: The native language magnet theory. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. McNeilage, & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 259-274). Dordrecht: Kluwer.
- Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, 255, 606-608.
- Lane, H. L. (1965). Motor theory of speech perception: A critical review. *Psychological Review*, 72, 275-309.
- Lively, S. E. (1993). An examination of the perceptual magnet effect. *Journal of the Acoustical Society of America*, 93, 2423.
- Lively, S. E., & Pisoni, D. B. (1997). On prototypes and phonetic categories: A critical assessment of the perceptual magnet effect in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, 23, 1665-1679.

- Lotto, A. J. (2000). Reply to “an analytical error invalidates the ‘depolarization’ of the perceptual magnet effect” [J. Acoust. Soc. Am. 107, 3576-3577 (2000)]. *Journal of the Acoustical Society of America*, 107, 3578-3580.
- Lotto, A. J., Kluender, K. R., & Holt, L. L. (1998). Depolarizing the perceptual magnet effect. *Journal of the Acoustical Society of America*, 103, 3648-3655.
- Macmillan, N. A., Braida, L. D., & Goldberg, R. F. (1987). Central and peripheral processes in the perception of speech and nonspeech sounds. In M. E. H. Schouten (Ed.), *The psychophysics of speech perception* (pp. 28-45). Dordrecht, the Netherlands: Martinus Nijhoff.
- Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. New York: Cambridge University Press.
- Meyer, D. E. (1970). On the representation and retrieval of stored semantic information. *Cognitive Psychology*, 1, 242-300.
- Miller, J. L., & Volaitis, L. E. (1989). Effect of speaking rate on the perceptual structure of a phonetic category. *Perception & Psychophysics*, 46, 505-512.
- Näätänen, R. (1992). *Attention and brain function*. Hillsdale, NJ: Erlbaum.
- Nosofsky, R. M. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13, 87-108.
- Opolko, F., & Wapnick, J. (1987). McGill University Master Samples [CD-ROM]. Montreal, Quebec, Canada: jwapnick@music.mcgill.ca
- Pisoni, D. B. (1993). Long-term memory in speech perception: Some new findings on

- talker variability, speaking rate and perceptual learning. *Speech Communication*, 13, 109-125.
- Pisoni, D. B., & Tash, J. (1974). Reaction times to comparisons within and across category. *Journal of the Acoustical Society of America*, 15, 285-290.
- Polka, L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, 20, 421-435.
- Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, 77, 353-363.
- Remez, R. E., Pardo, J. S., Piorkowski, R. L., & Rubin, P. E. (2001). On the bistability of sine wave analogues of speech. *Psychological Science*, 12, 24-29.
- Repp, B. H. (1977). Dichotic competition of speech sounds: The role of acoustic stimulus structure. *Journal of Experimental Psychology: Human Perception and Performance*, 3, 37-50.
- Repp, B. H. (1984). Categorical perception: Issues, methods, and findings. In N. Lass (Ed.), *Speech and language: Vol 10. Advances in basic research and practice* (pp. 244-335). Orlando, FL: Academic.
- Ritter, W., Deacon, D., Gomes, H., Javitt, D. C., & Vaughan, H. V., Jr. (1995). The mismatch negativity of event-related potentials as a probe of transient auditory memory: A review. *Ear and Hearing*, 16, 52-67.
- Samuel, A. G. (1982). Phonetic prototypes. *Perception & Psychophysics*, 31, 307-314.
- Schellenberg, E. G. (2002). Asymmetries in the discrimination of musical intervals:

- Going out-of-tune is more noticeable than going in-tune. *Music Perception*, 19, 223-248.
- Sharma, A., & Dorman, M. F. (1998). Exploration of the perceptual magnet effect using the mismatch negativity auditory evoked potential. *Journal of the Acoustical Society of America*, 104, 511-517.
- Studdert-Kennedy, M., Liberman, A. M., Harris, K. S., & Cooper, F. S. (1970). Motor theory of speech perception: A reply to Lane's critical review. *Psychological Review*, 77, 234-249.
- Sundberg, J. (1991). *The science of musical sounds*. San Diego, CA: Academic Press.
- Sussman, J. E., & Gekas, B. (1997). Phonetic category structure of [I]: Extent, best exemplars, and organization. *Journal of Speech, Language, and Hearing Research*, 40, 1406-1424.
- Sussman, J. E., & Lauckner-Morano, V. J. (1995). Further tests of the "perceptual magnet effect" in the perception of [i]: Identification and change/no-change discrimination. *Journal of the Acoustical Society of America*, 97, 539-552.
- Thompson, C. L., & Hollien, H. (1970). Some contextual effects on the perception of synthetic vowels. *Language and Speech*, 13, 1-13.
- Thyer, N., Hickson, L., & Dodd, B. (2000). The perceptual magnet effect in Australian English vowels. *Perception & Psychophysics*, 62, 1-20.
- Volaitis, L. E., & Miller, J. L. (1992). Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories. *Journal of the Acoustical Society of America*, 92, 723-735.

VITA

Graduate College
University of Nevada, Las Vegas

Melissa Kay Griffith

Home Address:

7340 West Russell Road
Apt 2029
Las Vegas, NV 89113

Degrees:

Bachelor of Science, Experimental Psychology, 2001
University of South Carolina, Upstate

Special Honors and Awards:

Chancellor's List Award, 2005
Who's Who Among Students in American Universities and Colleges, 2004
Excellence in Psychology Award, 2002
USAA All-American Scholar Collegiate Award, 2001, 2002
National Dean's List Award, 2001, 2002

Publications:

Hall, M. D. & Griffith, M. K. (2004). Attentional interference with musical timbres and vocoder synthesis. *Proceedings of the 8th International Conference on Music Perception & Cognition*, USA.
Hall, M. D., Wieberg, K., & Griffith, M. K. (2003). Analysis of anomalous vocalizations (University of Nevada, Las Vegas Auditory Perception Laboratory Research Report). Available from the National Institute for Discovery Science
Web site: <http://www.nidsci.org/news/acousticanalysis.php>.

Thesis Title: Evaluation of the Perceptual Magnet Effect and Categorical Perception for Musical Timbre

Thesis Examination Committee:

Chairperson, Dr. Karen Kemtes, Ph.D.
Committee Member, Dr. Michael Hall, Ph.D.
Committee Member, Dr. Gretchen Kambe, Ph.D.
Graduate Faculty Representative, Dr. Alice Corkill, Ph.D.