RESEARCH ON
SPOKEN LANGUAGE PROCESSING

Technical Report No. 13

November 30, 2005

Speech Research Laboratory
Department of Psychological and Brain Sciences
Indiana University
Bloomington, Indiana 47405

Supported by:
Department of Health and Human Services
U.S. Public Health Service
National Institutes of Health
Research Grant DC001i1
and
Training Grant DC00012
PERCEPTUAL LEARNING OF SPEECH PROCESSED THROUGH AN ACOUSTIC SIMULATION OF A COCHLEAR IMPLANT

Rose A. Burkholder

Submitted to the faculty of the University Graduate School in partial fulfillment of the requirements for the degree Doctor of Philosophy in the Department of Psychological and Brain Sciences, Indiana University November, 2005
Acknowledgements

This work was funded by NIH-NIDCD research grants R01 DC00064 and DC03937, training grant T32 DC00012, and the American Hearing Research Foundation. Numerous individuals who have helped make this research possible either through their professional or personal support. I have had the privilege of working with and alongside a number of talented graduate students and post-docs in the Speech Research Laboratory at Indiana University. Special thanks are due to Miranda Cleary for her invaluable advice and assistance in the earliest stages of my graduate career. She truly helped me make a smooth and successful transition into the lab and Psychology department.

I would also like to thank the great group of linguists that I’ve shared the lab with these past years: Cynthia Clopper, Caitlin Dillon, Susie Levi, and Steve Winters. Collectively, they have diversified and enriched my academic environment by exposing me to interesting and important research outside of my own field. Individually, they have each been a unique and very enjoyable source of social support as well.

Other past and present members of the Speech Research Lab that I would like to thank include: Kara Kohnen and Jen Karpicke for helping in data analysis running subjects; Darla Sallee for her administrative work; Luis Hernandez, Kuai Hinojosa, and Tim Johnston for maintaining the recording, computing, and testing facilities. Thanks are also due to Charlotte Reed and Lorraine Delhorne of the MIT Research Laboratory of Electronics for providing me with the environmental sound stimuli used in this project.

The present project would also certainly not have been possible without the help of my four fabulous dissertation committee members. I would like to thank Brian O’Donnell for his service to the committee and interest in an area of research quite different from his own. Special thanks are due to Olaf Sporns for his help in guiding me through both qualifying exams and this dissertation and providing assistance to help me advance my academic career.

Mario Svirsky has been especially instrumental in this project. I’d like to specifically thank him and Shivank Sinha for designing the signal processing system used in these studies and making it available to me. I am also grateful for Mario’s willingness to be a mentor to me despite having primary responsibilities in other academic departments and institutions. He has done a number of things to help me move my career beyond this step. The numerous introductions at conferences, opportunity to peer review for a major journal, and recommendations to colleagues on my behalf are greatly appreciated.

Last but not least, a million thanks are due to my dissertation committee chair and mentor David Pisoni. He has done what all other committee members have done and then some. The resources and travel opportunities he has provided have made it remarkable and rewarding to work under his tutelage. I found his enthusiasm and dedication as a mentor very motivating and without those elements this project, prior publications, and my current and future academic positions wouldn’t have been possible.

Finally, I am most thankful for the love and support of my best friend and fiancé Tibor Juhasz. He has encouraged my every pursuit with genuine interest and enthusiasm. I cannot contain the admiration and adoration I have for him in a reasonable amount of space. Simply put, the admiration I have for his work ethic, competitive spirit, character, and charisma has truly made this a better project. More importantly, the unending adoration I have for his heart, sense of humor, and soul has made me a better person.
**TABLE OF CONTENTS**

CHAPTER I: GENERAL INTRODUCTION AND OVERVIEW ............................................. 1

CHAPTER II: INFLUENCES OF SEMANTIC CONTEXT AND FEEDBACK ...................... 6

CHAPTER III: PERCEPTUAL ADAPTATION AND GENERALIZATION OF LEARNING ....... 43

CHAPTER IV: TRANSFER OF LEARNING TO SPEECH AND NONSPEECH TASKS ....... 79

CHAPTER V: SUMMARY AND CONCLUDING REMARKS ...................................... 119

REFERENCES ........................................................................................................ 125

APPENDIX A ........................................................................................................ 136

APPENDIX B ........................................................................................................ 140

APPENDIX C ........................................................................................................ 143
CHAPTER I: GENERAL INTRODUCTION AND OVERVIEW

Cochlear Implants: Implications for Neuroplasticity and Perceptual Learning

Significant amounts of research in the field of cognitive neuroscience have documented the remarkable experience dependent and developmental plasticity of sensory systems such as the auditory system (Bledsoe, Nagase, Miller, & Althuler, 1995; Illing, 2001; Linkenhoker & Knudsen, 2002; Miller & Knudsen, 2003; Palmer, Nelson, & Lindley, 1998; Robertson & Irvine, 1989). Auditory plasticity has most often been considered in the context of auditory deprivation or deafness in animals. However, recently a more complex model of auditory plasticity has been available for experimental investigation in both animals and humans. This model not only considers the effects of auditory deprivation but also the effects of introducing unique electrical stimulation to the auditory system through an auditory prosthetic device known as a cochlear implant (Kral, Hartmann, Tillein, Heid, & Klinke, 2002; Ponton, Don, Eggermont, Waring, Kwong, et al., 1996).

Deafness and cochlear implantation have emerged as important topics in the study of neuroplasticity in recent decades, providing a challenging new paradigm in which to examine the immediate and long term effects of sensory deprivation followed by a unique, artificial form of auditory stimulation. Deafness and cochlear implantation have introduced new and important perspectives to the study of neuroplasticity, because once a cochlear implant alleviates auditory deprivation, it supplies the brain with a new and unusual form of electrical stimuli that is not available under normal circumstances (Dorman & Wilson, 2004). The brain’s response to and ability to learn to understand this unique auditory stimulation is a particularly interesting aspect of cochlear implantation to study.

However, the examination of neuroplasticity in relation to deafness and cochlear implantation has often been skewed, focusing primarily on the effects of the auditory deprivation preceding implantation rather than the effects of implantation itself. Despite this early bias, a great deal of vital information has been obtained about neuroplasticity within auditory cortex by examining the transition from silence to sound in both animal and human cochlear implant users (Giraud, Price, Graham, & Frackowiak, 2001; Giraud, Truy, & Frackowiak, 2001; Kral et al., 2002; Lee, Lee, Oh, Kim, Kim et al., 2001; Ponton et al., 1996). These studies have not only provided needed information to help assess the nature of neuroplasticity in auditory pathways initiated by cochlear implantation but have also provided essential clues about cognitive, behavioral, perceptual learning, and speech perception processes in both adult and pediatric users of cochlear implants.

The ability to perceive speech in one's native language is a robust skill for normal-hearing listeners even under numerous degraded conditions such as sinewave, synthetic, and noise-masked speech (Pisoni, Manous, & Dedina, 1987; Remez, Rubin, Pisoni, & Carrell, 1981; Sumby & Pollack, 1954). Although it is well documented that speech can be recognized under many conditions, it is not necessarily the case that speech is immediately identifiable under such conditions. Rather, periods of adjustment are frequently required for degraded speech to become intelligible. During this period of adjustment, it is presumed that auditory perceptual learning occurs. Perceptual learning is the process of modifying perceptual capabilities to promote adaptation to and successful, detection, discrimination, and interpretation of novel auditory, visual, and/or tactile stimulation (Clark, 2002). Thus, auditory perceptual learning is an important process to examine and understand in order to determine what aspects of the speech signal listeners become attuned to, what higher level cognitive processes or strategies they use when learning to understand degraded speech, and what limitations there may be on learning and plasticity.
Although early examinations of perceptual learning under degraded auditory conditions were largely theoretically motivated and interpreted with respect to their implications for models of speech perception, spoken-word recognition, and general human auditory capabilities (e.g. Duffy & Pisoni, 1992; Sheffert, Pisoni, Fellowes, & Remez, 2002) recent interest in auditory perceptual learning has been fueled by clinical interest in hearing-impaired listeners or deaf individuals using cochlear implants. For most normal-hearing listeners, exposure to severely degraded auditory conditions may occur only on an occasional basis such as when talking on a cellular phone or socializing at a noisy cocktail party. However, for listeners with hearing loss, the typical auditory environment is significantly attenuated and characterized by insufficient or distorted amplitude and frequency information, in addition to competing background noise that normal-hearing listeners also experience. Auditory distortion and degradation is particularly detrimental to hearing-impaired listeners using cochlear implants because the signal their brain receives is based on electrical stimulation rather than acoustic input. Therefore, individuals who receive cochlear implants must learn how to decode and understand a completely novel form of auditory stimulation.

Understanding the methods by which cochlear implant patients learn how to use their devices is important, because it may help explain why there are frequently large and unexplained individual differences in clinical outcomes. Individual differences in cochlear implant users often go unexplained even after relevant clinical and demographic characteristics are accounted for. Along with commonly considered clinical and demographic characteristics such as onset, etiology, and duration of deafness, length of device use, and number of active electrodes, cognitive skills such as learning, memory and attention have also been suggested to play a possible role in the large individual differences observed in cochlear implant patients (e.g. Pisoni, 2000; Pisoni, Cleary, Geers, & Tobey, 2000; Punch, Robbins, Myres, Pope, & Miyamoto, 1987). The methods or mechanisms cochlear implant patients first use to learn to understand the novel auditory signal transmitted from their device could be one of the most influential contributors to individual differences in speech perception and one of the most informative processes for researchers interested in auditory perceptual learning to examine.

Post-lingually deafened cochlear implant users are a particularly interesting group of listeners in which to study auditory perceptual learning because they must adjust to electrical hearing after having been only exposed to acoustic hearing (i.e. normal unamplified hearing or amplified hearing) during most of their lives. However, although they are theoretically an ideal population in which to examine auditory perceptual learning, there are several limitations to using this group of listeners. First, post-lingually deaf cochlear implant users represent a sparse and heterogeneous population making it hard to assemble even small pools of subjects with similar demographic and clinical characteristics, including those that may influence the nature or quality of the signal they hear (i.e. electrode number and placement, processing strategy, number of viable spiral ganglion cells, length and history of hearing impairment). It is also impossible to control cochlear implant users’ daily exposure to speech and sound prior to completing perceptual learning tasks.

To obtain an unadulterated measure of auditory learning for distinctly novel stimuli, such as the input produced by cochlear implants, it is naturally desirable to have access to a sample of participants who are naïve to the stimuli. Although it has been possible to measure speech perception in cochlear implant users immediately after their device has been turned on and prior to exposure to speech (see Svirsky, Silveira, Neuburger, Teoh, & Suarez, 2004) and to measure perceptual learning in cochlear implant users by modifying the settings on their device (Fu, Shannon, & Galvin, 2002), it is not practical or ethical to conduct extended or long-term perceptual learning studies on patients while restricting or controlling their auditory exposure. However, confounds related to prior experience are not as severe
when studying normal-hearing listeners, because it is unlikely that they would hear the experimental stimuli outside the context of an experiment. Therefore, normal-hearing listeners exposed to an acoustic model of what cochlear implant patients experience may reasonably substitute for actual cochlear implant users in studies examining perceptual learning of spectrally degraded speech. Normal-hearing listeners exposed to speech modeled after the input transmitted through a cochlear implant and post-lingually deaf cochlear implant users are similar because both groups of listeners must recognize speech that is unintelligible in comparison to what they ordinarily hear or have previously heard prior to the onset of hearing loss.

**Acoustic Simulations of Cochlear Implants**

Although examination of perceptual learning in actual cochlear implant users can be successfully conducted under some circumstances (Fu et al., 2002; Svirsky et al., 2004), a recent supplement to this research, using normal-hearing individuals, has also provided useful information about how individuals may learn to understand speech processed through a cochlear implant. Acoustic simulations of cochlear implants have enabled researchers to study how normal-hearing listeners adapt to degraded stimuli modeled after the input from cochlear implants and are useful tools to study normal-hearing listeners’ speech perception and perceptual learning skills under unusual sensory conditions in general (e.g., Blamey, Dowell, Tong, & Clark, 1984; Dorman & Loizou, 1997; Dorman, Loizou, & Fiske, 1998; Dorman, Loizou, Kemp, & Kirk, 2000; Dorman, Loizou, & Rainey 1997; Eisenberg, Martinez, Holoweky, Pogorelsky, 2002; Eisenberg, Shannon, Martinez, Wygonski, & Boothroyd, 2000; Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995).

Cochlear implant simulations utilize signal-processing algorithms based on cochlear implant speech processors to spectrally reduce sound into a limited number of channels. Although the first cochlear implants devices utilized only a single channel of spectral information, the current generation of cochlear implants includes multichannel devices that may provide up to 24 full channels of the speech spectrum. Acoustic simulations of cochlear implants emulate how speech is processed into “channels” by filtering the speech signal into adjacent frequency bins that carry spectral information important for speech. In a cochlear implant, the output of each of the frequency bins modulates electrical pulse trains that are directed to electrodes implanted in the cochlea. However, in acoustic models of cochlear implants the bandpass filter outputs modulate noise bands or other carrier signals instead of pulse trains. Thus, many acoustic models of cochlear implants resemble noise vocoder devices originally designed in the late 1930s (Dudley, 1939; Dudley, Reisz, & Watkins, 1939; Klatt, 1987).

In basic acoustic models of cochlear implants, the analysis frequency bins and the noise bands have the same frequency. However, in other acoustic models of cochlear implants, the frequency of the noise bands exceeds the frequency of the bins’ analysis filters (Kaiser & Svirsky, 2000). This mismatch in frequencies causes an upward shift in the sound’s perceived frequency. This manipulation is designed to mimic a basalward frequency shift that may result when the electrodes of a cochlear implant are unable to reach areas deep within the apical end of the cochlea that are tonotopically tuned for low frequencies. Acoustic models of cochlear implants that manipulate frequency shift have been useful to examine perceptual adaptation to different degrees of stimulus degradation in normal-hearing listeners (Fu & Galvin, 2003; Fu & Shannon, 1999a; Rosen, Faulkner, & Wilkinson, 1999; Svirsky, Sinha, Neuburger, & Talavag, 2003).

Because of their clinical applicability to cochlear implant users and versatility of configuration options, acoustic simulations of cochlear implants are useful methods in which to examine perceptual learning in normal-hearing individuals. Using acoustic simulations of cochlear implants in normal-
hearing listeners provides a unique opportunity to examine how normal auditory systems learn to process stimuli similar to the input from cochlear implants. Understanding the perceptual strategies and mechanisms that normal-hearing listeners use to adapt to speech processed through cochlear implant simulations is valuable for several reasons. First, the auditory perceptual learning mechanisms identified in normal-hearing listeners could be compared to learning in cochlear implant patients. Any similarities or differences between these groups may have implications for understanding the consequences of auditory deprivation on neural plasticity and auditory perceptual learning. In addition, comparisons between normal-hearing listeners and cochlear implant patients may provide valuable insight about the enormous individual differences in speech and language outcomes following cochlear implantation.

In order for deaf individuals to learn to use a cochlear implant or for normal-hearing listeners to understand acoustic simulations of cochlear implants, we assume that there is some neural plasticity within the auditory system that supports reorganization and perceptual learning (Eggermont & Ponton, 2003; Ponton, 2001). However, it is not certain whether cochlear implant patients utilize the same neural mechanisms as normal-hearing listeners to learn to understand degraded input from their device. Similarly, it is also unknown whether cochlear implants patients and normal-hearing listeners utilize similar behavioral strategies when learning to perceive degraded speech processed by a cochlear implant. Therefore, exposing normal-hearing listeners to acoustic simulations of cochlear implants is a valuable research tool, because it enables comparisons of learning between listeners with and without a history of auditory deprivation.

A second benefit of using acoustic simulations of cochlear implants to examine perceptual learning in normal-hearing adults is related to the current lack of a routine or universal method of rehabilitation for newly implanted patients. It is currently unclear what training methods are best for teaching or training cochlear implant patients how to use their device. Therefore, research evaluating different training and feedback methods in normal-hearing listeners exposed to acoustic simulations of cochlear implants could provide new insights into how more successful and efficient treatments and rehabilitative programs could be designed for cochlear implant patients.

Similarly, examining the robustness of auditory learning in listeners exposed to an acoustic simulation of a cochlear implant could also provide insight into what training methods and materials would be the most beneficial and efficient for cochlear implant users to use. Generalization and transfer of learning are two important processes to assess in order to determine how flexible and robust perceptual learning is (Karmarkar & Buonomano, 2003; Moore, Amitay, & Hawkey, 2003). Generalization of learning shows that the benefits of training are not restricted to the specific tokens on which listeners were trained. Rather, generalization indicates that perceptual learning was robust enough to extend to novel training-stimulus tokens or to novel stimuli taken from a novel stimulus category. For example, if listeners are able to identify novel sentences after being trained only on other sentences, generalization to new training-stimulus tokens is considered to have occurred. Similarly, if listeners are able to accurately identify isolated words after being trained only on sentences, generalization or "carry over" to a novel stimulus category is considered to have occurred.

Evidence of transfer of learning suggests that listeners have flexibility enabling them to extend what they have learned to entirely new tasks on which they were not trained. For example, Nygaard, Sommers, and Pisoni (1994) demonstrated transfer of training effects in a novel voice learning experiment. They trained listeners to identify novel voices by name and then asked them to identify noise-masked words spoken by either the familiar or unfamiliar voices that were counterbalanced among listeners to control for differences in speaker intelligibility. Listeners identified more novel words spoken
by the familiar voices, demonstrating that what they learned in the initial voice identification training transferred to a spoken word recognition task.

Taken together, generalization and transfer are two aspects related to perceptual learning that are important to examine in studies measuring listeners’ perceptual learning and adjustment to speech processed through acoustic simulations of cochlear implants. Determining normal-hearing listeners’ ability to generalize learning to new stimuli and signal processing strategies and to transfer learning to entirely different tasks may also provide insight into what training methods and materials would be the most beneficial to cochlear implant users. Unfortunately, previous perceptual learning studies using cochlear implant simulations have not adequately measured generalization and transfer of learning (Davis, Johnsrude, Hervais-Adelman, Taylor, & McGettigan, 2005; Rosen et al., 1999).

Taking these current gaps in the auditory perceptual learning research into consideration, several novel perceptual learning experiments utilizing an acoustic simulation of a cochlear implant in normal-hearing adults are reported here. Taken together, these experiments were designed to: 1) evaluate the effects of the amount of exposure and the quality of feedback on the magnitude of perceptual learning, 2) provide a clearer picture about semantic and contextual influences on perceptual learning of speech processed through an acoustic simulation of a cochlear implant, 3) determine the degree to which perceptual learning generalizes to new stimulus materials and new methods of signal degradation and finally, 4) assess the robustness of perceptual learning through transfer tasks designed to measure environmental sound and gender identification and speaker discrimination.

The results from the present set of studies have important theoretical implications for understanding how listeners learn to perceive speech under highly degraded conditions, for what they learn, and for what limitations there may be to learning and plasticity in general. An additional benefit of the studies reported here is that the findings from this research have some direct clinical application to actual cochlear implant users who often struggle when initially learning how to recognize speech and sounds through their cochlear implants. More specifically, the research reported in the following chapters provides some new insights into what training and feedback methods and materials and transfer tasks may be the most beneficial for cochlear implant users to be trained with and exposed to immediately after receiving their implant.

In the following chapters, three separate experiments examining auditory perceptual learning in normal-hearing adults listening to an acoustic simulation of a cochlea implant are reported. Chapter II reports the results from a study that investigated the effects of feedback and the semantic context of training stimuli on the perceptual learning of normal-hearing listeners exposed to an acoustic simulation of a cochlear implant. Chapter III reports the results of an experiment designed to examine listeners’ abilities to generalize auditory perceptual learning to frequency shifts in the acoustic model of the cochlear implant. In Chapter IV, the robustness of auditory perceptual learning is explored further in an experiment that assessed the transfer of speech identification training to environmental sound identification, gender identification, and speaker discrimination tasks. Finally, in Chapter V, a summary and interpretation of the results of all three studies is presented and the overall clinical and theoretical implications of this line of research are discussed.
CHAPTER II: INFLUENCES OF SEMANTIC CONTEXT AND FEEDBACK

Introduction

Variability in Speech Perception Abilities and Strategies in Cochlear Implant Users

Over the past quarter century, cochlear implants have developed as a successful treatment for many profoundly and severely deaf adults and children who derive little to no benefit from hearing aids (e.g. Eddington, Dobelle, Brackmann, Mladejovsky, & Parkin, 1977; House. Berliner, & Eisenberg, 1981; Michelson, 1971). However, the amount of success cochlear implant patients have with their device frequently depends on a variety of clinical and demographic characteristics such as onset, duration, degree, and etiology of deafness, electrode placement, and perceptual detection and discrimination skills (Blamey, Pyman, Gordon, Clark, Brown, Dowell et al., 1992; Donaldson & Nelson, 2000). Just as the success with these devices varies among sub-groups of patients who differ according to one or more of these criteria (i.e. pre-lingually vs. post-lingually deaf adults), the post-implant benefits observed among patients who have similar clinical histories often vary considerably as well.

Because a large degree of the individual differences between cochlear implant patients typically goes unexplained, even after accounting for obvious clinical and demographic variables, clinicians and researchers are often left to contemplate what other factors may be contributing to variability in a variety of outcome measures (Kirk, 2000). Several outcome measures in which individual differences are frequently observed among cochlear implant users include speech perception (Gantz, Tyler, Woodworth, Tye-Murray, & Fryauf-Bertschy, 1994; Sarant, Blamey, Dowell, Clark, & Gibson, 2001) and production (Kirk, Diefendorf, Riley, & Osberger, 1995; Miyamoto, Kirk, Robbins, Todd, Riley, et al., 1997), spoken word recognition (Collison, Munson, & Carney, 2004; Pisoni, Cleary, Geers, Tobey, 2000; Waltzman, Cohen, Gomolin, Green, Shapiro, et al., 1997), language development (Svirsky, Robbins, Kirk, Pisoni, & Miyamoto, 2000), speaker (Cleary, 2003; Cleary & Pisoni, 2002) and voice-gender discrimination (Fu, Chinchilla, & Galvin, 2004), music appreciation and perception (Gfeller & Lansing, 1991; 1992; Gfeller, Christ, Knutson, Witt, & Mehr, 2003; Stordahl, 2002), and environmental sound identification (Reed & Delhorne, 2005). Although some of the areas of performance in which cochlear implant patients differ are rather specific, it is unknown whether the causes of variability are domain specific as well or if differences are related to more general cognitive and linguistic variables. In addition, it is currently unknown whether variations in the rehabilitative strategies used with cochlear implant users contribute substantially to the progress that they make following cochlear implantation.

The present approach to determining what additional factors may contribute to the individual differences in cochlear implant performance has focused primarily on uncovering possible relationships between general cognitive and linguistic factors and speech perception skills (e.g. Collison et al., 2004; Gantz, Woodworth, Abbas, Knudson, & Tyler, 1993; Lyxell, Andersson, Andersson, Arlinger, Bredberg, et al., 1998; Pisoni et al., 2000; Punch, Robbins, Myres, Pope, & Miyamoto, 1987). In the handful of studies that have examined cognitive and linguistic correlates of outcome measures with cochlear implants, a wide variety of cognitive variables have been examined, including IQ, expressive vocabulary, language comprehension, working memory, and nonverbal intelligence. In an early study by Punch and colleagues (1987), IQ was identified as a significant predictor of adult cochlear implant users' speech perception scores. Reading-span, which is a measure of working memory, has also been found to be related to cochlear implant users' speech perception abilities (Lyxell et al., 1998). However, in a more recent study by Collison and colleagues (2004) nonverbal intelligence, expressive vocabulary, and
language comprehension were not related to the spoken-word recognition scores of a small, heterogeneous group of adult cochlear implant users.

The inconsistencies and difficulties in identifying cognitive process variables that predict an adult’s success with a cochlear implant make it hard to draw any conclusions about whether or not individual differences in cognitive and/or linguistic competence influence the performance outcomes of cochlear implant users. In addition, these inconsistencies make it difficult to determine what other unexplored variables should also be examined for possible relationships with cochlear implant outcome measures. Another source of variability in cochlear implant users’ outcome performance that has not been fully explored is the perceptual learning strategies that patients use to understand their new form of auditory input. Because there are currently few well established or universal methods of training cochlear implant patients to use their implant, listeners are often left on their own when trying to make sense of the new auditory signals they hear outside of a rehabilitative setting (Brown, Dowell, Martin, & Mecklenburg, 1990; Clark, 2003; McConkey-Robbins, 2000). Therefore, much of the unexplained variability in cochlear implant users’ performance may be related to the auditory perceptual learning process and the cognitive and linguistic skills that they specifically use during the auditory perceptual learning process.

Cognitive Correlates of Speech Perception in Normal-Hearing Populations

To understand how perceptual learning and other cognitive and linguistic variables may contribute to speech perception, spoken word recognition, and other auditory skills of cochlear implant users, it is important to first consider how these processes operate and interact in normal-hearing listeners. Individual differences in speech perception (Suprenant & Watson, 2001; Watson, Qui, Chamberlain, & Li, 1996), spoken word recognition (Marslen-Wilson & Warren, 1994), and sound identification (Gygi, Kidd, Watson, 2004) have all been documented in normal-hearing listeners. Several cognitive and linguistic variables such as working and long-term memory (Cowan, 1996; Gupta, 2003; Pisoni, 1993), auditory-visual integration skills (Lachs, 2002; Vatikiotes-Bateson, Eigsti, Yano, & Munhall, 1998; Watson et al., 1996), vocabulary size (Pisoni, Nusbaum, Luce, & Slowiaczek, 1985), and lexical knowledge (Norris, McQueen, & Cutler, 2003; Pisoni et al., 1985) have been suggested to partially account for individual differences in speech perception and language processing in normal-hearing listeners.

In addition, auditory training under normal listening conditions has been found to result in perceptual learning that modifies normal-hearing listeners’ general auditory (Leek & Watson, 1988) and speech perception abilities (Logan, Lively, & Pisoni, 1991; Tremblay, Kraus, Carrell, McGee, 1997; Wang, Spence, Jongman, & Sereno, 1999; Wayland & Guion, 2004). For instance, Logan and colleagues (1991) demonstrated that Japanese listeners, who are frequently unable to discriminate or identify the English phonemes /r/ and /l/, can be trained to perceive these contrasts. The Japanese listeners’ improvement in identifying /r/ and /l/ after training provides evidence that auditory perceptual learning has modified speech perception abilities. In a follow-up study, it was also determined that perceptual learning by these Japanese listeners was so robust that it even carried over to their own productions of the phonemes /r/ and /l/ (Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997).

Taken together, evidence of cognitive, linguistic, and training effects on speech perception in normal-hearing listeners, even in clear auditory conditions, suggests that similar variables may be important to investigate in cochlear implant users. More thorough examinations of cognitive, linguistic, and training effects on cochlear implant users’ abilities to learn to understand speech through their device may help account for the individual differences that are frequently observed in this group of individuals.
and may lead to advances in the rehabilitative strategies that are available to them. Studies considering cochlear implant users' abilities to adapt to their novel form of auditory stimulation may also contribute to the current body of research on speech perception and auditory perceptual learning in general.

**Speech Perception in Degraded Auditory Conditions: Simulating Cochlear Implants**

Although examining speech perception, spoken word recognition, and auditory perceptual learning in normal-hearing listeners can conceivably be done under clear listening conditions, it is much more appropriate to examine this topic in normal-hearing listeners exposed to degraded auditory conditions. One obvious benefit of assessing speech perception and perceptual learning in degraded listening conditions is that ceiling effects, which drastically reduce variability in normal listening conditions, can be avoided. An additional benefit of measuring speech perception and auditory perceptual learning skills of normal-hearing listeners in degraded auditory conditions is that the nature and amount of signal degradation can be adjusted to emulate the auditory conditions that some listeners are exposed to naturally because of hearing-loss or use of a cochlear implant. For example, digital signal processing techniques have been used to model various degrees of hearing loss in normal-hearing listeners (Humes, Espinoza-Varas, & Watson, 1988). Similarly, a variety of methods to spectrally degrade speech have recently been developed in order to simulate the auditory input relayed to cochlear implant users through their devices (Blamey, Dowell, Tong, & Clark, 1984; Dorman & Loizou, 1997; Dorman, Loizou, & Rainey, 1997; Fu & Shannon, 1999; Kaiser & Svirsky, 2000; Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995).

Acoustic simulations of cochlear implants have recently been employed as research tools in studies with normal-hearing individuals to address several questions relevant to cochlear implant users and speech perception and auditory perceptual learning in general. First, acoustic simulations of cochlear implants have been used to verify the effects of spectral degradation on speech perception. Acoustic models that utilize fewer channels and thus provide less spectral information are harder for listeners to understand (Dorman & Loizou, 1997). Similarly, models that implement a large frequency shift in the speech signal are also very difficult for listeners to perceive (Fu & Shannon, 1999; Rosen, Faulkner, & Wilkinson, 1999). These studies demonstrate the basic limitations on the auditory system's ability to identify severely degraded and spectrally mismatched speech.

In addition to examining variables directly related to the nature and amount of signal degradation, speech perception studies using acoustic simulations of cochlear implants have also examined cognitive correlates of normal-hearing listeners' abilities to perceive the degraded speech (Chiu, Eng, Strange, Yampolsky, & Waters, 2002; Eisenberg, Shannon, Martinez, Wygonski, & Boothroyd, 2002). Investigations of what training methods may be the best for enabling listeners to assimilate to speech processed though an acoustic simulation of a cochlear implant have also been conducted (Chiu & McCabe, submitted; Davis, Johnsrude, Hervais-Adelman, Taylor, & McGlone, 2005; McCabe & Chiu, 2003). Finally, the ability to perceive speech processed through acoustic simulations of cochlear implants has also been linked to the lexical and contextual information contained in the stimuli (Davis et al., 2005; Dorman & Loizou, 1997; Eisenberg, Martinez, Holoweczy, & Pogorelsky, 2002; Rosen et al., 1999; Shannon, Fu, & Galvin, 2004).

In order to fully understand what strategies listeners may use when learning to perceive speech processed through a real or simulated cochlear implant, it is important to first consider the cognitive processes that are used to support speech perception in clear conditions and in other degraded auditory conditions such as noise. It has been suggested that there are specific top-down processing strategies used for speech perception that depend on linguistic knowledge (Nittouer & Boothroyd, 1990) and
phonological and syntactic rules (Miller & Isard, 1963; Norris, McQueen, & Cutler, 2003). The phonological rules of a language are the most fundamental influence on speech perception and spoken word recognition. However, lexical knowledge based on properties such as word frequency and familiarity and neighborhood density (Luce & Pisoni, 1998; Meyer & Pisoni, 1999) also influence speech perception at the phonemic level. Therefore, knowledge of language and perception of individual words can help determine what sound pattern a listener reports hearing, particularly in noisy conditions.

A second mechanism of top-down processing in speech perception involves the use of sentence context and semantic information to determine what word or words may have been heard (Miller & Isard, 1963). It is clearly evident why both of these strategies would be useful for parsing speech that is degraded by noise or transformed in another manner. Simply put, lexical knowledge and semantic cues often provide enough information to compensate for or ameliorate the detrimental effects of degraded auditory input and allow listeners to fill in the empty gaps in the degraded speech signal they are hearing.

Cognitive Strategies for Perceiving Degraded Speech: Lexical Knowledge

Word familiarity (Miller, Heise, & Lichten, 1951) and word frequency (Howes, 1957), which ultimately reflect a listener’s lexical knowledge, have been identified as important lexical properties that contribute to speech perception in noise. These early studies suggested that the lexical knowledge and lexical patterns within language facilitate speech perception under degraded auditory conditions. These findings were some of the earliest to suggest that the lexical knowledge of individuals and the lexical patterns within their language were both intimately linked to spoken word recognition performance in degraded auditory conditions.

Recently, the interest in lexical influences on speech perception has been extended to research using cochlear implant simulations with normal-hearing listeners. Eisenberg and her colleagues (2002) found that normal-hearing children were better at identifying lexically easy words in isolation and in sentences than lexically hard words when the stimuli were processed through a cochlear implant simulation. The results of this study replicated previous findings obtained during the development of the Lexical Neighborhood Test which defined and identified lexically easy and hard words (Kirk, Pisoni, & Osberger, 1995; Kirk, Eisenberg, Martinez, & Hay-McCutcheon, 1999; Luce & Pisoni, 1998). Lexically easy words are words that are frequently used in language and come from sparse lexical neighborhoods where there is little competition from other phonetically similar words. Therefore, there are few words that could be confused with lexically easy words. In contrast, lexically hard words have low familiarity and usage but many more confusable neighbors.

The study by Eisenberg and her colleagues (2002) was particularly interesting because the normal-hearing children they examined performed very similarly to deaf children who use cochlear implants. The pediatric cochlear implant users in their study also identified lexically easy words more frequently than lexically hard words. This is a noteworthy result, because it suggests that the nature of the signal degradation was sufficient to elicit similar patterns of behavioral responses in deaf children using cochlear implants and their normal-hearing peers.

However, not all children using cochlear implants identified lexically easy words more accurately than lexically hard words when they were embedded in sentences. Pediatric cochlear implant patients that performed poorly on other clinical speech perception and language processing tasks did not appear to use their lexical knowledge as effectively as high performing cochlear implant users and their normal-hearing peers. The authors suggested that the low performing children may not have been as good at using sentence context to aid in the identification of words because of linguistic and/or cognitive
constraints. The results of this study suggest that linguistic variables and speech perception are related in pediatric cochlear implant users and that lexical knowledge may contribute to the individual differences that are observed in this group of listeners and, perhaps, in adult listeners as well.

Several studies have examined lexical influences on adult cochlear implant users’ speech perception skills (Collison et al., 2004; Kaiser, Kirk, Lachs, & Pisoni, 2003). Kaiser and colleagues (2003) found that postlingually deafened adults with cochlear implants identified lexically easier words more accurately than lexically hard words in an audiovisual word recognition task. However, in a recent study exploring the relationship between adult cochlear implant users’ linguistic skills and word recognition abilities, it was determined that cochlear implant users were less sensitive to the neighborhood density of words when completing a gated-word recognition task than normal-hearing adults (Collison et al., 2004).

This result is not consistent with what was found previously in comparisons among pediatric cochlear implant users and their normal-hearing peers (Eisenberg et al., 2002) and in the study testing audiovisual word recognition skills of adult cochlear implant users (Kaiser et al., 2003). Because of the current inconsistencies in determining how lexical influences contribute to cochlear implant users’ speech perception, additional research focusing on how lexical knowledge and other top-down processes affect speech perception in this population should be conducted. Similarly, research using normal-hearing listeners exposed to speech transformed through acoustic simulations of cochlear implants should provide new knowledge about the influence of lexical information on cochlear implant users’ speech perception abilities.

Speech Perception and Perceptual Learning without Lexical Representations

Although there is a great deal of evidence documenting the influence of lexical knowledge on normal-hearing listeners’ speech perception skills in adverse auditory conditions (Miller et al., 1951; Howes, 1957), few studies have actually attempted to measure speech perception in degraded auditory conditions using stimuli without lexical representations. Nonword repetition techniques have recently been used to examine speech perception and phonological working memory skills in the absence of lexical knowledge in normal-hearing listeners in clear auditory conditions (Burkholder & Pisoni, 2003; Gathercole, Willis, Baddeley, & Emslie, 1994; Gupta, 2003; Metsala, 1999), in hearing-impaired listeners who use cochlear implants (Cleary, Dillon, & Pisoni, 2002; Dillon, Burkholder, Cleary, & Pisoni, 2004; Willstedt-Svensson, Lofqvist, Almqvist, & Sahlen, 2004), and in normal-hearing listeners exposed to speech processed through an acoustic simulation of a cochlear implant (Burkholder, Pisoni, & Svirskey, 2004).

Several studies have found that repeating nonwords is an extremely difficult task for many prelingually deaf pediatric cochlear implant users (Cleary et al., 2002; Dillon et al., 2004). Dillon and colleagues (2004) determined that if pediatric cochlear implant users’ nonword repetitions were scored strictly as correct or incorrect, only 5% of their responses would be completely correct. Even when judged on a 7-point ratings scale, some pediatric cochlear implant users perform near floor on the task (Cleary et al., 2002; Dillon et al., 2004). In addition, other pediatric cochlear implant users cannot even provide enough nonword repetition responses to be included in any meaningful statistical analysis. Taken together, these examples illustrate the difficulty of completing a complex speech perception task when the test stimuli lack lexical representation in long-term memory.

Nonword repetition is a very difficult task because top-down lexical knowledge cannot be used reliably to complete any of the subcomponent processes involved in the task. After hearing a nonword,
listeners must rely primarily on their ability to encode fine acoustic-phonetic detail using bottom-up processing and sub-lexical knowledge. Listeners must then hold this acoustic-phonetic detail in short-term memory by rehearsing it in the phonological loop of working memory. Finally, listeners must reassemble the novel phonological sequence into a fluent verbal response. Accomplishing this sequence of events with nonword patterns has proven to be quite difficult for pre-lingually deaf children using cochlear implants.

Repeating nonwords under severely degraded conditions is also very hard for normal-hearing listeners exposed to a cochlear implant simulation. In a pilot study using an 8-channel model with a frequency shift designed to mimic a 6.5mm basalward shift on the cochlea, Burkholder and colleagues (2004) found that normal-hearing adult listeners also performed near floor when asked to repeat the same nonwords as pediatric cochlear implant users. This result is important when considered in comparison to the pediatric cochlear implant users’ nonword repetition scores because it suggests that it is unlikely that the children performed so poorly because of speech production problems. Rather, their poor performance was most likely related to the degraded auditory input and the lack of lexical representations in the stimuli. However, a drawback of comparing normal-hearing adults’ and pediatric cochlear implant users’ nonword repetition performances is that the normal-hearing listeners had relatively little exposure to the degraded speech in comparison to the children using cochlear implants. Perhaps if normal-hearing adults were given more exposure to and training with the processed speech they would have performed better even on a nonword repetition task.

Preliminary support for this idea was provided by an additional finding in the pilot study (Burkholder et al., 2004). After a brief period of open- and closed-set word recognition training, the nonword repetition scores of the normal-hearing adults, derived from a new set of nonwords, improved. This result is theoretically important, because it suggests that training on familiar words processed through a cochlear implant simulator may generalize to novel nonwords. Thus, even with minimal lexical context, some fine acoustic-phonetic and sub-lexical detail may be learned in the laboratory in a short period of time.

In addition, even though perceptual learning through word training relied heavily on lexical knowledge, it may be generalizable to a speech perception task that requires more bottom-up processing driven primarily by the acoustic-phonetic signal and sub-lexical knowledge. Although the finding that listeners’ nonword repetitions were more accurate after training suggests that perceptual learning occurred, the possibility that a large part of this improvement is due to procedural learning cannot be ruled out completely. Listeners’ nonword repetition scores may have improved simply because they became more adept at generating fluent nonword responses. In such a case, procedural rather than perceptual learning was likely to have occurred. In order to determine more conclusively whether auditory perceptual learning occurs when stimuli lack lexical representations, studies that specifically manipulate the lexical properties of the training stimuli should be conducted.

Recently, several studies examining the role of training and lexical representations on learning how to perceive speech processed through a cochlear implant simulation have been conducted by Davis and his colleagues (Davis et al., 2005; Hervais-Adelman, Carlyon, Davis, & Johnsrude, 2004). These authors examined whether training stimuli were required to have lexical representations in order for normal-hearing listeners to learn how to perceive speech processed through an acoustic simulation of a cochlear implant. To answer this question, they trained separate groups of listeners with words or nonwords. In addition, they trained separate listener groups with meaningful sentences, sentence length stimuli comprised entirely of nonwords (i.e. Cho tekeen garund pid ga sumeeun), and sentences in which
nonwords were embedded among common function words to retain syntactic structure (i.e. *The teken agund to the simeem*).

During training, the participants listened passively to the stimuli and then received feedback in the form of the unprocessed meaningful or nonword sentence followed by an additional repetition of the degraded stimulus. The assessment of learning was based on the listeners' abilities to transcribe meaningful sentences or words that were similar in length and complexity to the training stimuli. Results indicated that nonword training was not as successful as word training (Hervais-Adelman et al., 2004). In addition, although a group of listeners trained on nonword sentences improved throughout the sentence transcription session, their performance started below that of the listeners trained on meaningful sentences and sentences in which nonwords were embedded among the closed-class function words. In fact, their initial performance was only comparable to the performance of naïve listeners who did not receive any form of training or exposure to the degraded speech in a previous experiment (Davis et al., 2005).

These results indicate that training on nonwords or nonword "sentences" provided little if any benefit to listeners when they were later required to transcribe real words or meaningful sentences (Davis et al., 2005; Hervais-Adelman et al., 2004). Based on these findings, it appears that normal syntactic sentence structure and lexical representations of at least the closed-class function words in the training sentences may be required to obtain any improvement in speech perception scores. This result demonstrates that removing information needed for top-down processing of degraded speech is not only detrimental to the ability to understand the spoken message but also detrimental to the ability to learn how to better understand degraded speech and generalize learning to new stimuli.

Cognitive Strategies for Perceiving Degraded Speech: Semantic Predictability

An alternative way to measure auditory perceptual learning in the absence of sufficient top-down cues is to use sentences that have intact lexical information but lack semantic meaning. In addition to lexical effects, there are robust contextual influences on speech perception for speech presented in degraded and unnatural auditory conditions or articulated by individuals who have speech intelligibility problems. It has been determined that the perception of words embedded within sentences masked by noise is highly related to how predictable words are based on preceding sentence context (Kalikow, Stevens, & Elliott, 1977). The ability to accurately complete a sentence verification task in natural and synthetic speech conditions is also related to the predictability of sentence context (Pisoni, Manous, & Dedina, 1987). In addition, listeners report higher intelligibility scores for the speech of normal-hearing and deaf children when there are more context cues available in the speech stimuli (McGarr, 1981).

Previous research suggests that even listeners as young as 3- to 5-years-old benefit from the use of linguistic context and predictability when the auditory signal is degraded (Fallon, Trehub, & Schneider, 2002; Nittroer & Bootbryd, 1990). In addition, several studies have reported that both adults and children are sensitive to contextual information when listening to acoustic simulations of cochlear implants (Burkholder, 2004; Dorman & Loizou, 1997; Eisenberg et al., 2002; Rosen et al., 1999; Shannon et al., 2004). These studies have shown that word recognition scores are consistently better when words are embedded in sentences rather than presented in isolation (Dorman & Loizou, 1997; Eisenberg et al., 2002; Rosen et al., 1999; Shannon et al., 2004). These findings suggest that sentence context aids in identifying words and sentences processed through an acoustic simulation of a cochlear implant.
In addition, the semantic predictability of sentences has also been found to be related to listeners’ abilities to perceive speech processed through an acoustic simulation of a cochlear implant (Burkholder, 2004) or through similar noise-band filtering techniques (Stickney & Assmann, 2001). Both studies used sentences varying in final word predictability to examine specifically how semantic context influenced normal-hearing listeners’ abilities to identify degraded speech. Semantic predictability of the stimuli used in these studies depended on how predictable a sentence’s last word was based on earlier words. Results indicated that listeners were more accurate in identifying the last word of sentences with high rather than low predictability. These findings suggest that sentence context clues that are derived in a top-down manner are critical to decoding a spoken message presented in degraded auditory conditions that mimic input from cochlear implants.

Semantic context of sentences is known to influence speech perception so heavily that, in order to obtain measures of speech perception that are unconfounded by listeners’ abilities to guess words based on context, researchers have used stimuli with limited or no semantic predictability. Stimuli that completely lack semantic predictability and meaning but that are still grammatically correct are referred to as anomalous sentences or syntactic prose (i.e. The deep buckle walked the old crow) (Herman & Pisoni, 2003; Malgady & Johnson, 1977; Marslen-Wilson, 1985). Anomalous sentences enable examinations of speech perception and auditory perceptual learning under conditions in which important top-down semantic information has been blocked or eliminated.

Although lexical knowledge can still be used to identify individual words in anomalous sentences, the words can no longer be identified by relying on sentence context. Rather, anomalous sentence identification relies more on bottom-up processing based on acoustic-phonetic, sub-lexical information, and lexical information in the absence of normal semantic support. Because of the lack of appropriate semantic support, anomalous sentences are difficult to identify in both clear and noisy auditory conditions (Miller & Isard, 1963). In addition, it has also been demonstrated that it is difficult to correctly recall (Marks & Miller, 1964) and recognize (Malgady & Johnson, 1977) anomalous sentences or syntactic prose.

However, Davis and colleagues (2005) have recently suggested that anomalous sentences may be sufficient to elicit perceptual learning of speech processed through an acoustic simulation of a cochlear implant. Listeners who were passively exposed to syntactic prose and then given feedback revealing what they heard performed significantly better during the first and last half of a short test phase than naive listeners and listeners exposed to sentences composed entirely of nonwords or nonwords embedded grammatically among function words. In addition, listeners that heard syntactic prose performed as well as listeners trained on meaningful sentences. This result suggests that perceptual learning for degraded auditory stimuli, such as those generated by an acoustic simulation of a cochlear implant, relies primarily on lexical rather than syntactic information and semantic predictability.

Although Davis and colleagues (2005) suggested that perceptual learning may occur even in the absence of semantic coherence, the methods used in their study may have been inappropriate to confirm this accurately. The amount of learning that occurred within separate listener groups was not adequately assessed in this study for several reasons. First, no learning data were collected during the first period of exposure and training with syntactic prose and sentences. Listeners were never required to identify any of the sentences that they were exposed to during the training period. Thus, no firm conclusions can be made about whether the listeners were really learning during the period that they heard the training stimuli.
Secondly, the assumption that the listeners trained on syntactic prose had learned during the training period was based simply on their word identification scores for the first and second half of the sentences heard during the test period. However, feedback indicating the identity of the test sentences was also provided during this test period. Therefore, this feedback, rather than the training on sentences and syntactic prose, may have been the actual cause of the learning observed in the test period.

In addition, in the study by Davis and his colleagues (2005) that examined lexical influences on perceptual learning with an acoustic simulation of a cochlear implant, no data were collected to determine if the listeners’ performance was due to group differences that existed prior to training and testing. In addition to not collecting learning data during training, pre-test data were not collected to ensure that the groups began the experiment with similar abilities to perceive speech through an acoustic simulation of a cochlear implant. Therefore, it is uncertain if some training groups performed better during the test phase simply because the individuals randomly assigned to the group were more skilled in identifying the speech from the start. Finally, the training and test phases of the study were conducted using only 20 sentences apiece. Such brief periods of training and testing may not be sufficient to conclusively determine the impact of semantic context on perceptual learning for speech processed through an acoustic simulation of a cochlear implant.

Davis and colleagues’ (2005) experimental design also limits the assessment of the overall robustness of auditory perceptual learning gained through passive training on anomalous sentences. Based on their accuracy in identifying meaningful sentences during the test phase, it was presumed that listeners who heard syntactic prose during passive training had learned something about the degraded speech signal. However, in this test phase, both the category of stimuli used (meaningful rather than anomalous sentences) and the task (reporting what was heard rather than simply listening passively) were different. Thus, for listeners trained on anomalous sentences, generalization of perceptual learning to a new stimulus category, not perceptual learning overall, was the variable that was assessed.

In contrast, for listeners trained on meaningful sentences, only the task changed. No attempt was made to test this group’s ability to generalize their perceptual learning to a new set of stimuli such as anomalous sentences. A comparison of the listener groups’ ability to generalize learning to new categories of stimuli would be useful to determine if the robustness of the learning between these two groups is really equal. Similarly, no measure of retention of learning or effects of stimulus repetition was tested in this study (Davis et al., 2005).

A more comprehensive examination of perceptual learning of degraded speech should also include tests of whether listeners trained on anomalous sentences could learn to identify specific tokens on which they were trained. Similarly listeners trained on anomalous sentences should also be asked to identify novel anomalous sentences during the test phase. This measure may help determine if these listeners can generalize their learning to novel training tokens. In addition, generalization and short-term retention of learning should also be assessed in the training group who received meaningful sentences to provide a more accurate comparison of the quality of learning achieved by each group.

Taking the current shortcomings of research examining perceptual learning in the absence of semantic context into consideration, the present study was designed to achieve several goals. The primary goal of this research was to determine the effects that training with semantically anomalous sentences has on the perceptual learning of speech processed through an acoustic simulation of a cochlear implant. To more accurately and thoroughly assess perceptual learning in the absence of semantic context, comparisons of the actual course of learning during training periods using meaningful and anomalous
sentence were made. In addition, generalization and short-term retention of learning were also compared in these groups.

Determining whether lexical knowledge and semantic context or "top-down" knowledge may facilitate or even be necessary for training with an acoustic simulation of a cochlear implant is important because the answer may have implications for what strategies listeners rely on when faced with degraded and unusual auditory conditions. Although previous research strongly suggests that lexical and contextual support play an important role in deciphering degraded speech signals, few studies have adequately examined how perceptual learning of degraded speech is influenced when lexical meaning and context are removed by using nonwords and anomalous sentences as training stimuli (Davis et al., 2005; Hervais-Adelman et al., 2004). Carrying out these manipulations is not only useful in order to expand upon general theories of speech perception but may also have some direct clinical applications to cochlear implant users and other listeners who have hearing, speech, or language impairments.

There is increasing interest within the speech and hearing sciences in determining what therapeutic or training materials are most effective in treating patients with speech and/or hearing impairments. In fact, recent findings have suggested that phonological disorders in children may be more effectively and rapidly treated by using nonword training materials (Gierut, 1998; Morissette & Gierut, 2002). Presumably, this type of treatment is successful because children learn to modify their phonological inventories independently of their lexical representations by focusing perceptual learning on sub-lexical information in speech. Training without lexical representations prevents children from confining their learning to simply the words they were trained on or the words that they already know. Rather, by training children on unfamiliar phonological patterns that have no lexical representation, they can more easily and quickly generalize sub-lexical knowledge to different phonological environments.

It is also important to explore whether stimuli lacking semantic meaning would also be useful for training normal-hearing listeners to understand speech modeled after the input received by profoundly deaf individuals using cochlear implants. By using training materials that lack semantic information, such as semantically anomalous sentences, listeners may be forced to rely more on lower-level, acoustic-phonetic and sub-lexical phonological information to identify speech during training rather than top-down processing based on sentence context. It is possible that focusing attention and perceptual learning on acoustic-phonetic details during training may lead to an enhanced ability to identify speech processed through an acoustic simulation of a cochlear implant after training. Thus, although performance during a training period using semantically anomalous sentences may be poorer than training with meaningful sentences, speech perception performance after training with anomalous sentences may possibly exceed traditional training.

Previous research suggests that the amount of learning acquired while listening to anomalous sentences should at least be equivalent to that of meaningful sentences (Davis et al., 2005). However, this conclusion was made based on a very brief period of training and testing and insufficient learning data. Therefore, the effects of using anomalous sentences to train listeners to understand speech processed through an acoustic simulation of a cochlear implant could potentially be more robust. By obtaining actual learning and speech perception data during longer training and testing periods, these hypotheses can be more adequately tested. In addition, the robustness of perceptual learning acquired through training on anomalous sentences can more accurately be tested in the present study by examining generalization of learning to new training tokens and assessing short-term retention of learning.
Learning to Perceive Degraded Speech: Effects of Training and Feedback

In addition to determining whether the presence of top-down information in training materials can influence the quantity and quality of auditory perceptual learning, it is also important to determine what training or feedback methods are most successful for teaching listeners to identify speech processed through acoustic simulations of cochlear implants. Unlike individuals who are implanted with artificial hips or other prostheses, cochlear implant recipients, particularly post-lingually deaf adults, rarely receive long-term and systematic rehabilitative treatment to improve speech perception (Brown et al., 1990; Clark, 2003; McConkey-Robbins, 2000) or other auditory skills such as music appreciation (Gfeller, 2001; Gfeller, Mehr, & Witt, 2001; Gfeller, Witt, Adamek, Mehr, Rogers, et al., 2002). Although a small selection of interactive training programs are available through manufacturers of cochlear implants (Rehabilitation Manual, Cochlear Corporation, 1998), there does not appear to be a consensus among these professionals, clinicians, and patients about what is the most beneficial form of treatment for newly implanted individuals. Therefore, examining the effects of different training methods on normal-hearing listeners’ abilities to learn to understand speech processed through acoustic simulations of cochlear implants may provide new insights into what training methods would also be most useful to newly implanted cochlear implant patients.

Several previous studies have examined the efficacy of several different types of training procedures on normal-hearing listeners exposed to speech processed through acoustic simulations of cochlear implants. These studies were conducted, in part, to help determine what methods of training may also be useful to deaf cochlear implant users (Chiu & McCabe, submitted; Davis et al., 2005; McCabe & Chiu, 2003). The interest in the type of training used to help normal-hearing listeners understand speech modeled after input relayed through cochlear implants has focused primarily on the training tasks used and the feedback given.

Chiu and McCabe (submitted) specifically examined what training tasks or combination of training tasks were most beneficial to normal-hearing listeners trying to understand the novel auditory stimuli. They compared an eclectic, interactive training regime that was modeled after a cochlear implant manufacturers’ own training program with a simpler, computer-based training method. The interactive training program included a variety of vowel, consonant, syllable, and sentence identification tasks. The computer-based training used only sentence identification tasks. The authors reported that the simpler and self-paced computer-based training method was as adequate as the more diverse training program designed to include on-line interaction with a “trainer” or a clinician. Their results suggest that relatively simple, self-guided training may be just as useful to cochlear implant patients as training supervised in a clinical setting.

Davis and colleagues (2005) also examined the impact of simple, self-paced training on normal-hearing listeners’ abilities to learn to understand speech processed through an acoustic simulation of a cochlear implant. In addition, they specifically considered what type of feedback is best to use during this type of training. After making multiple comparisons of methods using written and clear auditory feedback, they determined that feedback that induces a “pop-out effect” is most effective.

“Pop-out” can be described as the experience of re-hearing a degraded stimulus token after its identity has already been revealed. Presumably, when a listener specifically knows what the identity of a sentence is, when he/she listens to it again, the percept of parts of or even the entire sentence may clearly pop-out (Remez, Rubin, Pisoni, & Carrell, 1981; Remez, Rubin, Berns, Pardo, & Lang, 1994). To induce a pop-out effect, Davis and colleagues (2005) gave unprocessed auditory or written feedback to the listeners and then simply replayed the degraded sentence again. They found that feedback that creates a
PERCEPTUAL LEARNING WITH AN ACOUSTIC SIMULATION

A pop-out experience was effective regardless of whether it used written or unprocessed auditory feedback. However, in subsequent experiments examining the effects of semantic context on normal-hearing listeners' abilities to learn to understand speech processed through an acoustic simulation of a cochlear implant, pop-out feedback using only auditory repetitions of the stimuli were used by Davis and colleagues (2005).

Although the pop-out feedback was determined to be effective regardless of whether written or auditory information was given, only one of these forms of feedback is actually feasible with cochlear implant patients. Unlike normal-hearing listeners, deaf cochlear implant users cannot be presented with a clear repetition of a stimulus token. Therefore, examinations of perceptual learning by normal-hearing listeners trained to understand speech processed through an acoustic simulation of a cochlear implant should include some form of visual written feedback that could also be used with cochlear implant users.

A form of feedback applicable to cochlear implant users is necessary in any perceptual learning study that attempts to determine how effective and robust training may be and to determine what training materials or tasks are most conducive to learning. Therefore, the present study examined the effects of semantic context of training materials on learning using a feedback condition that paired written text with a repetition of degraded auditory stimuli. This experiment tests the hypothesis that using this type of feedback is beneficial both because it is more clinically applicable than auditory feedback and because it is likely to generate the pop-out effect reported by Davis and colleagues (2005). In addition, using this clinically applicable feedback may provide new insights into how auditory perceptual learning may be affected when training stimuli lack semantic context that is normally critical for the top-down information processing that is used in speech perception.

**Method**

**Experimental Design**

Figure 2.1 shows a schematic representation of the experimental method. A pre-test and post-test design with training was used. The experimental procedure was divided into five phases: (1) Familiarization, (2) Pre-training, (3) Training, (4) Post-training, and (5) Generalization. The specific procedures conducted during each phase of the experiment will be discussed in detail in the Experimental Procedures. Six groups of participants were used. Each group differed according to what training stimuli were provided and what method of feedback was used during training.

**Participants**

In each of the six groups of participants, 24 young adults were used for a total of 144 participants. Participants were undergraduate students enrolled in an introductory psychology course at Indiana University. All participants were monolingual, native speakers of American English and reported that they were free of any speech, hearing, language, and attentional disorders. Participants received partial course credit for their participation in this study.

**Stimuli and Materials**

**Stimuli.** Well known nursery rhymes (e.g. Jack and Jill; Humpty Dumpty) were used for a brief period of familiarization with the acoustic simulation. The stimuli used before, during, and after training and in the generalization phase were meaningful and/or anomalous sentences taken from the Boys Town sentence and anomalous sentence list (Stelmachowicz, Hoover, Lewis, Kortekaas, & Pittman, 2000), the Harvard sentences list (IEEE, 1969), and a set of Harvard sentences that were converted into anomalous
sentences (Herman & Pisoni, 2000). The sentence stimuli used in the current study appear in Appendix A.

<table>
<thead>
<tr>
<th>(1) Familiarization</th>
<th>(2) Pre-test</th>
<th>(3) Training Phase</th>
<th>(4) Post-test</th>
<th>(5) Generalization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Passive Listening</td>
<td>Experimental Groups</td>
<td>Sentence transcription</td>
<td>Sentences transcription</td>
<td>Transcription</td>
</tr>
<tr>
<td>Nursery rhymes with text</td>
<td>(1-2) Sentence transcription</td>
<td>20 meaningful sentences</td>
<td>20 meaningful sentences</td>
<td>(1) 20 Tokens</td>
</tr>
<tr>
<td>NO feedback</td>
<td>Unprocessed stim as FB</td>
<td>NO feedback</td>
<td>NO feedback</td>
<td>(trained category)</td>
</tr>
<tr>
<td></td>
<td>Processed stim + text as FB</td>
<td></td>
<td></td>
<td>10 old</td>
</tr>
<tr>
<td></td>
<td>(3-4)Anomalous sentence transcription</td>
<td>Unprocessed stim as FB</td>
<td>NO feedback</td>
<td>10 new</td>
</tr>
<tr>
<td></td>
<td>Processed stim + text as FB</td>
<td>NO feedback</td>
<td></td>
<td>(2) 20 Novel tokens</td>
</tr>
<tr>
<td>Control Groups</td>
<td>Sentence transcription</td>
<td>NO feedback</td>
<td></td>
<td>(untrained set)</td>
</tr>
<tr>
<td>(5) Sentence transcription</td>
<td>NO feedback</td>
<td></td>
<td></td>
<td>(3) 50 Novel meaningful sentences</td>
</tr>
<tr>
<td>(6) Sentence transcription</td>
<td>NO exposure</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 2.1. Schematic representation of experimental procedure used in the current study.

The anomalous sentences were generated by substituting words in the original Harvard sentences with other words taken from the same syntactic category (noun, verb, etc.). All Harvard meaningful and anomalous sentences were between seven and 11 words long and contained five key content words that were scored during data analysis. The Boys Town meaningful and anomalous sentences each contained four words, all of which were key words scored during data analysis.

One set of twenty Harvard sentences was used to assess performance before and after a period of training. Training stimuli consisted of 80 Harvard sentences and anomalous sentences and 50 Boys Town sentences and anomalous sentences. Two blocks of novel Harvard sentences and one block of anomalous sentences were used during the generalization phase.

All of the stimulus materials were recorded by the author, a female speaker with a Midwestern dialect. Recordings were made in a sound-attenuated booth (IAC Audiometric Testing Room, Model 402A) using a Shure head-mounted microphone (SM98) placed approximately two inches away from the mouth. Recordings were digitized online (16-bit analog-to-digital converter (DSC Model 240)) at 22,050 Hz and stored directly on a personal computer in .wav format using an individualized version of a speech acquisition program (Dedina, 1987). Three tokens of each stimulus were recorded. The repetition judged to be of the highest quality by the experimenter was used.

Using sound editing software (Adobe Audition) all usable sound files were segmented to the zero crossing closest to the beginning and end of the waveform. The root mean square (RMS) amplitude for all the files was then obtained. Based on the RMS amplitude measure, all stimuli were equalized to an amplitude of 65 db(A) using the Level16 software program (Tice & Carrell, 1998).
In a pilot study, ten listeners were asked to transcribe 100 sentences spoken by the speaker used in this study to establish speech intelligibility. Results showed that nearly 95% ($M = 94.72$, $SD = 1.23$) of key words were identified correctly when listeners heard the sentences produced by this speaker in the clear. This analysis indicates that the speaker has intelligible speech. This is an important finding, because it indicates that any difficulties that listeners may have in identifying the speech when it is processed by the acoustic simulation of a cochlear implant are due to the nature of the signal degradation and not due to initially poor speaker intelligibility.

**Simulation Strategy.** All auditory stimuli were processed offline using a personal computer. The signal processing procedure used for the cochlear implant simulation was adapted from the real-time signal processing methods designed by Kaiser and Svirsky (2000). The signal was lowpass filtered with a cutoff frequency of 12,000 Hz. A bank of eight bandpass, analysis filters was then used to divide the auditory signal into its corresponding frequency components. The temporal envelope was extracted from the output of the filters and then used to modulate noise bands created through a bank of eight synthesis filters with the same cutoff frequencies as the original analysis filters. The cutoff frequencies for the analysis and synthesis filter banks are listed in Table 2.1.

**Table 2.1.** Frequency boundaries of the analysis and synthesis filters used in the acoustic model of the cochlear implant.

<table>
<thead>
<tr>
<th>Frequency Boundaries</th>
</tr>
</thead>
<tbody>
<tr>
<td>854 – 1467 Hz</td>
</tr>
<tr>
<td>1467 – 2032 Hz</td>
</tr>
<tr>
<td>2032 – 2732 Hz</td>
</tr>
<tr>
<td>2732 – 3800 Hz</td>
</tr>
<tr>
<td>3800 – 5150 Hz</td>
</tr>
<tr>
<td>5150 – 6622 Hz</td>
</tr>
<tr>
<td>6622 – 9568 Hz</td>
</tr>
<tr>
<td>9568- 11000 Hz</td>
</tr>
</tbody>
</table>

Because the frequency of the noise bands and the output of the filters were matched, this acoustic model most accurately represents the input transmitted to cochlear implant patients who hypothetically have deep enough electrode insertions to enable well matched tonotopic frequency and electrode placement. Such a placement would presumably result in little if any basalward frequency shift. In other circumstances, a frequency shift may result from cochlear implantation if electrodes are unable to be placed far enough into the apical end of the cochlea which is tonotopically tuned for low frequency sounds. When electrodes cannot be implanted this deep and are instead placed more basalward within the
cochlea, electrical stimulation for low frequency sounds is directed to neural fibers ordinarily tuned for somewhat higher frequency sounds. This mismatch may result in sounds being perceived as a higher frequency than normal.

The choice of an eight channel acoustic model was based on studies showing that adult cochlear implant users achieve asymptotic speech perception scores when they have this number of active channels (Fishman, Shannon, & Slattery, 1997; Friesen, Shannon, Baskent, & Wang, 2001). Additionally, the best adult cochlear implant patients achieve word identification scores similar to those obtained by normal-hearing listeners exposed to an 8-channel acoustic simulation of a cochlear implant, when the acoustic model is implemented without any frequency shift (Dorman, Loizou, Fitzke, & Tu, 2000).

Experimental Procedures

Participants were tested in individual perceptual testing stations in the Speech Research Laboratory at Indiana University. Each booth was equipped with a Gateway PC (P5-133) using a 15” monitor (Vivitron15). Auditory stimuli were presented over calibrated headphones (Beyer Dynamics DT100) at approximately 70dB(A) SPL as measured by a standard sound-level meter. Stimuli were always presented in random order within blocks to the participants.

Familiarization phase. Testing began with a brief period of familiarization in which participants simply listened to and read silently along with six popular nursery rhymes. The text of these nursery rhymes appeared on the top half of the monitor in a 30pt. sans serif font 500ms prior to the start of each utterance. The text of the nursery rhymes remained visible for the entire duration of the utterance. Listeners did not have the option to replay any of the familiarization stimuli.

Pre- and post-training phase. During pre- and post-training phases, listeners transcribed 20 Harvard sentences by typing their responses onto the computer’s keyboard. Responses could be entered immediately after each utterance ended and when a prompt and cursor appeared on the monitor. The same set of 20 sentences was used for both pre- and post-training. Participants did not receive any feedback during these two phases and did not have the option to repeat the sentence after it was played.

Training phase. During the training phase, listeners heard either normal semantically coherent sentences or semantically anomalous sentences (i.e. The deep buckle walked the old crowd). Listeners assigned to the anomalous sentence condition were told that they would be hearing sentences that did not make any sense and were given the written text of an example anomalous sentence to familiarize them with the task. After hearing each sentence, listeners typed their responses according to the same procedure used in the pre- and post-training phases.

There were two sub-groups of listeners within the normal sentence and anomalous sentence training groups. Each sub-group within the sentence and anomalous sentence group received a different type of feedback. Feedback was provided after listeners were completely done entering their responses. In one feedback condition, the same token of each sentence that listeners heard was played back in its unprocessed form. This condition will be referred to as the “unprocessed” auditory-only feedback condition.

In a second feedback condition, the processed sentence was played again after the orthographic text of the sentence appeared in the middle of the screen. The text of the sentence appeared 250ms prior
to the onset of each utterance and remained visible until each utterance stopped playing. This form of feedback will be referred to as the "processed + text" feedback condition.

In addition, two other control groups also heard and transcribed sentence materials during the training period. One group transcribed the same sentences processed through the acoustic simulation of the cochlear implant but did not receive any form of feedback. This group will be referred to as the "no feedback" group. This control group was used to help measure the effect that mere exposure to the processed speech without accompanying feedback has on learning. A second group transcribed the same sentences in their unprocessed form, without feedback. Thus, this group received no exposure to processed speech during the training period. This group will be referred to as the "no exposure" group. This control group was used to measure simple practice effects that may result in an increase in keyword recognition scores from pre-training to post-training.

**Generalization phase.** Following training and the post-training evaluation, three separate blocks of stimuli were presented in counterbalanced order to measure generalization. One block contained 50 novel Harvard sentences, and the other block contained 20 Harvard sentences. A third block contained 20 Harvard anomalous sentences. In the two blocks that contained 20 meaningful or anomalous sentences, half of the stimuli were presented to the listeners during training, and half were completely novel. Old stimuli were replayed to listeners to evaluate their short-term retention of training materials and to determine the effect that repetition of training materials has on keyword recognition scores. During generalization, listeners transcribed what they heard using the same methods used in the pre- and post-training task.

**Data Analysis**

The data output files containing the correct responses and each participant's responses for each experimental phase were imported into data spreadsheets for scoring. Each sentence was scored by determining the percentage of key words that were identified correctly. Key words containing obvious spelling and typographical errors were scored as correct. Errors in spelling were scored as correct only if the response did not resemble a different word or combination of words. Similarly, typing errors were only scored as correct if the response did not resemble a different word.

A typing error was scored as correct if a target letter was substituted by any letter immediately surrounding it or if letters immediately surrounding the target letter were inserted into the response. Responses in which the correct letters were transposed were also considered typographical errors and scored correctly. Insertions of unnecessary spaces within words were also scored as correct if the response did not result in a different word. Homophones (i.e. "rain" for "reign") of key words were also accepted as correct. However, key words reported in an incorrect tense (i.e. "place" for "placed") or with other incorrect affixes (i.e. "shoes" for "shoe") were scored as incorrect.

Scoring was completed by the experimenter and a trained undergraduate research assistant. Interrater reliability was established by computing the Cronbach’s alpha value of scores assigned to all the responses given by the first ten listeners in the study (Bland & Altman, 1997). The Cronbach’s alpha value obtained was .97, indicating that inter-rater reliability for scoring key words correct in this study was very high. Any ambiguous scoring decisions were resolved at the experimenter’s discretion.
Results

The descriptive statistics for each groups’ performance on each phase of the experiment appear in Table 2.2. To determine if all six groups of listeners had roughly the same abilities to identify speech processed through an acoustic simulation of a cochlear implant prior to training, a one-way ANOVA was conducted on the pre-training scores. This analysis revealed no difference \( F(5, 143) = .81, p = .543 \) in the pre-training scores, indicating that the groups were similar to one another prior to training.

Table 2.2. Means and standard deviations of the percent of key words correctly identified by the six groups of listeners before, during, and after training (left three columns) and during generalization (right three columns).

<table>
<thead>
<tr>
<th></th>
<th>Pre-Training</th>
<th>Post-Training</th>
<th>Trained Stim</th>
<th>Novel Stim (Old Category)</th>
<th>Novel Stim (New Category)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sentences</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unexposed</td>
<td>.37 (.11)</td>
<td>.97 (.06)</td>
<td>.47 (.14)</td>
<td>.71 (.16)</td>
<td>.42 (.12)</td>
</tr>
<tr>
<td>Untrained</td>
<td>.36 (.16)</td>
<td>.50 (.15)</td>
<td>.52 (.18)</td>
<td>.63 (.13)</td>
<td>.49 (.12)</td>
</tr>
<tr>
<td>Auditory FB</td>
<td>.35 (.10)</td>
<td>.51 (.11)</td>
<td>.53 (.11)</td>
<td>.82 (.11)</td>
<td>.48 (.11)</td>
</tr>
<tr>
<td>Proc. + Txt. FB</td>
<td>.41 (.10)</td>
<td>.60 (.08)</td>
<td>.62 (.10)</td>
<td>.86 (.12)</td>
<td>.57 (.10)</td>
</tr>
<tr>
<td><strong>Nonsense Sentences</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Auditory FB</td>
<td>.38 (.09)</td>
<td>.40 (.07)</td>
<td>.59 (.10)</td>
<td>.50 (.13)</td>
<td>.41 (.09)</td>
</tr>
<tr>
<td>Proc. + Txt. FB</td>
<td>.39 (.13)</td>
<td>.45 (.09)</td>
<td>.61 (.12)</td>
<td>.55 (.11)</td>
<td>.44 (.11)</td>
</tr>
</tbody>
</table>

Pre- and Post-training Data

Figure 2.2 shows the pre- and post-training scores of the six groups. When just post-training scores were submitted to a 2 x 3 (training stimuli x feedback type) ANOVA, a main effect was found for type of feedback \( F(2, 143) = 6.41, p = .002 \). Listeners who received any form of feedback \( M = .59, SD = .14 \) performed better than both groups of listeners who did not receive feedback \( M = .50, SD = .16 \).

By using a one-way ANOVA testing the effect of group, significant differences between the groups’ performance after training appeared more robust. A post-hoc Tukey analysis based on this test indicated that both the listeners trained on meaningful and anomalous sentences who received processed + text feedback performed significantly better than all of the four other listener groups. In addition, listeners trained on anomalous sentences who received the unprocessed auditory feedback \( M = .59, SD = .10 \) had significantly higher post-training scores than the group who received no exposure \( M = .47, SD = .14 \).

To assess the magnitude of improvement that occurred from pre- to post-training, a repeated-measures ANOVA was conducted using pre- and post-test transcription scores as the two levels of the within-subjects factor. The between-subjects factors were the type of training materials used and the type
statistics of the listeners' performance during the training task. Figure 2.3 displays the learning curves for each of the six groups for performance before, during, and after training. Figures 2.4-2.5 illustrate the individual learning curves of the 24 listeners included in each group given exposure to the processed speech during training.

![Learning Curves Graph]

**Figure 2.3.** Keyword recognition performance before, during, and after training for all listener groups exposed to speech processed through an acoustic simulation of a cochlear implant during the training period. Error bars represent the standard error of the mean.

**Effects of training stimuli and feedback.** A 2 x 3 (training stimuli x feedback) univariate ANOVA was conducted on the keyword recognition scores obtained during training. Main effects were observed for both training materials ($F(1, 119) = 37.45, p < .001$) and type of feedback used ($F(2, 119) = 8.21, p < .001$) during training. Listeners trained on meaningful sentences ($M = .53, SD = .12$) performed significantly better than listeners trained on anomalous sentences ($M = .42, SD = .08$). Post-hoc analysis indicated that listeners trained with processed + text feedback ($M = .52, SD = .11$) performed significantly better than listeners trained with auditory feedback alone ($M = .45, SD = .11$). The interaction effect between training stimuli and type of feedback was not significant.

A oneway ANOVA was conducted with accompanying post-hoc tests to specifically determine which groups performed better during training than others. Listeners trained on sentences who received processed + text feedback performed significantly better than all the other groups. Both the listeners who received no feedback during training and those who were trained on sentences with unprocessed auditory feedback had higher keyword identification scores than listeners trained on anomalous sentences with unprocessed auditory feedback alone.

The training data were also analyzed to determine the amount of improvement during the training period. Figure 2.6 displays performance during the first, middle, and last phase of the sentence training period for five of the six groups. The values for performance during each phase of training were obtained by splitting the 130 training sentences into fifths each containing 26 sentences and then determining the overall average percent correct for each group for each phase of training. Thus, Figure 2.6 presents the performance on the first, third, and fifth or first, middle, and last block of 26 sentences heard by the listeners.
of feedback provided. Results indicate a significant main effect of session (pre-test vs. post-test) \( F(1, 139) = 969.67, p < .001 \) and type of feedback \( F(2, 139) = 3.47, p = .034 \). The session by feedback type interaction was also significant \( F(1, 139) = 10.58, p < .001 \) indicating that not all groups improved equally from pre- to post-training. All other interactions were not significant.

![Bar chart showing percent correct](image)

**Figure 2.2.** Pre- and post-training performance of listener groups with no feedback (left panel), unprocessed auditory feedback (middle panel), and processed + text feedback (right panel). The change in percentage points from pre-training to post-training appears above the bars representing each group. Error bars represent the standard error of the mean.

A post-hoc Tukey test indicated that listeners who received processed + text feedback improved more from pre- to post-test than the control group of listeners who received no exposure to the processed speech during the training period. Although the control group’s keyword recognition scores did significantly improve by ten percentage points from pre- to post-test \( F(1, 23) = 96.73, p < .001 \), this improvement is simply due to a practice effect. Groups receiving no feedback or feedback in the form of unprocessed auditory stimuli did not differ from the control group that was not exposed to the processed speech during the training period.

To assess how listener groups differed from one another, a repeated-measures ANOVA was conducted using group as the between-subjects factor and assessment period (pre- or post-test) as the within-subjects factor. This analysis revealed differences in the amount of improvement from pre-test to post-test among the six groups of listeners. The main effect of group assignment was significant \( F(5, 138) = 2.55, p = .031 \), and there was an interaction between assessment session and group \( F(5, 138) = 9.85, p < .001 \). A post-hoc analysis indicated that listeners who received processed + text feedback improved more from pre-test to post-test than the control group of listeners that were given no exposure to processed speech during the training period \( p < .001 \).

**Training and Learning Data**

Because the amount and time course of auditory perceptual learning occurring during the training period played a role in pre- to post-training increases in performance, differences in performance during training were evaluated using several methods. Table 2.2 shows a summary that includes the descriptive
Figure 2.4. Learning curves for listeners trained with sentences using (a) no (b) unprocessed and (c) processed + text feedback. Error bars represent the standard error of the mean.
Figure 2.5. Learning curves for listeners trained on nonsense sentences using (a) unprocessed auditory feedback and (b) processed + text feedback. Error bars represent the standard error of the mean.

The control group, which received no exposure to the acoustically altered speech during the training period, was omitted from Figure 2.6 and all analyses of training performance because they simply transcribed the sentences in their unprocessed form. However, this group’s transcription accuracy was evaluated to provide a measure of the speaker’s intelligibility in the clear for this specific set of training sentences. The intelligibility of these sentences in normal auditory conditions was high ($M = .97$, $SD = .15$) indicating that the other groups’ speech perception performance during training was primarily influenced by the nature of the signal degradation rather than the intelligibility of the speaker.
**Figure 2.6.** Keyword recognition performance during the first, middle, and final phase of the entire training period for listeners trained with sentences with no feedback and with unprocessed auditory feedback or processed + text feedback (left panel) and by listeners trained with nonsense sentences with unprocessed auditory feedback or processed + text feedback (right panel). The change in percentage points from the first phase of training to the final phase appears above the bars representing each group. Error bars represent the standard error of the mean.

A repeated-measures ANOVA was conducted on the five groups' training performance using the three phases of training as the within-subjects factor and the type of feedback and training materials as the between-subjects factors. Results revealed a main effect of training phase ($F(2, 114) = 80.91, p < .001$). Listeners in all groups improved throughout training.

Main effects were also observed for the two between-subjects effects. Listeners who received feedback performed better overall during the training period than the control group who received no feedback ($F(2, 114) = 8.30, p < .001$). A post-hoc analysis indicated that two feedback conditions also differed from one another ($p = .012$). Listeners receiving the processed + text feedback performed better during these three phases of training ($M_1 = .45, SD_1 = .12; M_2 = .53, SD_2 = .13; M_3 = .57, SD_3 = .14$) than listeners receiving only unprocessed auditory feedback ($M_1 = .39, SD_1 = .12; M_2 = .46, SD_2 = .13; M_3 = .50, SD_3 = .14$). However, individually neither of these two feedback groups differed significantly from the control group that received no feedback during training.

A significant main effect of the training stimuli was also observed ($F(2, 114) = 38.60, p < .001$). Listeners trained on normal sentences ($M = .54, SD = .11$) performed better overall during training than listeners trained on nonsense sentences ($M = .43, SD = .08$). A test of within-subjects contrasts revealed an interaction between training phase and training stimuli ($F(1, 114) = 4.08, p = .046$). Listeners trained on meaningful sentences demonstrated larger gains during the training period than listeners trained on nonsense sentences.

**Effects of stimulus set.** Because the training period included two different types of sentence materials in disproportionate numbers, the training data from each group of sentences was replotted and analyzed separately in order to determine if there was an effect of sentence materials on the amount of
learning achieved during training. Eighty Harvard normal and anomalous sentences ranging from seven to 11 words in length and 50 less complex Boys Town anomalous and meaningful sentences that contained only four words each were the training materials. Figure 2.7 shows performance during the first, middle, and final phases of training on the 80 Harvard sentences and the 50 Boys Town sentences.

(a) Harvard Sentences

![Chart showing performance of Harvard Sentences](chart)

Feedback Condition

(b) Boys Town Sentences

![Chart showing performance of Boys Town Sentences](chart)

Feedback Condition

Figure 2.7. Keyword recognition performance for three phases of training with (a) Harvard and (b) Boys Town sentences and nonsense sentences. The change in percentage points from the first phase to the final phase of training appears above the bars representing each group. Error bars represent the standard error of the mean.
A repeated-measures ANOVA was conducted using the three phases of training as one within-subjects factor and sentence set (Harvard or Boys Town) as the other within-subjects factor. The between-subjects factors were the type of training stimuli (anomalous or meaningful sentences) and type of feedback used. Results indicate main effects for both training phase ($F(2, 114) = 73.69, p < .001$) and sentence set ($F(1, 114) = 274.95, p < .001$).

As expected, the shorter and less linguistically complex Boys Town sentences and anomalous sentences were easier for the listeners to identify ($M = .62, SD = .14$) than the Harvard sentences ($M = .38, SD = .21$). As found in the previous analysis, there were significant main effects of feedback ($F(2, 114) = 13.24, p < .001$) and training stimuli ($F(1, 114) = 72.18, p < .001$). Tests of within-subjects contrasts revealed a three-way interaction effect between phase of training, feedback, and training stimuli ($F(1, 114) = 4.01, p = .048$). Overall, listeners trained on sentences with processed + text feedback perform better and learn more throughout training than listeners trained on anomalous sentences and listeners who received clear auditory feedback.

**Generalization Data**

The degree to which perceptual learning during training could be generalized to new stimuli was also tested. To examine generalization of learning, several test phases were conducted after training. In these generalization tests, listeners were asked to identify new tokens without receiving feedback. Listeners heard novel tokens taken from the same category of stimuli on which they were trained (meaningful or anomalous sentences) and also heard novel stimuli from the category on which they were not trained (meaningful or anomalous sentences). The two rightmost columns of Table 2.2 provide a summary of the descriptive statistics of performance during the generalization phases.

Figure 2.8 shows keyword identification scores of all six listener groups for novel sentences and anomalous sentences. For the sentence training groups, generalization of learning was inferred from performance on novel Harvard sentences and on completely new anomalous Harvard sentences. For the anomalous sentence training group, the measure of generalization of learning was based on the listeners' performance on novel anomalous sentences similar to what were used in training and on meaningful sentences. Therefore, for all listeners, generalization of learning to novel training tokens and to a novel stimulus category was indexed by measuring their performance on separate blocks of Harvard meaningful and anomalous sentences administered after training and the brief post-training assessment.

To assess performance during the generalization phase, a repeated-measures ANOVA was conducted using the generalization stimuli (meaningful or anomalous sentences) as a within-subjects factor and the type of training materials and feedback used as between-subjects factors. Results revealed a main effect for the type of stimuli heard in the generalization phase. Overall, the key words in sentences ($M = .51, SD = .12$) were easier ($F(1, 139) = 426.41, p < .001$) for listeners to identify during generalization than the keywords in anomalous sentences ($M = .39, SD = .16$). A three-way interaction was found between the generalization stimuli, the type of feedback received during training, and the training materials heard ($F(1, 139) = 7.85, p = .006$). This three-way interaction reflects the slightly poor performance on novel sentences by listeners trained on sentences who received clear auditory feedback.

An interaction between the generalization stimuli and the training stimuli also approached significance ($F(1, 139) = 3.10, p = .080$). Listeners trained on anomalous sentences appeared to perform better on the novel sentences than listeners trained on meaningful sentences. The type of feedback received was the only between-subjects variable that had a consistent effect on performance during generalization ($F(2, 139) = 6.39, p = .002$). A post-hoc Tukey HSD test indicated that listeners receiving
no feedback performed worse during the generalization phase than listeners receiving either type of feedback ($p < .001, p = .007$).

![Graph showing percent correct performance across training conditions](image)

**Figure 2.8.** Performance during generalization phases using novel sentences and anomalous sentences by listener groups with no feedback (left panel), unprocessed auditory feedback (middle panel), and processed + text feedback (right panel). Error bars represent the standard error of the mean.

To more accurately assess how the six groups' speech perception scores differed during the generalization phase, a second repeated-measures ANOVA was conducted using group assignment as the between-subjects variable and the generalization stimuli as the within-subjects variable. In addition to the previously reported main effect of stimuli in the generalization phase ($F(1, 139) = 6.39, p = .002$), an interaction was observed between the generalization stimuli and group assignment ($F(5, 139) = 2.95, p = .015$). The main effect of group assignment was also significant ($F(5, 139) = 4.65, p = .001$). Post-hoc analysis found differences between the control group given no exposure to the processed speech and several other trained groups. Both groups trained on meaningful sentences and nonsense sentences who received processed + text feedback performed better during the generalization phase than the control group that received no exposure to the processed speech ($p < .001, p = .004$). Of the listeners trained with clear auditory feedback alone, only the group trained on anomalous sentences performed better than the control group who received no exposure.

Although comparing the groups' abilities to identify novel stimuli after training provides some insight into how much learning generalized to new stimuli, the listeners' performance on novel training tokens and on a novel stimulus category alone does not provide the most accurate measure of generalization. A better assessment of the amount of generalization can be made when performance during the two blocks of the generalization phase is compared to performance during training and during post-training. Figure 2.9 displays the mean word recognition scores obtained during training, generalization of learning to novel anomalous sentences and novel meaningful sentences, and post-training. If performance during the generalization phases is equal to or exceeds performance during the training period, it can be concluded that generalization to new tokens or categories has occurred. Therefore, the magnitude of the difference between training and generalization scores can be used as one index of the extent of the generalizability of learning.
Figure 2.9. Generalization of learning during the training period to anomalous and meaningful sentences by listeners trained on sentences without feedback and with unprocessed auditory feedback and processed + text feedback (left panel) and by listeners trained on nonsense sentences with unprocessed feedback and processed + text feedback (right panel). Error bars represent the standard error of the mean.

The magnitude of generalization can also be inferred by examining the difference between post-training scores and generalization scores. Post-training scores were obtained using the same materials used in pre-training. Due to the familiarity of the stimuli, listeners should attain higher keyword recognition scores during post-training than during the generalization phases that follow. The magnitude of this difference is also a good indication of whether or not generalization was robust in listeners. Listeners who display generalization scores that differ the least from the post-training scores are considered to have generalized learning the most.

To determine the magnitude of the differences between training and post-training performance and generalization performance in the listeners, two repeated-measures ANOVAs were conducted. In the analyses that compared training performance to performance on new training tokens and a new stimulus category, only the training data from the 80 Harvard sentences were used. This restriction was made because the generalization phase only used sentences from this set. However, the training phase also included sentences from the easier Boys Town sentence set. The inclusion of these easier sentences in the training set may have inflated training scores making the full training data set an inappropriate comparison for generalization data. The data from the listener group who received no exposure to the processed speech during the training period was also excluded from this analysis, because their performance during the training period was obtained using unprocessed sentences.

To test the difference between training and generalization performance, a repeated-measures ANOVA was conducted using experiment phase (training, generalization to novel anomalous sentences, and generalization to novel meaningful sentences) as the within-subjects factor and training stimuli and feedback as between-subjects factors. A main effect of experiment phase was found ($F(2, 114) = 97.08, p < .001$). The interaction effect between experiment phase and the training stimuli used was also significant ($F(2, 114) = 18.18, p < .001$). This interaction reveals lower performance on anomalous sentences by the listeners trained on meaningful sentences. Thus, listeners trained on meaningful
sentences were unable to generalize learning to anomalous sentences. All other within-subjects interactions were nonsignificant.

The between-subjects effect of type of feedback was also significant \( F(2, 114) = 9.88, p < .001 \). A post-hoc Tukey test indicated that listeners who received processed + text feedback performed better overall than the listeners who received unprocessed auditory feedback and no feedback \( (p = .013, \ p = .001) \). The main effect of the training stimuli approached significance \( F(1, 114) = 3.57, p = .061 \).

An identical 3 X 2 repeated-measures ANOVA was also conducted to compare post-training and generalization performance. A main effect of experiment phase was found \( F(2, 114) = 204.31, p < .001 \). The three-way interaction between experiment phase, training stimuli, and feedback was also significant \( F(2, 114) = 3.14, p = .045 \). This interaction indicates that listeners trained on anomalous sentences with clear auditory feedback were able to generalize learning to meaningful sentences more than listeners trained on meaningful sentences. All other interaction effects were nonsignificant. The main effect of feedback was also significant \( F(2, 114) = 3.86, p = .024 \). Post-hoc analysis indicated that the group of listeners trained with processed + text feedback performed better overall than listeners who received no training \( (p < .001) \).

**Stimulus Repetition Data**

A third block of stimuli that listeners were tested on during the generalization phase included tokens that were presented to listeners during the initial training period. Refer to Table 2.2 for descriptive statistics of listeners’ keyword recognition scores for stimuli on which they were previously trained. The inclusion of this group enabled an examination of short-term retention of learning and repetition effects. Figure 2.10 illustrates the differences between performance on novel tokens taken from the category of stimuli trained on and old tokens presented during training for each of the six listener groups.

A repeated-measures ANOVA was used to determine the effect of stimulus repetition on the ability to identify speech processed through an acoustic simulation of a cochlear implant. The repetition of stimuli was the within-subjects factor, while the type of training/test materials and feedback used were the between-subjects factors. Results show a large main effect of repetition \( F(1, 139) = 382.77, p < .001 \). Stimuli that were presented during the training period \( (M = .67, SD = .18) \) were transcribed more accurately during the generalization phase than stimuli that were novel \( (M = .47, SD = .23) \). Both two-way interaction effects were significant. Listeners trained on meaningful sentences displayed a much larger repetition effect than listeners trained on nonsense sentences \( F(1, 139) = 89.18, p < .001 \). Repetition effects were also larger for listeners who were given feedback during training than for listeners who did not receive training \( F(1, 139) = 9.03, p < .001 \).

Overall main effects were also obtained for the training stimuli and feedback used. Listeners trained on sentences performed better on both old and new stimuli than listeners trained on nonsense sentences \( F(1, 139) = 75.33, p < .001 \). A post-hoc analysis was conducted to more conclusively determine the main effect of feedback \( F(1, 139) = 15.78, p < .001 \). Results indicated that listeners who received processed + text feedback performed better on this phase of the experiment than listeners who received only clear auditory feedback \( (p = .004) \). The control groups that received no feedback did not differ from either of the two feedback groups.

In addition to comparing performance on the old training tokens to performance on the novel tokens, comparisons were made between the accuracy in identifying these same sentences during training and after training. Figure 2.11 shows the average scores obtained on the same ten sentences during and
after training. The control group who received no exposure to the processed stimuli during training is not included in Figure 2.11 and was not used for the data analysis, because their scores during training are based on the unprocessed tokens.

(a) Meaningful Sentences

(b) Anomalous Sentences

Figure 2.10. Word recognition scores on old and new stimuli by listener groups trained on meaningful (a) or anomalous sentences (b). The difference in percentage points between the old and new stimuli appears above the bars representing each group. Error bars represent the standard error of the mean.
Figure 2.11. Word recognition scores on the same ten sentences during and after training for listeners trained on sentences with no feedback, unprocessed feedback, and processed + text feedback (left panel) and trained on nonsense sentences with unprocessed feedback and processed + text feedback (right panel). The change in percentage points from the first repetition to the second is shown above the bars representing each group. Error bars represent the standard error of the mean.

A repeated measures ANOVA was conducted using experiment phase (during or after training) as the within-subjects factor and training stimuli and feedback as the between-subjects factors. In general, listeners were much better at identifying the keywords in the set of sentences after training \( (M = .67, SD = .11) \) than during \( (M = .47, SD = .19) \) training \( (F(1, 114) = 325.70, p < .001) \). An interaction effect was found between the training stimuli used and the amount of improvement between the first and second repetition of the sentences \( (F(1, 1114) = 41.94, p < .001) \). The interaction effect between feedback and the amount of improvement from the first to second was also significant \( (F(1, 114) = 15.94, p < .001) \). The three-way interaction was not significant.

Listeners trained on normal sentences \( (M = .65, SD = .17) \) performed better overall \( (F(1, 114) = 109.70, p < .001) \) on this set of sentences than the listeners trained on anomalous sentences \( (M = .46, SD = .19) \). A main effect of feedback was also found \( (F(1, 114) = 15.75, p < .001) \). Listeners who received any form of feedback performed better than the listeners who did not receive feedback. The between-subjects interaction effect was not significant. A post-hoc Tukey test determined that listeners who heard processed + text feedback were more accurate in identifying words in these sentences than listeners who received auditory-alone feedback \( (p = .001) \).

Predictors of Performance

In addition to the group mean comparisons and analyses of variance, additional tests were also conducted to identify predictors of listeners' performance during several phases of the experiment. The data from five out of the six listener groups were fit to linear regression models to determine what variables accounted for the most variance in the percentage of keywords identified during and after training and during the generalization phase. A linear regression was also used to predict short-term retention of learning or gains in performance due to stimulus repetition. Pre-training, training, and post-
training scores were used as continuous variables and the training materials and type of feedback used were categorical variables in the analyses. All of these variables were used as predictor variables in the regression models where appropriate.

The control group that was given no exposure to the processed stimuli during training was omitted from the linear regression because no valid training data were available for them. Using the enter method, five linear regressions were performed to identify significant predictor variables for listeners' performance during and after training and for listeners' performance on novel, meaningful and anomalous sentences and old stimuli used in the training period. Table 2.3 shows a summary of the results of the regression analyses. The dependent variables being predicted in each regression analysis are listed in bold print. The significant predictor variables found in each analysis are grouped below the boldfaced headings.

**Table 2.3.** Results of five separate linear regressions used to identify predictors of performance during training and post training, short-term retention of learning, and generalization to novel sentences and anomalous sentences. The dependent variables being predicted in each regression analysis are listed in bold print. The significant predictor variables found in each analysis are grouped below the boldfaced headings.

<table>
<thead>
<tr>
<th>Predictors of Performance During Training</th>
<th>β</th>
<th>$r^2 = .72$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-training performance</td>
<td>.66***</td>
<td></td>
</tr>
<tr>
<td>Training materials</td>
<td>.56***</td>
<td></td>
</tr>
<tr>
<td>Type of feedback</td>
<td>.23***</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Predictors of Post-training Performance</th>
<th>$r^2 = .82$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training performance</td>
<td>.58***</td>
</tr>
<tr>
<td>Training materials</td>
<td>-.41***</td>
</tr>
<tr>
<td>Pre-training performance</td>
<td>.42***</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Predictors of Short-term Retention of Learning</th>
<th>$r^2 = .70$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training performance</td>
<td>.31*</td>
</tr>
<tr>
<td>Training materials</td>
<td>.62***</td>
</tr>
<tr>
<td>Type of feedback</td>
<td>.28***</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Predictors of Generalization to New Sentences</th>
<th>$r^2 = .55$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training performance</td>
<td>.48***</td>
</tr>
<tr>
<td>Training materials</td>
<td>-.35**</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Predictors of Generalization to New Nonsense Sentences</th>
<th>$r^2 = .47$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training performance</td>
<td>.58***</td>
</tr>
<tr>
<td>Training materials</td>
<td>-.34**</td>
</tr>
</tbody>
</table>

The predictor variables for three out of the five models accounted for a substantial amount of variance (70% to 82%) in performance during and after training and for variance in short-term retention of learning. Performance during training was predicted by pre-training scores and by the type of feedback and training stimuli used. Post-training scores and the short-term retention of old training stimuli were both predicted, in part, by the type of training materials used and by listeners' performance during training. Additional variance of post-training scores was accounted for by pre-training scores. The type of
feedback used during training was also a significant predictor of performance on old training stimuli presented during the generalization phase. Roughly half of the variance was accounted for when performance on novel, anomalous and normal sentences was predicted from training performance and training materials.

Although the type of training materials used was a significant predictor for several of the dependent variables, the standardized beta coefficients indicate that the nature of the prediction was not always the same. Positive standardized beta coefficients associated with the training material variable were obtained in the regressions predicting training performance and short-term retention of learning. In accordance with the group mean comparisons and analyses of variance tests, this result indicates that training with sentences predicts better performance during training and predicts a larger benefit from stimulus repetition. However, negative beta coefficients were found when training stimuli used to predict post-training performance and generalization of learning to novel sentences and anomalous sentences. This result indicates that training with the harder anomalous sentences that lack semantic interpretation predicted better keyword recognition scores during the post-training phase and during generalization to novel sentences and anomalous sentences.

Discussion

Rapid and Robust Auditory Perceptual Learning

The results of this study support three main conclusions about auditory perceptual learning of spectrally degraded speech modeled after the input received by cochlear implant users. First, this study demonstrated that word recognition performance improves very rapidly over a short period of time regardless of the amount of exposure to the stimuli, the presence or quality of feedback, or the semantic content of training materials. However, a second finding demonstrated that the magnitude and robustness of perceptual learning achieved by listeners during this time period is strongly influenced by the amount of exposure, the presence of feedback and the type of feedback received. Finally, a third finding indicates that, although it may be initially much harder to perceive and remember training materials that lack a semantic interpretation, there may be subtle benefits to being trained on anomalous sentences compared to meaningful sentences.

Perceptual learning of speech processed through an acoustic simulation of a cochlear implant appears to be rapid and robust, because even the listeners given no exposure to the processed speech between pre- and post-training assessment displayed a significant increase in keyword identification from pre- to post-training. However, because the same set of sentences was used during the pre- and post-training phase, this improvement is assumed to represent a simple practice effect. Simply giving some listeners exposure to the degraded auditory stimuli during a control training condition did not result in any improvements that were significantly greater than the practice effect alone. This result indicates that the specific kinds of feedback may be an important factor in producing large perceptual learning effects for spectrally degraded speech. This conclusion was supported by the finding that listeners trained on either meaningful or anomalous sentences who received any type of feedback during training achieved significantly higher post-training scores than listeners trained on sentences who received no form of feedback.

The absence of feedback was also found to be detrimental to performance during the generalization phase. However, more detailed post-hoc analyses examining the no-feedback listeners according to whether they received exposure or not indicated that only the listeners who had no feedback and no exposure differed in their generalization performance. Thus, it appears that simple short-term
exposure to speech processed through an acoustic simulation of a cochlear implant does produce a small but still significant amount of auditory perceptual learning. In addition, this learning generalizes to new tokens that are similar to the ones used during training.

After being exposed to training sentences without receiving feedback, listeners were also able to transcribe training sentences more accurately the second time they were presented. This finding indicates that listeners successfully encoded and stored the sentences even in their degraded form. In addition, listeners who received no exposure to degraded speech during training also showed a large benefit of stimulus repetition. This is an interesting finding, because it suggests that encoding a sentence in its unprocessed form can later aid in identifying the same sentence in a degraded form. However, overall, benefits of stimulus repetition were greatest when feedback is provided during perceptual learning.

Feedback Effects on Perceptual Learning

In addition to the mere presence of feedback, the specific type of feedback administered to listeners was found to have selective effects on various aspects of perceptual learning and spoken word recognition performance. The degree of improvement observed from pre-training to post-training, which reflects the strength of the perceptual learning effect, was significantly greater in one of the two feedback groups used in this study. Processed + text feedback resulted in the greatest amount of improvement following training.

The amount of learning that occurred during the training period was also related to the type of feedback used. The trends observed here were similar to those found in the pre- and post-training analysis. Listeners who received processed + text feedback showed larger improvements during training and performed better overall during training than listeners who only heard the unprocessed auditory stimulus during training.

Two out of the three analyses examining generalization of learning also indicated that listeners trained with degraded auditory and written text feedback performed better than other listeners. When word recognition scores for novel sentences and anomalous sentences were compared to scores obtained during training and post-training, listeners who received degraded auditory and written text as feedback showed greater generalization. Evidence of greater generalization in this training group indicates that perceptual learning was more robust overall.

In addition to the ability to quickly adapt to speech processed through an acoustic simulation of a cochlear implant and generalize this kind of perceptual learning, the processed + text feedback also produced larger repetition effects. That is, listeners who received this form of feedback were more accurate in transcribing sentences that they had heard previously than other listeners. This result suggests that encoding tokens during training and retaining a representation of them afterwards appears to be easier when listeners have the opportunity to match degraded acoustic-phonetic signals to the intact phonological representations elicited through reading orthographic text. Overall, the current results suggest that the benefits of even a short training regimen using processed auditory feedback paired with orthographic text are robust, widespread, and superior to the benefits provided by just unprocessed auditory feedback or mere exposure to the spectrally degraded speech without feedback.

Taken together, the findings indicating that the processed + text feedback is superior to unprocessed auditory feedback are informative and theoretically and clinically significant. The feedback condition in which written text was paired with a repetition of the training tokens is the only feedback condition examined in this study that could feasibly be used in a clinical setting with cochlear implant
users. Unlike the normal-hearing listeners used in this study, deaf cochlear implant users do not have the option to hear the clear auditory stimulus as feedback. Demonstrating that a clinically applicable form of feedback is superior to more traditional auditory-alone feedback suggests that deaf cochlear implant users may still have an optimal opportunity to improve their speech perception abilities within short training periods despite the fact that all possible training and feedback options are not available to them.

However, simply determining that one type of feedback is better than another provides little insight into the perceptual learning process that is used by listeners who received each type of feedback. Theoretically, it is important to consider why the degraded auditory feedback with written text is superior to clear auditory feedback and why listeners receiving this type of feedback appeared to learn more. The degraded auditory and written feedback differed in several ways from the auditory-alone feedback. Listeners who received the degraded auditory and written text feedback heard more repetitions of the processed speech than the listeners who heard clear auditory feedback. However, listeners who received no exposure to the degraded speech did not differ overall from listeners who received exposure to the degraded stimuli without feedback. Thus, it is likely that a factor other than the amount of exposure to degraded speech stimuli enabled listeners receiving degraded auditory and written text feedback to learn more about the unfamiliar speech sounds they heard.

According to work by Davis and colleagues (2005), feedback is most effective when it elicits a “pop-out” effect that enables listeners to covertly or subvocally experience a spoken utterance while simultaneously hearing the degraded stimulus. In the present study, when listeners received the degraded auditory and written text feedback, they presumably read along with the sentence as it was being played. Through this process, listeners should be able to map the degraded acoustic-phonetic information in the signal with familiar phonological representations with reasonable accuracy to experience pop-out. Learning to match even small amounts of the degraded acoustic-phonetic information onto phonological representations of familiar words may have given the listeners receiving the degraded auditory and written feedback an advantage over listeners receiving auditory-alone feedback.

The effectiveness of the processed + text feedback condition and the phenomenon of “pop out” can be explained through the transfer appropriate processing (TAP) theory of learning and memory (e.g. Craik, 1994; Lockhart, 2002; Morris, Bransford, & Franks, 1977; Neath & Surprenant, 2003; Rajaram, Srinivas, & Roediger, 1998; Roediger, Buckner, & McDermott, 1999). This theory posits that learning and memory can be enhanced if the appropriate cognitive processes are executed during the training task or during learning. For an auditory perceptual learning task such as identifying speech processed through an acoustic simulation of a cochlear implant, encoding the stimulus is one of the most fundamental and important processes. Encoding is the mechanism through which the auditory stimulus is first perceived and interpreted (Craik, 1994).

However, in order to successfully learn how to identify a degraded speech signal, listeners must also be able to retrieve the correct sub-lexical and lexical representations of the stimulus as they initially encode it or when they receive feedback. According to TAP, retrieval is the most successful when the original conditions of encoding are “reinstated “ at the time of test (Craik, 1994). Thus, in auditory training tasks, feedback that enables accurate retrieval of the identity of the stimulus while reinstating encoding should be the most useful. This principle was demonstrated in the current study. The best feedback condition in the current study paired the original degraded stimulus with written text. A positive consequence of this feedback is that the initial conditions of encoding and learning were reinstated and simultaneous encoding and identification of the stimulus facilitated a “pop out” experience (Davis et al., 2005). Therefore, the “pop out” effect reported by Davis and his colleagues can be largely attributed to the fact that the training task and feedback encouraged transfer appropriate processing.
Effects of Semantic Context on Perceptual Learning

In the present study, listeners also appeared to have advantages related to the type of stimuli on which they were trained. As expected, during training, listeners trained on sentences performed better than most listeners trained on anomalous sentences, regardless of whether feedback was provided during sentence training. This result is consistent with many earlier studies examining speech perception for stimuli that deviate from normal linguistic rules (Malgady & Johnson, 1977; Marks & Miller, 1964; Miller & Isard, 1963). In addition, listeners trained on sentences demonstrated a larger repetition effect for stimuli that they heard again after training than listeners trained on anomalous sentences. The difficulty in encoding and retaining the content of anomalous sentences in memory (Malgady & Johnson, 1977; Marks & Miller, 1964) is no doubt one cause of the small repetition effects observed in the anomalous sentence training groups.

However, listeners trained on anomalous sentences were not disadvantaged in all phases of this experiment. At the post-training phase, listeners who were trained on anomalous sentences did not perform any worse than listeners trained on sentences. In addition, some listeners trained on anomalous sentences actually achieved higher post-training scores than listeners who received no exposure to the degraded speech during training. Taken together, these results suggest that learning to perceive spectrally degraded speech modeled after the input from cochlear implant users can effectively occur even when training stimuli are stripped of meaningful semantic context. These results replicate Davis et al.'s (2005) previous findings and, importantly, did so using a form of feedback that could actually be used with cochlear implant users.

Listeners trained on anomalous sentences also appeared to be able to generalize their learning to novel stimuli more than listeners trained on meaningful sentences. Although main effects of training stimuli were not evident in the repeated-measures tests of generalization, several interactions were observed between the training stimuli used and degree of generalization. A two-way interaction effect between training stimuli and generalization of learning to anomalous sentences was found. Listeners trained on sentences did not generalize their learning to anomalous sentences as well as listeners trained on anomalous sentences. A three-way interaction found that listeners trained on anomalous sentences with unprocessed auditory feedback performed better on meaningful sentences during generalization than listeners trained on meaningful sentences with the same type of feedback. Thus, it appears that listeners' abilities to generalize what they have learned during anomalous sentence training may be related to the specific type of feedback they received.

Results of a regression analysis also indicate that training on stimuli that lack appropriate semantic information may actually benefit post-training performance and generalization. A regression analysis used to predict post-training scores found that the stimuli heard during training affected performance after training. Negative beta weights in this analysis indicated that training with anomalous sentences predicted better post-training performance. Similarly, the type of training materials used was also a significant but weak predictor of the amount of generalization to novel meaningful and anomalous sentences. Training regimes using anomalous sentences were also associated with more robust generalization of learning to novel training tokens and a novel stimulus category.

These results are of some theoretical interest because they suggest that being trained on harder speech materials which lack appropriate semantic information may lead to better speech perception performance after training. Training with sentences that are not semantically predictable may be especially effective because they may encourage listeners to rely more on bottom-up processing to encode and store the stimulus materials. While engaging more in bottom-up processing to identify words...
that are simply strung together in no predictable manner, listeners may become more aware of the fine acoustic-phonetic and sub-lexical changes that result from the acoustic simulation of the cochlear implant.

The suggestion that harder and unfamiliar training materials can lead to more robust learning and generalization is not new. Martin and Gierut (2004) proposed that using nonword training materials to treat children with phonological disorders may result in faster and more generalizable and permanent learning. However, in treating a phonological disorder, the primary goal is to change speech production. In contrast, in the present study, changes in speech perception ability were examined. Thus, the current evidence showing that anomalous sentences are at least, or perhaps even more, effective training stimuli than normal sentences suggests that the benefits of using training materials that lack semantic or lexical meaning may be observed in both speech perception and production and auditory perceptual learning. Clinically, this is an important finding because it suggests that it may be useful to employ semantically anomalous sentences as training materials with cochlear implant users. These listeners may also successfully learn from training that draws attentional focus to sub-lexical acoustic-phonetic and phonological information rather than to lexical and semantic cues. Similarly, cochlear implant users may also be able to generalize this learning robustly and rapidly to new stimuli.

The finding that listeners trained on anomalous sentences display better generalization to new stimuli is not completely conclusive in this study. Only the results of the regression analysis support the proposal that training with anomalous sentences may lead to more robust learning than training with meaningful sentences. In the various ANOVAs, effects of training stimuli on generalization performance only approached significance. One explanation for the weak support of better generalization of learning by listeners trained on anomalous sentences is the combination of stimuli used to measure generalization in this study. It is difficult to determine precisely how the semantic content in sentences influenced generalization of learning to novel stimulus categories when the generalization stimuli also differ in semantic content. In other words, it is not ideal to assess generalization to a novel stimulus category by comparing the sentence training groups’ anomalous sentence scores to the anomalous sentence training groups’ sentence scores. The primary problem in making this comparison is that the two stimulus sets are not equally intelligible.

To make a fairer comparison of generalization between the sentence and anomalous sentence groups, additional test conditions could be considered in which all listeners need to identify new categories of the same stimuli such as CVCs or isolated words during the generalization phase. Previous research has found that speech perception training on sentences generalizes to isolated words and that training on specific phonemes in one environment generalizes to the same phonemes in different environments (Greenspan, Nusbaum, & Pisoni, 1988; Schwab, Nusbaum, & Pisoni, 1985). However, it is unknown whether training on anomalous sentences would also generalize to isolated words. The results of the present study suggest that anomalous sentence training should generalize to isolated words at least to the same extent as meaningful sentence training. A future direction of research examining effects of semantic content on the robustness of auditory perceptual learning is to include additional measures of generalization of learning. Similarly, assessments of transfer of learning in listeners trained on anomalous sentences could also be valuable tools to help determine the role that semantic context has on the quality and quantity of perceptual learning.

Clinical Implications and Future Directions

Despite the problems noted in measuring the effects of semantic content on generalization to new stimuli, the present study did yield several novel results that have clinical implications for how listeners
with cochlear implants may initially understand and learn to understand speech when the semantic content is reduced and speech perception must rely more on the degraded acoustic-phonetic detail. The utilization of top-down linguistic processes in perceiving speech through actual cochlear implant devices is expected in post-lingually deaf adults because they have acquired spoken language normally and have intact lexicons and stored auditory representations on which to map new sensory input. However, it has been suggested that post-lingually deafened adult cochlear implant users may actually rely less on some lexical information when completing gated-word recognition tasks than normal-hearing adults (Collison et al., 2004). In addition, in some cases, cochlear implant users may not be able to use lexical knowledge or sentence context at all to perceive speech. For example, pre-lingually deaf children learning their first spoken language or post-lingually deaf adults learning a second language will not be able to use prior lexical and contextual information to decipher speech signals.

Similarly, cochlear implant users may frequently encounter everyday communicative situations in which other cues that they would ordinarily use to understand speech are unavailable. For example, while using a telephone or listening to a public-service announcement, both lip-reading and pragmatic cues are largely unavailable. Therefore, it is important to understand how perceiving speech through a cochlear implant is affected when an important cue to understanding speech is reduced and the speech signal must be decoded primarily by using bottom-up processing of acoustic-phonetic and sub-lexical information. The results of the present study are encouraging because they suggest that training cochlear implant users to understand laboratory speech that lacks top-down processing cues could be successful and perhaps improve speech perception in everyday listening conditions that may lack important contextual cues normally used to perceive speech.

Viewed from a different perspective, the present results also support the importance of being able to use semantic content and lexical information to initially understand speech processed through an acoustic simulation of a cochlear implant. This finding also has implications for cochlear implant users. Collison and colleagues (2004) recently found that adult cochlear implant users were not as sensitive to lexical information when completing some speech perception tasks as normal-hearing adults. This result suggests that these listeners may not only be at a disadvantage because they are exposed to degraded speech but also because they fail to completely use important linguistic knowledge. Thus, cochlear implant users may also benefit from training programs designed to enhance their ability to make use of lexical, semantic, and other important cues when they are available in the speech signal.

Conclusions

In conclusion, the present results show that both the type of training stimuli and feedback used contribute to the ability to learn, remember, and generalize learning to new signals processed through an acoustic simulation of a cochlear implant. Feedback that enabled listeners to read the content of the stimuli while they rehearsed the degraded token was almost universally better than auditory-alone feedback. Degraded auditory and written feedback may be more effective because it makes the spoken message “pop out” from the degraded signal (Davis et al., 2005). “Pop out” may help listeners become more successful at identifying spectrally degraded speech because it reinstates the conditions of original encoding and learning while simultaneously enabling accurate retrieval of the identity of the speech (e.g. Craik, 1994; Lockhart, 2002; Morris et al., 1977; Neath & Surprenant, 2003; Rajaram et al., 1998). The effectiveness of the processed + text feedback is encouraging because it can be implemented with cochlear implant users. Thus, the findings concerning feedback effects in this study may have implications for the development of clinical treatments for cochlear implant patients.
Finding that training with anomalous sentences can be as or even more successful than training with meaningful sentences also has implications for how cochlear implant users could be trained to understand speech when important semantic and lexical cues are unavailable. Theoretically, the results of this study reconfirm earlier findings that semantic information is important to the top-down processing skills used in identifying speech. However, semantic information is not necessary for listeners to learn to understand degraded speech more accurately. In fact, several results of this study suggest that learning to understand speech processed through an acoustic simulation of a cochlear implant may even be more robust when the training stimuli lack semantic context. Learning to understand spectrally reduced speech may be more robust when semantic cues are removed, because under such circumstances, listeners will need to rely more heavily on using the degraded acoustic-phonetic signal to identify words. Directing the listeners’ conscious attention to the phonological properties of the speech signal may make it easier to generalize learning to new types of stimuli than traditional training methods which may rely more on interpreting higher-level semantic and lexical cues.
CHAPTER III: PERCEPTUAL ADAPTATION AND GENERALIZATION OF LEARNING

Introduction

The perceptual learning processes involved in perceiving novel and degraded speech signals have been examined using a variety of different manipulations and methods to degrade or alter speech stimuli. Normal-hearing listeners have been found to be able to recognize speech and speakers when the speech signal is synthesized (Cooper, Delattre, Liberman, Borst, & Gerstman, 1952; Daffy & Pisoni, 1992; Pisoni, Manous, & Dedina, 1987), reversed (Sheffert, Pisoni, Fellowes, & Remez, 2002), time-compressed (Altmann & Young, 1993; Dupoux & Green, 1997), “morphed” (Specht, Rimol, Reul, & Hugdahl, in press), and recreated using time-varying sinusoidal patterns (Remez, Rubin, Pisoni, & Carrell, 1981; Sheffert et al., 2002). Taken together, these findings suggest that speech is a very robust signal because it can be recognized even when the natural vocal quality is eliminated and important phonetic properties are temporally distorted.

Recently, interest in listeners’ abilities to understand speech based primarily on temporal cues rather than on intact spectral information has also developed (e.g. Davis, Johnsrude, Harvais-Adelman, Taylor, & McGgettigan, 2005; Eisenberg, Shannon, Martinez, Wygonski, & Boothroyd, 2000; Shannon, Fu, & Galvin, 2004; Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995). Spectrally reduced speech that retains temporal cues is a particularly interesting form of degraded speech to examine, because it can be used as a model of what deaf cochlear implant users might hear through their devices. Thus, studying normal-hearing listeners’ abilities to perceive spectrally degraded speech may provide valuable new clues about how deaf individuals learn to understand the novel auditory stimuli transmitted through their device.

Although cochlear implants have often been found to be an effective and remarkable method of treating profound deafness in both adults and children, they do not simply restore hearing to a normal acoustic form (e.g. Gantz, Tyler, Woodworth, Tye-Murray, & Fryauf-Bertschy, 1994; Lyxell, Andersson, Andersson, Arlinger, Bredberg, & Harder, 1998; Miyamoto, Kirk, Robbins, Todd, Riley, & Pisoni, 1997; Waltzman, Cohen, Gomolin, Green, Shapiro, & Hoffman, 1997). Rather, the auditory signals they provide to listeners are based on electrical input and are highly degraded compared to the intact and original acoustic speech signals (Clark, 2002; Dorman & Wilson, 2004).

The auditory signals generated through cochlear implants are highly degraded, because the original acoustic signals from which they are derived are processed and reduced through several steps. Figure 3.1 shows a simple schematic illustration of a cochlear implant device implanted into the auditory system. Acoustic stimuli detected by an external microphone are first channeled into a speech processing system through a small radiofrequency transmitting cord that connects the two components. The speech processor then filters the acoustic signal into multiple and contiguous frequency bands (Dorman & Wilson, 2004).

An implanted receiver is responsible for converting the transmitted signal into electrical currents. Thus, the cochlear implant mimics the signal transduction process that intact and viable hair cells would normally undertake. This signal transduction process begins when small changes in the amplitude of the signal, collectively known as the amplitude envelope, are detected and extracted in each frequency channel. Using this amplitude envelope, the spectral information in each channel is then converted into amplitude-modulated electrical pulses. These pulses are then transmitted to the spiral ganglion cells of
the cochlea and auditory nerve through an array of active electrodes (Balkany, Hodges, Miyamoto, Gibben, & Odabassi, 2001; Dorman & Wilson, 2004).

**Figure 3.1.** Illustration of cochlear implant device and pathway of auditory stimulation. The elements of the device and auditory pathway include: 1) microphone/speech processor 2) transmitting coil 3) external transmitter 4) internal receiver/electrode array 5) auditory nerve 6) cortex.

Although the signal processing schemes of cochlear implants are designed to preserve acoustic properties important for perceiving speech and are able to filter the speech signal into as many as 24 frequency bands, listeners may still receive relatively uninterpretable signals for a variety of reasons. Sparse spiral ganglion cell survival, channel or electrode interactions, and the depth of electrode insertion into the cochlea are some factors that may contribute to the large individual differences in the intelligibility of auditory signals (e.g. Blamey, Pyman, Gordon, Clark, Brown, Dowell et al., 1992; Donaldson & Nelson, 2000; Dorman & Wilson, 2004). Electrodes are often only implanted two-thirds of the way into the structure or to a depth of around 15 mm to 20 mm (Clark, 2003). Because the cochlea is tonotopically arranged for frequency coding, the inability to implant electrodes deep enough or in a uniform spatial arrangement can adversely impact cochlear implant users' perception of frequency.

As a result of shallow electrode insertion, epical neurons tuned for some frequencies lower than 1000 Hz may go unstimulated. Because much of this low-frequency information is important for speech perception, the electrical currents generated from the low frequency channels are often directed to tonotopic areas tuned for slightly higher frequencies instead. This mismatch in frequency and electrode placement has been hypothesized to be one aspect of cochlear implantation that could make speech particularly hard for post-lingually deaf cochlear implant users to understand (Blamey, Dooely, Parisi, & Clark, 1996; Dorman, Loizou, & Rainey, 1997; Fu & Galvin, 2003; Fu & Shannon, 1999a, b; Harnsberger, Svirsky, Kaiser, Pisoni, Wright, & Meyer, 2001; Rosen, Faulkner, & Wilkinson, 1999; Svirsky, Silveira, Neuburger, Teoh, & Suarez, 2004). A basalward frequency shift may be detrimental to post-lingually deaf cochlear implant users because it is one additional variable that makes electrically generated speech sound remarkably different than the acoustically generated speech the listeners heard prior to going deaf.
Because many post-lingually deaf cochlear implant users are exposed to spectrally degraded and frequency shifted auditory signals after having normal hearing for some period of time before the onset of their deafness, they are a unique and theoretically interesting population in which to examine auditory perceptual learning. In addition to a general interest in how cochlear implant users may learn to understand the degraded auditory input generated through their devices (Clark, 2002; Davis et al., 2005), a specific interest in cochlear implant users' abilities to adapt to the frequency shifts generated through their devices has recently emerged (Fu & Shannon, 1999a, b; Fu, Shannon, & Galvin, 2002; Harnsberger et al., 2001; Svirsky et al., 2004). Investigations of cochlear implant users' abilities to adapt to frequency shifted speech have provided some insights into how spectral mismatch affects listeners' speech perception performance and have also suggested possible theoretical explanations of auditory perceptual learning in cochlear implant users.

Adaptation to Frequency Shifts by Cochlear Implant Users

The normal auditory system can accommodate up to a 3-mm frequency shift during recognition or categorization of speech stimuli processed by an acoustic simulation of a cochlear implant due to variations in indexical properties of speakers such as age and gender (Fu et al., 1999). Therefore, cochlear implant users that have a mapping strategy that falls within 3 mm of normal tonotopic arrangement may adapt relatively quickly. In contrast, a spectral and tonotopic mismatch in excess of 3 mm will likely be more difficult and time consuming for listeners to adjust to and maintain perceptual constancy.

Fu and Shannon (1999a) evaluated cochlear implant users' vowel recognition scores when using stimuli preprocessed to emulate both small and large frequency shifts. They found that listeners performed best on stimuli that most closely matched their current clinical speech processing algorithm. However, there was evidence that the cochlear implant users were able to adapt to small frequency shifts that were equivalent to or less than a 3-mm longitudinal shift along the cochlea. This result is consistent with the proposal that it should be no harder for cochlear implant users to adjust a 3-mm frequency shift than it is for normal-hearing listeners to adjust to a child's voice or a female's voice after hearing a male voice. Based on the results of their study, Fu and Shannon (1999a) proposed and discussed several hypotheses for how cochlear implant users may learn to understand spectrally mismatched speech through their device.

The "accommodation hypothesis" suggests that, through experience alone, listeners may slowly accommodate the inappropriate pairings of spectral information and tonotopic location and form their own unique internal auditory representations of speech. These new auditory representations would be derived from an "adapted electric map" of the spectral and tonotopic pairings. Alternatively, the "tonotopic matching hypothesis" suggests that these internal auditory representations of speech are correctly formed only when there is a relatively accurate match between spectral information and tonotopic location. Thus, this hypothesis suggests that speech will be best understood when cochlear implant users have frequency-to-electrode placement that closely mimics a normal acoustic or tonotopic map (Fu & Shannon, 1999a).

Because the cochlear implant users in Fu and Shannon's (1999a) study did not perform any worse on small frequency shifts, the accommodation hypothesis seems to more accurately reflect how listeners are able to understand speech after receiving a cochlear implant (Fu & Shannon, 1999a). However, because their study was conducted over a short testing session, it is not certain whether there were truly limits to cochlear implant users' abilities to adapt to larger frequency shifts or whether simple time constraints prevented them from fully adapting. Longitudinal perceptual learning studies are clearly
necessary to determine more accurately whether large frequency shifts are too difficult for cochlear implant users to adapt to.

In another study, Fu and colleagues (2002) examined cochlear implant users’ ability to adapt to frequency shifts over an extended period of time by experimentally altering their mapping or processing strategies. The settings on cochlear implants can be non-invasively modified to change a variety of parameters such as stimulation rate, number of active electrodes, and frequency-to-electrode assignment. Fu and colleagues examined cochlear implant users’ abilities to adapt to frequency shifts by fitting them with speech processors that shifted the normal frequency-to-electrode assignment of their devices basally by 2 to 4 mm. These processors were worn continuously by the patients for a three-month period.

The speech perception performance of these patients was monitored for the three months immediately following the introduction of the experimental shift. Fu and colleagues (2002) observed a gradual improvement in listeners’ abilities to identify consonants and sentences taken from the Hearing in Noise Test (HINT). However, the listeners’ abilities to identify vowels and transcribe a second set of sentence stimuli when listening to the novel shift failed to approach baseline performance. These results suggest that there may be limits to cochlear implant users’ abilities to adapt fully to frequency shifts that exceed the amount of shift that they are normally accustomed to. Even after three months of consistent exposure to a new mapping strategy, listeners’ speech perception scores remained lower than they were prior to remapping.

Although these previous studies did determine the degree to which cochlear implant users could accommodate frequency shifts and also quantitatively assessed the decrease in speech perception of large frequency shifts, they did not explore in depth how frequency shifts may qualitatively affect cochlear implant users’ speech perception skills. In order to examine the impact of frequency shifts on speech perception more qualitatively and to determine the extent to which frequency shifts could be adapted to, Harnsberger and colleagues (2001) measured the perceptual vowel spaces of experienced cochlear implant users. No evidence of abnormally shifted vowel spaces was found in the large majority of these listeners. These results indicate that the patients may have fully adapted to the frequency shift caused by their device.

However, although the relatively normal vowel spaces of these cochlear implant patients indicate that they adapted to the frequency shift, they do not indicate when or how this adaptation may have occurred. The cochlear implant patients in Harnsberger and colleagues’ (2001) study already had at least one year of experience with their cochlear implant. Therefore, they likely adapted to the frequency shift sometime during their first year of cochlear implant use.

To more accurately determine the time course of perceptual learning and adjustment to a frequency shift in cochlear implant users, Svirsky and colleagues (2004) conducted a longitudinal examination of cochlear implant users’ vowel spaces. Vowel space measurements were made using a method-of-adjustment procedure in which listeners matched synthetic vowels to a target vowel. Vowel spaces for the cochlear implant users were first measured on the day their implant was activated. The initial measurements were also made before the patients heard any speech sounds. Thus, the initial reports of vowel space in this sample of cochlear implant users provided a better assessment of how a frequency shift may initially affect vowel perception and categorization without any experience using the implant.
Svirsky and colleagues (2004) found that the initial vowel spaces of the cochlear implant patients were atypical compared to normal-hearing listeners. However, the atypical patterns observed were not consistent among listeners and were not compatible with the assumption of a basalward frequency shift. Although there was an apparent mismatch between listeners’ vowel categories and the electrical stimulation that they were receiving, this mismatch was not specifically attributable to a basalward frequency shift.

Despite being unable to confirm the cause of the cochlear implant patients’ atypical vowel spaces, Svirsky and colleagues (2004) did determine that the patients eventually learned to categorize vowels more normally. After repeatedly completing the vowel categorization task throughout a two-year period without receiving feedback, most listeners had developed near normal vowel spaces. However, the amount of time it took for the listeners to adjust their vowel spaces varied greatly.

Several of the cochlear implant patients in Svirsky’s and colleagues’ study (2004) showed evidence of normalized vowel spaces in as little as one day to one month after initial stimulation. For other listeners, the adaptation process took nearly the entire two-year period. Evidence of such a large range of variability in the time course of perceptual learning and adaptation to frequency shifts in these cochlear implant users suggests that the processes involved in learning how to recognize speech sounds could be particularly informative and interesting to examine in cochlear implant users. Examination of the auditory perceptual learning processes used by cochlear implant patients to adapt to frequency shifts may have important clinical implications for understanding why cochlear implant users often vary considerably on other measures of speech perception and spoken word recognition (e.g. Blamey, Pyman, Gordon, Clark, Brown, et al., 1992; Pisoni, 2000; Pisoni, Cleary, Geers, & Tobey, 2000; Sarant, Blamey, Dowell, Clark, & Gibson, 2001).

**Perceptual Adjustment to Frequency Shifts in Acoustic Simulations of Cochlear Implants**

Research using normal-hearing listeners exposed to acoustic simulations of cochlear implant could also provide valuable new insights into how cochlear implant users initially encode degraded speech and how they may adapt to new stimuli and signal processing strategies. In addition to examining cochlear implant users’ abilities to identify speech with a variety of different frequency shifts, Fu and Shannon (1999a) also assessed normal-hearing listeners’ abilities to identify speech processed through acoustic simulations of cochlear implants that included frequency shifts. They found that normal-hearing listeners performed similarly to cochlear implant users. Both normal-hearing listeners and cochlear implant patients maintained their baseline levels of performance even when small frequency shifts of 3 mm and under were introduced.

In addition, both normal-hearing listeners and cochlear implant patients demonstrated a large decline in performance as the frequency shift was increased beyond 3 mm. This result suggests that large frequency shifts may be too difficult for listeners to adapt to easily. However, neither the normal-hearing listeners nor cochlear implant users received any feedback during testing or any form of training. Thus, it is unclear whether these listeners were unable to adapt to large frequency shifts because of fundamental auditory system or encoding constraints or because the appropriate conditions for auditory perceptual learning were simply not made available to them.

When training is provided to listeners, it does appear that normal-hearing listeners exposed to acoustic simulations of a cochlear implant can eventually adapt to relatively large frequency shifts. Rosen and colleagues (1999) measured normal-hearing listeners’ abilities to identify words in sentences processed through four-channel cochlear implant simulations that had either no frequency shift or a 6.5-
mm frequency shift. Results indicated that listeners were able to identify over 60% of key words in sentences correctly without a frequency shift. However, when a 6.5-mm shift was introduced, the listeners' key word recognition scores plummeted to only 1% correct.

Rosen and colleagues (1999) then tested if listeners were able to identify speech processed with a large frequency shift more accurately if they received training on that frequency shift. During nine 20-minute training sessions, the authors used interactive continuous discourse tracking and on-line signal processing to train the listeners to better understand the frequency shifted speech (DeFillippo & Scott, 1978). Continuous discourse tracking is an interactive procedure in which listeners work with a speaker or "trainer" to understand speech spoken in live-voice. Listeners simply report what they think the speaker said. Any incorrect portions of the utterance are repeated by the speaker until the listener reports them correctly.

In the study by Rosen and colleagues (1999), the trainer's speech was processed through the acoustic simulation of the cochlear implant online and transmitted to the listeners. The listeners' task was to repeat the sentence. The trainer monitored each listener's repetitions by confirming words repeated correctly and repeating the words that were still incorrect. After this form of training, listeners' abilities to recognize consonants, vowels, and words in sentences improved. The results of this study suggest that, although introducing a large frequency-shift initially impairs speech perception performance, frequency-shifted speech may not remain entirely unintelligible if normal-hearing listeners are given enough training and experience with the frequency-shifted signal.

The method of training used to help normal-hearing listeners adapt to spectrally degraded and frequency shifted speech has also been found to influence listeners' speech perception and perceptual learning skills. Gradual and incremental training with spectral mismatch has been found to help normal-hearing listeners adapt more quickly to speech that is processed using a large frequency shift (Svirsky, Sinha, Neuburger, & Talavage, 2003). Svirsky and colleagues found that normal-hearing listeners who were initially trained with no frequency shift and then gradually trained on larger and larger frequency shifts performed better on vowel, consonant, and sentence recognition tests than listeners who were only ever trained on a large 6.5-mm frequency shift.

The course of treatment for actual cochlear implant users is similar to the normal-hearing listeners trained only on the large frequency shift in Svirsky's and colleagues' study (2003). That is, most cochlear implant patients may initially be exposed to some degree of frequency shift rather than a correct frequency-electrode assignment. Mismatches between frequency and electrode placement are tolerated in cochlear implantation in order to preserve low frequencies important for speech perception. Thus, finding that listeners trained exclusively on one large frequency shift show slower speech perception gains than listeners trained with incremental shifts may have important clinical implications.

The results of Svirsky's and colleagues' (2003) incremental training study suggest that it may actually be more useful to expose cochlear implant users to spectrally mismatched speech gradually rather than immediately after implantation. By initially using no frequency shift and then gradually introducing it later by remapping frequency to electrode assignment, cochlear implant users may have a better opportunity to understand speech in its less complex spectrally degraded form. Gaining fundamental knowledge and specific experiences with the spectrally degraded speech signal through training with no frequency shift could feasibly bootstrap perceptual adjustment when frequency shifts are introduced at a later time.
Taken together, previous research examining auditory adaptation to frequency shifts in cochlear implant users and normal-hearing listeners exposed to acoustic simulations of cochlear implants indicates that training can successfully help listeners adapt to frequency shifts. Thus, perceptual learning using gradual and incremental training may be one of the most efficient ways for listeners to understand frequency shifted speech (Svirsky et al., 2003). However, listeners could potentially be more efficient at understanding frequency shifted speech if perceptual learning for speech processed with no frequency shift or a small frequency shift simply generalized to larger frequency shifts.

If training on spectrally degraded speech with no frequency shift or a small frequency shift generalizes to speech processed with a large shift, additional incremental training on intermediate shift conditions may not even be necessary. Instead, if training with no frequency shift or small frequency shift conditions is robust and generalizable to large frequency shift conditions, shorter training periods with fewer intermediate frequency shift conditions may be sufficient for listeners to adapt to a larger frequency shift. The generalizability of perceptual learning of spectrally degraded speech to spectrally mismatched or frequency shifted speech is clearly an important research problem that has not yet been fully explored in cochlear implant patients or normal-hearing listeners exposed to acoustic simulations of cochlear implants (Fu & Galvin, 2003).

**Generalization of Perceptual Learning to Spectral Mismatch**

In addition to the clinical implications, theoretically important findings about auditory perceptual learning may be uncovered when examining cochlear implant users’ or normal-hearing listeners’ abilities to generalize learning to frequency shifts. Generalization is an important aspect of auditory perceptual learning to investigate, because it can provide a good indication of how robust learning really is (Karmarkar & Buonomano, 2003; Moore, Amitay, & Hawkey, 2003). In addition, examinations of listeners’ abilities to generalize learning to new auditory stimuli may help determine what they learn about the auditory signal on which they were trained and what properties carry over to generalization and transfer conditions.

Generalization of learning in cochlear implant users has most often been examined indirectly through studies using a variety of test materials such as consonants, vowels, isolated words, and sentences (Burkholder, 2005; Davis et al., 2005). However, studying adaptation to frequency shifts in cochlear implant users provides a unique opportunity to assess generalization of learning to new auditory signals and signal processing as well. By examining generalization of learning to new auditory signals and signal processing parameters that violate the normal frequency-to-place mapping in the cochlea, clues to understanding how cochlear implant speech is encoded and learned by both normal-hearing and deaf listeners may be uncovered.

An ability to generalize learning to new frequency shifts by cochlear implant users or normal-hearing listeners’ exposed to acoustic simulations of cochlear implants suggests that they may learn to identify degraded speech using abstract information. In contrast, failure to generalize learning to new frequency shifts suggests that listeners may learn to identify speech through their device based on the specific characteristics of the training stimuli. Examinations of the generalization of learning may be useful to help determine more conclusively what the limitations of perceptual learning and the auditory system really are.

Previous findings suggest that cochlear implant users’ speech perception skills and perceptual learning abilities may show some specificity or advantage for stimuli that most closely resemble what they have experience with (Fu & Shannon, 1999a). However, cochlear implant users also appear to be
able to generalize what they have learned about speech signals through everyday use with their speech processors to small but novel frequency shifts (Fu & Shannon, 1999a; Fu et al., 2002). Thus, evidence of both specificity for and generalization to spectral mismatch has been found in cochlear implant users. This apparently contradictory set of findings should be explored further in studies that utilize training and generalization tasks to assess frequency shift accommodation in cochlear implant users.

It is also unclear whether normal-hearing listeners trained to understand spectrally reduced speech can generalize learning to speech processed with varying degrees of spectral mismatch. Although evidence of generalization to different stimulus materials such as meaningful and anomalous sentences has been found in normal-hearing listeners trained with spectrally degraded speech (Burkholder, 2005; Davis et al., 2005), these findings do not reveal the specific acoustic properties of the acoustic signal listeners may be attuned to. Examinations of listeners' abilities to generalize learning among different frequency shifts or different forms of signal processing are necessary to determine more clearly what listeners become most attuned to when being trained with acoustic simulations of cochlear implants.

In a recent study, normal-hearing listeners' abilities to generalize learning to a different vocoder carrier signal in an acoustic simulation of a cochlear implant were examined (Davis, Taylor, Johnsude, & Carlyon, 2004). Using acoustic models of cochlear implants with noise-band and pulse-train carriers, Davis and colleagues tested the generalization of learning from one degraded auditory signal to another by training listeners on one condition and then testing them on a novel condition. Results indicated an asymmetric pattern of generalization. Listeners who received training on noise-vocoded speech were able to generalize what they learned to the pulse-vocoded speech. However, listeners trained on pulse-vocoded speech did not appear to generalize what they learned to noise-vocoded speech. Thus, it appears that the robustness of auditory perceptual learning may depend on what the specific characteristics of the training stimuli are and what was learned about the auditory signals during training.

Although the carrier signals used in acoustic simulations of cochlear implants have been found to interact with normal-hearing listeners' generalization abilities, other variations in signal processing appear to have more consistent effects on the robustness of learning. In another recent study, listeners were able to generalize learning among acoustic simulations of cochlear implants that were processed using different frequency filter ranges (Davis, Hervais-Adelman, Taylor, Carlyon, & Johnsude, 2005). Listeners trained on noise-band processors filtered into low frequency regions (50 to 1406 Hz) generalized their learning to stimuli filtered into high frequency regions (1593 to 8000 Hz) and vice versa. Thus, it appears that auditory perceptual learning with an acoustic simulation of a cochlear implant is not specific to the particular frequency ranges in which the stimuli were filtered as long as analysis filters and noise bands are frequency matched.

Research by Davis and colleagues (2004; 2005) that examined generalization of perceptual learning with acoustic simulations of cochlear implants is theoretically interesting, because they suggest that listeners may primarily learn from the general properties of degraded speech rather than from specific aspects of the training stimuli when listening to noise-vocoded speech. However, this example of robust perceptual learning is not necessarily relevant to cochlear implant users. The frequency information available to cochlear implant users is not as limited as the frequency filters used by Davis and colleagues (2005) to examine generalization of learning. In addition, measuring generalization of learning between different carrier signals (Davis et al., 2004) may not be clinically applicable to cochlear implant users, because the design of a cochlear implant currently confines the method of stimulation to electrical pulses.
However, frequency-to-electrode mapping can be reconfigured in cochlear implant users. Thus, frequency shifts may be more appropriate variables with which to study the robustness of learning in cochlear implant users. In addition, normal-hearing listeners trained on speech processed through an acoustic simulation of a cochlear implant should also be tested to determine their abilities to generalize perceptual learning among different kinds of frequency-shifted speech.

Recently, Fu and Galvin (2003) carried out a study using 20-channel acoustic simulations of cochlear implants with basalward and apical (upward and downward) frequency shifts. Listeners were trained on sentences shifted basally and then tested on the trained shift in addition to a variety of novel basalward and apical frequency shifts. Thus, in this study, listeners' abilities to generalize learning to novel frequency shift conditions were tested. The authors suggested three alternatives for how listeners would perform during testing.

First, listeners may improve on frequency shifts taken from the shift class on which they were trained and perform worse on all frequency shifts from the untrained category. This result would suggest that a bias towards the trained shift was created during perceptual learning that resulted in the differentiation of the familiar frequency shift from novel ones. Such a perceptual bias may be due to the formation of new auditory representations of speech based specifically on some aspect of the trained frequency shift (Fu & Galvin, 2003). A perceptual bias may also be a result of encoding specificity for the trained frequency shift.

Alternatively, listeners' performance on the novel shift class may remain unchanged as performance on the trained shift improves. According to Fu and Galvin (2003), this result would suggest "local adaptation" without a specific bias or encoding specificity to the training stimuli. During local adaptation, a bias to the training stimuli may be avoided because old auditory representations of speech are retained.

The prior two sets of results would indicate that generalization to novel shift classes does not occur and that auditory perceptual learning is highly specific to the training shift category. In contrast, if listeners improve on stimuli processed with basalward and apical frequency shifts, generalization of learning has presumably occurred. Such a result would suggest that adaptation is more global in nature and occurs because training on a specific shift helps listeners learn to process and encode the basic acoustic properties of spectrally degraded speech (Fu & Galvin, 2003).

Fu and Galvin (2003) found that, after training with basally shifted speech, listeners performed better than baseline on stimuli that were upwardly shifted but did no better on stimuli that were downwardly or apically shifted. Listeners also showed the greatest improvement in the shift condition on which they were trained. More recently, Fu, Nogaki, & Galvin (2005) found evidence that listeners trained with a large frequency shift also do not generalize their learning to spectrally degraded speech that is unshifted. Taken together, the results of these studies indicate that learning may have been specific to the properties of the shift rather than to the general qualities of spectral degradation that were present in both upwardly and downwardly shifted speech.

In Fu and Galvin's earlier study (2003), listeners were also retrained on downward shifts and were then retested on all the shift conditions. Improved performance was observed for the downward shift conditions. However, listeners' performance on upwardly shifted speech showed no decrement after they were retrained on downwardly shifted speech. Thus, it appears that listeners retained what they learned during the first period of training despite being retrained on a different signal and showing prior specificity of learning to one method of shift.
Overall, these previous studies examining speech perception by normal-hearing listeners exposed to acoustic simulations of cochlear implants provide a complicated and somewhat contradictory account of the robustness of auditory perceptual learning to new signal processing conditions (Davis et al., 2004; Fu & Galvin, 2003). Although evidence of generalization between different signal carriers has been obtained, research specifically examining generalization of learning between and within upward and downward frequency shifts suggests that listeners' generalization abilities may be more limited. Instead of finding widespread generalization among frequency shifts, Fu and his colleagues (2003; 2005) found evidence of specificity of learning. Failure to generalize learning to a different class of frequency shift or to no frequency shift indicates that listeners may encode and learn from specific properties of the degraded speech signal rather than learn primarily about the more general properties of the processed speech signals.

Because the research examining generalization of learning to frequency shifts in normal-hearing listeners exposed to acoustic simulations of cochlear implants has been somewhat limited and has produced conflicting results, additional examinations of this topic should be conducted (Davis et al., 2004; Fu & Galvin, 2003). In addition, previous studies examining normal-hearing listeners' abilities to generalize learning to new cochlear implant simulation strategies have not considered all relevant aspects of adapting and generalizing learning to frequency shifts. Although Fu and Galvin (2003) assessed normal-hearing listeners' abilities to generalize learning from large apical and basalward shifts to no shift and numerous other shift conditions, they did not determine if training with spectrally matched but degraded speech would generalize to speech that was processed with various degrees of frequency shift.

Svirsky and colleagues (2003) found that incremental training that began with no frequency shift and gradually progressed to a large 6.5-mm frequency shift resulted in more rapid learning than training that began immediately with the large frequency shift. However, it is possible that if listeners have the ability simply to generalize learning from training with no-shift or small-shifts to larger shifts, the adaptation process would progress even more rapidly. In the current study, normal-hearing listeners' abilities to generalize learning to a small and large basalward frequency shift after being trained without a frequency shift were examined using an 8-channel acoustic simulation of a cochlear implant. In addition, similar to Fu and Galvin's (2003) study, listeners trained with a frequency shift were tested on a novel frequency shift and no shift condition to assess generalization of auditory perceptual learning and to determine the robustness of training with a frequency shift.

The current study has direct clinical relevance because it may provide further support for altering the initial mapping strategies that cochlear implant patients receive (Svirsky et al., 2003). Currently cochlear implant users are first exposed to auditory signals with some degree of spectral mismatch in order to preserve low frequency information that is important for speech perception. However, if training with no frequency shift generalizes to frequency shifts in normal-hearing listeners, cochlear implant patients may also benefit from training with spectrally degraded but matched signal processing strategies immediately after implant activation and prior to receiving a frequency-to-electrode mapping strategy that implements a frequency shift.

In addition, the present investigation may be more clinically applicable than previous studies examining generalization of learning with an acoustic simulation of cochlear implants. In one previous examination of generalization of learning after training with an acoustic simulation of a cochlear implant, generalization among carrier signals (e.g. noise bands, sinewaves, pulse trains), which remain constant in cochlear implants, was examined rather than generalization to frequency shifts (Davis et al., 2004). In an earlier investigation of generalization of learning among frequency shifts, a 20-channel acoustic model of a cochlear implant was used (Fu & Galvin, 2003).
A 20-channel acoustic model of a cochlear implant may provide more spectral information to the normal-hearing listeners than many cochlear implant patients actually receive and successfully use. Previous research has found that adult cochlear implant users typically achieve asymptotic speech perception scores with eight active channels (Fishman, Shannon, & Slattery, 1997; Friesen, Shannon, Baskent, & Wang, 2001). In addition, normal-hearing listeners' speech perception performance is most similar to cochlear implant users’ speech perception performance when eight channels are used in an acoustic model of a cochlear implant (Dorman, Loizou, Fitzke, & Tu, 2000). Thus, using an 8-channel acoustic model may be a more appropriate method to assess generalization of learning among unshifted and shifted speech that is spectrally degraded.

The present study also has theoretical significance and may provide complementary information to earlier studies examining normal-hearing listeners’ abilities to generalize learning among different methods of signal processing (Davis et al., 2004; Fu & Galvin, 2003). Testing generalization among frequency shifts should provide new insights about what listeners actually become attuned to and what they learn during training with the degraded speech processed by a cochlear implant simulation. Testing generalization among frequency shifts should reveal what properties of speech listeners are sensitive to. Thus, the goal of the current study is to test normal-hearing listeners’ abilities to generalize auditory perceptual learning of spectrally degraded speech to different degrees of spectral mismatch.

Hypotheses and Predictions: Encoding Specificity and Generalization of Learning

Based on earlier work by Rosen (1999), Svirsky (2003), and their colleagues we expected that listeners trained on a large 6.5-mm frequency shift would perform the worst during training. However, the predictions for how each listener group will perform in the generalization phases are less clear. Previous studies testing listeners’ abilities to recognize spectrally degraded and frequency shifted speech have found evidence of both specificity and generalization of learning (Davis et al., 2004; Fu & Galvin, 2003; Fu et al., 2005). Given the variability in the findings from previous studies examining adaptation to spectrally mismatched speech, there are several feasible outcomes in the current study.

Encoding Specificity. One possible result is that listeners will show complete encoding specificity to the frequency shift on which they were trained. If learning is specific to the trained frequency shift, listeners’ performance on the two novel shifts should be significantly worse than performance during training. In addition, listeners should also perform significantly better on the trained frequency shift than on the novel frequency shifts during the generalization period. Such a pattern of performance is also an indication that no generalization to novel frequency shifts occurred.

However, encoding specificity may be influenced by the amount of frequency shift listeners receive during training. It is predicted that listeners trained with the large 6.5-mm frequency shift will not demonstrate encoding specificity. Encoding specificity is not expected in this group of listeners because speech processed with a 6.5-mm shift sounds highly degraded and unnatural. Thus, speech processed with a large frequency shift will be harder for listeners to encode in general. In addition, previous studies have shown that it is difficult for normal-hearing listeners to learn to understand spectrally degraded speech with a large basalward frequency shift (Rosen et al., 1999; Svirsky et al., 2003). In contrast, listeners trained with no frequency shift or the small 3-mm frequency shift should be able to encode the stimuli on which they were trained more accurately and demonstrate encoding specificity.

Generalization. An alternative result of the present study is that listeners will show generalization to novel frequency shifts. If generalization of learning does occur, listeners should perform no worse on sentences processed with a novel shift during the generalization period than on sentences
presented during the training period. Some degree of generalization is expected in all listener groups in this study.

Based on findings by Fu and colleagues (2002), listeners trained with no frequency shift should generalize learning to the smaller 3-mm frequency shift. However, a more interesting question is whether these listeners will also generalize learning to the larger 6.5-mm frequency shift. Evidence of generalization to a larger frequency shift in normal-hearing listeners trained with an acoustic simulation of a cochlear implant may suggest that cochlear implant users could benefit from experiencing spectrally matched mapping strategies before receiving mapping strategies expanded to include low frequency information and a frequency shift.

Although not as clinically important or interesting, the generalization abilities of the other listener groups may also provide insight into the robustness of auditory perceptual learning in general. Listeners trained on the 6.5-mm shift may show more generalization to the two novel shift conditions than the other two listener groups because the novel stimuli will be less distorted than the stimuli on which they were trained. Listeners in the intermediate shift condition may also show robust generalization because their training stimuli do not differ much from the two novel conditions. Because the 3-mm frequency shift condition is perceptually close to the two novel conditions, the changes in shift may be less drastic to listeners trained on an intermediate 3-mm shift and easier for them to generalize to.

Taken together, these two factors, the difficulty of the training stimuli and the perceptual difference or “distance” between the training and test stimuli, are expected to play a role in the listeners' abilities to generalize learning to new frequency shifts. Generalization of learning should be more robust when listeners are tested on materials that are less degraded than the training materials. In addition, listeners may also generalize learning more effectively to the novel stimulus set that is most perceptually similar to the stimuli on which they were trained.

**Combined Specificity and Generalization.** Another possible outcome is that individual listener groups will show some evidence of both encoding specificity and generalization of learning. Listeners who perform the best on their trained shift during the generalization phase may not necessarily have word-recognition scores for novel shifts that are significantly worse than word-recognition scores obtained during training. In other words, it is possible that performance on new frequency shifts will be equal to training while performance on the trained shift will exceed performance in all other conditions. Such a result may provide evidence that generalization and specificity of learning are not necessarily mutually exclusive as previous research has suggested (Fu & Galvin, 2003).

**Experiment I**

**Method**

**Experimental Design**

Figure 3.2 shows a detailed schematic representation of the experimental design used for Experiment I. The experimental procedure was divided into five phases: (1) Familiarization, (2) Pre-training, (3) Training, (4) Post-training, and (5) Generalization. The specific procedures carried out during these five phases will be described in detail in the Experimental Procedures section. Three independent groups of participants were used. The groups differed according to the amount of frequency mismatch used in the acoustic model of the cochlear implant during training.
(1) Familiarization (2) Pre-test (3) Training Phase (4) Post-test (5) Generalization

<table>
<thead>
<tr>
<th>Passive Listening</th>
<th>Sentence transcription</th>
<th>Sentence transcription training</th>
<th>Sentence transcription</th>
<th>Sentence transcription</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nursery rhymes with text</td>
<td>20 meaningful sentences</td>
<td>100 Harvard sentences</td>
<td>20 meaningful sentences</td>
<td>150 Harvard sentences</td>
</tr>
<tr>
<td>NO feedback</td>
<td>NO feedback</td>
<td>(1) No shift</td>
<td>(1) 50 trained shift</td>
<td>(1) 50 trained shift</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2) 3-mm shift</td>
<td>(2) 50 1st new shift</td>
<td>(2) 50 1st new shift</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3) 6.5-mm shift</td>
<td>(3) 50 2nd new shift</td>
<td>(3) 50 2nd new shift</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Processed stim + text given as feedback</td>
<td>NO feedback</td>
<td>NO feedback</td>
</tr>
</tbody>
</table>

**Figure 3.2.** Schematic representation of the experimental procedure used in the current study.

**Participants.** Each of the three groups of participants consisted of 20 young adults. Participants were recruited from a database containing email addresses of undergraduate students who had previously indicated in an online Psychology survey that they were interested in participating in paid studies. All participants were monolingual native speakers of American English and reported that they were free of any speech, hearing, language, and attentional disorders. Participants were paid $12 for their participation in this study.

**Stimuli and Materials**

**Stimuli.** Well known nursery rhymes (e.g. Jack and Jill, Humpty Dumpty) were used in the initial familiarization phase with the acoustic simulation. Twenty Harvard sentences were used in the pre- and post-training phases (IEEE, 1969). One-hundred Harvard sentences were used for training. One-hundred fifty Harvard sentences, split into three different blocks of 50, were used during the generalization phase. All Harvard sentences were between seven and 11 words long and contained five key content words that were scored during data analysis. The sentence stimuli used in the current study appear in Appendix B.

All stimuli were recorded by the author, a female speaker with a Midwestern dialect. Recordings were made in a sound-attenuated booth (IAC Audiometric Testing Room, Model 402A) using a Shure head-mounted microphone (SM98) placed approximately two inches from the mouth. Recordings were digitized online (16-bit analog-to-digital converter (DSC Model 240)) at 22,050 Hz and stored directly on a personal computer in .wav format using a version of a speech acquisition program (SAP; Dedina, 1987). Three tokens of each stimulus were recorded. The repetition judged to be of the highest quality by the experimenter was used in testing.

Using sound editing software (Adobe Audition) all usable sound files were segmented to the zero crossing closest to the beginning and end of the waveform. The root mean square (RMS) amplitude for all the files was then obtained. Based on the RMS amplitude measure, all stimuli were equated to an amplitude of 65 db(A) using the Level16 software program (Tice & Carroll, 1998).

In a pilot study, ten listeners were asked to transcribe 100 sentences spoken by the speaker used in this study to establish baseline speech intelligibility. Results showed that the keywords were identified correctly at 95% ($M = 94.72$, $SD = 1.23$) when listeners heard sentences produced by this speaker in quiet. This analysis indicates that the speaker has intelligible speech. This is an important finding,
because it indicates that any difficulties that listeners may have in identifying the speech when it is processed by the acoustic simulation of a cochlear implant are due to the nature of the signal degradation and not due to initially poor speaker intelligibility.

**Signal Processing.** All auditory stimuli were processed offline using a personal computer. The signal processing procedure used for the cochlear implant simulation was adapted from the real-time signal processing methods designed by Kaiser and Svirsky (2000). Three separate but similar signal processing strategies were used in order to generate stimuli varying in the amount of frequency shift.

For all three conditions, the signal was first lowpass filtered with a cutoff frequency of 12,000 Hz. A bank of eight bandpass analysis filters was then used to divide the auditory signal into eight frequency channels. The choice of an eight channel acoustic model was based on studies showing that adult cochlear implant users achieve asymptotic speech perception scores when they have this number of active channels (Fishman et al., 1997; Friesen et al., 2001). In addition, the best adult cochlear implant patients achieve open-set word identification scores similar to those obtained by normal-hearing listeners exposed to an 8-channel acoustic simulation of a cochlear implant (Dorman et al., 2000).

The temporal envelope was extracted from the output of the eight filters and then used to modulate noise bands created through a bank of eight synthesis filters. A basalward frequency shift was implemented in this acoustic model by using noise bands that had a higher frequency than the analysis filters. Different degrees of frequency shift were created by varying the amount of mismatch between the frequencies of the analysis and synthesis filters. In this experiment, frequency shifts representing 0 mm, 3 mm, and 6.5 mm of basalward misplacement along the length of the cochlea were used. The cutoff frequencies for the analysis and synthesis filter banks for each shift condition are listed in Table 3.1.

The frequency shifts generated through these signal processing methods were intended to model frequency shifts that may result after actual cochlear implantation. A frequency shift may result from cochlear implantation if electrodes are not placed far enough into the apical end of the cochlea which is tonotopically tuned for low frequency sounds. When electrodes cannot be implanted this deep and are instead placed more basalward within the cochlea, electrical stimulation for low frequency sounds is directed to neural fibers ordinarily tuned for higher frequency sounds. This mismatch may result in sounds being perceived at a slightly higher frequency than normal.

**Experimental Procedures**

Participants were tested in individual perceptual testing stations in the Speech Research Laboratory at Indiana University. Each booth was equipped with a Gateway PC (P5-133) using a 15” monitor (Vivitron15). Auditory stimuli were presented over calibrated headphones (Beyer Dynamics DT100) at approximately 70dB(A) SPL as measured by a standard sound-level meter. Stimuli were always presented in random order within blocks to the participants.

**Familiarization phase.** Testing began with a brief period of familiarization in which participants simply listened to and read silently along with six popular nursery rhymes that were processed using no frequency shift. The text of these nursery rhymes appeared on the top half of the monitor in a 30pt. sans serif font 500ms prior to the start of each utterance. The text of the nursery rhymes remained visible for the entire duration of the utterance. Listeners did not have the option to replay any of the familiarization stimuli. No explicit response was required during this phase.
Table 3.1. Frequency boundaries of the analysis and synthesis filters used to generate the eight frequency bands and the three different shift conditions of the acoustic model of the cochlear implant.

<table>
<thead>
<tr>
<th>Frequency Boundaries</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>No shift</strong></td>
</tr>
<tr>
<td><strong>Analysis Filters</strong></td>
</tr>
<tr>
<td>854 – 1467 Hz</td>
</tr>
<tr>
<td>1467 – 2032 Hz</td>
</tr>
<tr>
<td>2032 – 2732 Hz</td>
</tr>
<tr>
<td>2732 – 3800 Hz</td>
</tr>
<tr>
<td>3800 – 5150 Hz</td>
</tr>
<tr>
<td>5150 – 6622 Hz</td>
</tr>
<tr>
<td>6622 – 9568 Hz</td>
</tr>
<tr>
<td>9568–11000 Hz</td>
</tr>
<tr>
<td><strong>3-mm shift</strong></td>
</tr>
<tr>
<td>508 – 914 Hz</td>
</tr>
<tr>
<td>914 – 1287 Hz</td>
</tr>
<tr>
<td>1287 – 1749 Hz</td>
</tr>
<tr>
<td>1749 – 2454 Hz</td>
</tr>
<tr>
<td>2454 – 3346 Hz</td>
</tr>
<tr>
<td>3346 – 4319 Hz</td>
</tr>
<tr>
<td>4319 – 6266 Hz</td>
</tr>
<tr>
<td>6266 – 11000 Hz</td>
</tr>
<tr>
<td><strong>6.5-mm shift</strong></td>
</tr>
<tr>
<td>250 – 500 Hz</td>
</tr>
<tr>
<td>500 – 730 Hz</td>
</tr>
<tr>
<td>730 – 1015 Hz</td>
</tr>
<tr>
<td>1015 – 1450 Hz</td>
</tr>
<tr>
<td>1450 – 2000 Hz</td>
</tr>
<tr>
<td>2000 – 2600 Hz</td>
</tr>
<tr>
<td>2600 – 3800 Hz</td>
</tr>
<tr>
<td>3800 – 6800 Hz</td>
</tr>
</tbody>
</table>

Pre- and post-training phase. During pre- and post-training phases, listeners were required to transcribe 20 Harvard sentences by typing their responses onto the computer's keyboard. Responses could be entered immediately after each utterance ended and when a prompt and cursor appeared on the monitor. The same set of 20 Harvard sentences was used for both pre- and post-training, and the sentences were processed using no frequency shift. Participants did not receive any feedback during these two phases, and they did not have the option to repeat the sentence after it was played.

Training phase. During the training period, all listeners received training on a new set of 100 Harvard sentences, but the amount of frequency shift that each of the three listener groups received (0 mm, 3 mm, or 6.5 mm) varied. Listeners reported what they heard in each sentence following the same procedure used in the pre- and post-training sessions. All listeners received feedback in which the same processed sentence was played again after the written text of the sentence appeared on the screen. The text of the sentence appeared 250ms prior to the onset of each utterance and remained visible until each
utterance stopped playing. Listeners did not have the option to replay the feedback. The decision to use this type of feedback was based on the findings of an earlier study which indicated that this form of feedback results in the largest learning effects when compared to auditory-alone feedback and several control groups receiving either no feedback or no exposure to the degraded speech (Burkholder, 2005). Because comparisons between this feedback group and control groups have previously been made, no control group was included in the present study (Burkholder, 2005).

**Generalization phase.** During the generalization phase, three blocks of 50 novel Harvard sentences were presented to the listeners. Each block contained sentences processed by either the 0-mm shift, 3-mm shift, or 6.5-mm-shift model. All three sets of sentences were processed using each shift, and the combinations of sentence set and shift condition were counterbalanced among listeners to control for possible sentence effects. In the first phase of generalization, listeners always heard the block of sentences processed in the shift condition on which they were trained. The order of the final two blocks of sentences was counterbalanced among listeners. Half of the listeners in each group heard the smaller novel shift first followed by the larger shift. Half of the listeners heard the larger shift followed by the smaller shift.

**Data Analysis**

The data output files containing the correct responses and each participant’s responses for each experimental phase were imported into spreadsheets for scoring. Each sentence was scored by determining the percentage of key words that were identified correctly. Key words containing obvious spelling and typographical errors were scored as correct. Errors in spelling were scored as correct only if the response did not resemble a different word or combination of words. Similarly, typing errors were only scored as correct if the response did not resemble a different word.

A typing error was scored as correct if a target letter was substituted by any letter immediately surrounding it or if letters immediately surrounding the target letter were inserted into the response. Responses in which the correct letters were transposed were also considered typographical errors and scored correctly. Insertions of unnecessary spaces within words were also scored as correct if the response did not result in a different word. Homophones (i.e. “rain” for “reign”) of key words were also accepted as correct. However, key words reported in an incorrect tense (i.e. “place” for “placed”) or with other incorrect affixes (i.e. “shoes” for “shoe”) were scored as incorrect.

Scoring was completed by the experimenter and a trained undergraduate research assistant. Interrater reliability between the two scorers was previously determined to be high by computing the Cronbach’s alpha value of scores assigned to 10 listeners completing an identical task in a similar study (Burkholder, 2005). Any ambiguous scoring decisions were resolved at the experimenter’s discretion.

**Results**

The descriptive statistics for each groups’ performance on each phase of the experiment appear in Table 3.2. The data from two control groups used in a previous perceptual learning study were also included (Burkholder, 2005). To determine if all three groups of listeners had roughly the same abilities to identify speech processed through an acoustic simulation of a cochlear implant at the beginning of training, a one-way ANOVA was conducted on the pre-training scores. The analysis revealed no difference \( F(2, 63) = 1.72, p = .188 \) in the pre-training scores, indicating that all three groups were similar to each other prior to training.
Training and Perceptual Learning Data

Because the amount and time course of learning that occurred during the training period played a large role in pre- to post-training increases in performance, performance during training was determined and evaluated through several methods. Table 3.2 provides a summary of the listeners’ performance during the training task. Figure 3.4 illustrates the learning curves of each listener group which include performance before, during, and after training.

![Graph showing performance over phases of training with error bars representing the standard error of the mean.]

**Figure 3.4.** Performance of listener groups trained with no shift, a 3-mm shift, or a 6.5-mm shift before training, during five phases of training, and after training. Error bars represent the standard error of the mean.

The values for performance during each phase of training were obtained by splitting the 100 training sentences into fifths each containing 20 sentences. The overall average percent correct for each group for each phase of training was then determined. Thus, Figure 3.4 presents the groups’ pre- and post-test performance and performance on the first, third, and fifth or first, middle, and last block of 20 sentences heard by the listeners during training. Figure 3.5 illustrates the individual learning curves of the 20 listeners in each of the groups.

A one-way ANOVA was conducted to determine the effect of the training stimuli’s shift on keyword recognition scores during training. The main effect of frequency shift used during training was significant ($F(2, 63) = 52.41, p = .001$). A post-hoc Tukey test indicated that listeners trained on the 6.5-mm shift ($M = .30, SD = .06$) performed significantly worse than listeners who were trained with no shift ($M = .52, SD = .10$) or a 3-mm shift ($M = .51, SD = .08$).

The training data were also analyzed to determine the amount of improvement during the training period. Figure 3.6 displays performance during the first, middle, and last phase of the sentence training period for the three groups. A repeated-measures ANOVA was conducted on the three groups’ training performance using the three phases of training as the within-subjects factor and the amount of frequency shift used during training as the between-subjects factor. Results indicate a main effect of training phase ($F(2, 58) = 88.24, p < .001$).

The interaction between training phase and frequency shift used during training was also significant ($F(4, 56) = 3.87, p = .005$). This interaction effect reflects the finding that listeners trained with a 6.5-mm shift improved significantly less over the course of training than listeners trained with no shift or a 3-mm shift. The main effect of frequency shift used during training was significant ($F(2, 58) =$
Table 3.2. Means and standard deviations (SD) of the percent of key words correctly identified by listeners trained with no shift, a 3-mm shift, or a 6.5-mm shift before, during, and after training (left three columns) and during the generalization phase (right three columns). Previously acquired data from two control groups is also included.

<table>
<thead>
<tr>
<th>Shift Condition</th>
<th>Pre-Training (no shift)</th>
<th>Training</th>
<th>Pst-Training (no shift)</th>
<th>No shift</th>
<th>3mm shift</th>
<th>6.5mm shift</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unexposed</td>
<td>.37(.11)</td>
<td>.97(.06)</td>
<td>.47(.14)</td>
<td>.42(.12)</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>Untrained</td>
<td>.36(.16)</td>
<td>.50(.15)</td>
<td>.52(.18)</td>
<td>.49(.12)</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>No shift</td>
<td>.40(.11)</td>
<td>.52(.10)</td>
<td>.59(.11)</td>
<td>.50(.09)</td>
<td>.54(.11)</td>
<td>.30(.10)</td>
</tr>
<tr>
<td>3mm shift</td>
<td>.37(.09)</td>
<td>.51(.08)</td>
<td>.58(.10)</td>
<td>.46(.12)</td>
<td>.67(.07)</td>
<td>.32(.16)</td>
</tr>
<tr>
<td>6.5 mm shift</td>
<td>.33(.14)</td>
<td>.30(.06)</td>
<td>.51(.19)</td>
<td>.41(.12)</td>
<td>.59(.18)</td>
<td>.32(.14)</td>
</tr>
</tbody>
</table>

Pre- and Post-training Data

Figure 3.3 shows the pre- and post-training scores of the three shift groups. To assess the magnitude of improvement that occurred from pre- to post-training, a repeated-measures ANOVA was conducted using pre- and post-test assessment scores as the two levels of the within-subjects factor. The between-subjects factor was the amount of frequency shift in the training stimuli. Results indicate a significant main effect of assessment session (pre-test or post-test) \((F(1, 61) = 461.26, p < .001)\). The interaction was not significant, indicating that the amount of improvement from pre- to post-test did not differ between the three groups. The main effect of the amount of frequency shift in the training materials was also not significant.

![Graph](image)

**Figure 3.3.** Pre- and post-training performance by listener groups trained with no shift, a 3-mm shift, or a 6.5-mm shift. The change in percentage points from pre-training to post-training is shown above the bars representing each group. Error bars represent the standard error of the mean.
27.37, p < .001). Post-hoc analysis indicated that listeners trained with a 6.5-mm shift performed worse at all phases of training than listeners trained with either a 3-mm frequency shift (p < .001) or no shift (p < .001).

(a) no frequency shift

(b) 3-mm frequency shift

(c) 6.5-mm frequency shift

Figure 3.5. Learning curves for listeners trained with (a) no shift, (b) a 3-mm, and (c) a 6.5-mm shift before, during, and after training. Error bars represent the standard error of the mean.
Figure 3.6. Performance of listener groups trained with no shift, a 3-mm shift, or a 6.5-mm shift on the first, middle, and last phase of training. The change in percentage points from the first to last phase of training is shown above the bars representing each group. Error bars represent the standard error of the mean.

The interaction between training phase and frequency shift used during training was also significant ($F(4, 56) = 3.87, p = .005$). This interaction effect reflects the finding that listeners trained with a 6.5-mm shift improved significantly less over the course of training than listeners trained with no shift or a 3-mm shift. The main effect of frequency shift used during training was significant ($F(2, 58) = 27.37, p < .001$). Post-hoc analysis indicated that listeners trained with a 6.5-mm shift performed worse at all phases of training than listeners trained with either a 3-mm frequency shift ($p < .001$) or no shift ($p < .001$).

Generalization Data

The degree to which perceptual learning during training generalized to new stimuli and new frequency shifts was also tested. To examine generalization of perceptual learning, several additional conditions were used after training. In these generalization tests, listeners were asked to identify new sentences without receiving feedback. Listeners heard novel sentences with the same amount of frequency shift used during training as well as novel stimuli processed with the two unfamiliar frequency shift conditions. The three right-hand columns in Table 3.2 provide a summary of the descriptive statistics of performance during the generalization phases.

The generalization data were first analyzed to determine if there were order effects on the three groups’ abilities to identify speech processed using novel frequency shifts. During the generalization phase, all listeners in each group were first tested on the same frequency shift that they heard during the training period. The presentation of the two novel shifts was counterbalanced in each group. Thus, in the group trained using no frequency shift, half of the listeners were exposed to the intermediate shift followed by a large shift, and half heard the large frequency shift before the intermediate shift. In the group of listeners trained with a 6.5-mm frequency shift, half were exposed gradually to no shift, and half heard no frequency shift immediately after hearing the familiar shift condition.
Three separate repeated-measures ANOVAs were conducted to determine if there were any order effects for the three listener groups. Shift condition was the within-subjects factor, and order of presentation was the between-subjects factor. A main effect of shift was found in all three analyses. However, the interaction effect and main effect of order of presentation were not significant in any of the groups. Thus, listeners' performance on particular shift conditions was not significantly related to the order in which each shift was presented during generalization.

Figure 3.7 shows the three group's performance on the three separate shift conditions administered during generalization. A repeated-measures ANOVA was run using frequency shift of the test stimuli as the within-subjects factor and frequency shift of the training stimuli as the between-subjects factor. The frequency shift of the test stimuli had a significant effect on keyword recognition ($F(2, 58) = 98.22, p < .001$). The interaction effect between the frequency shift in test stimuli and frequency shift in training stimuli was also significant ($F(2, 58) = 3.66, p = .008$). The between-subjects main effect of shift used during training was not significant.

![Graph showing performance](image)

**Figure 3.7.** Generalization phase performance of listeners trained with no shift, a 3-mm shift, or a 6.5-mm shift. Error bars represent the standard error of the mean.

The interaction between the frequency shift in test stimuli and frequency shift in training stimuli was explored further by examining the separate within-subjects contrasts for the generalization data of each group. Listeners trained with no frequency shift performed worst on the sentences processed with a 6.5-mm shift during the generalization phase. Performance on the 6.5-mm shift condition was worse than performance on the no shift ($F(1, 19) = 12.25, p = .002$) and 3-mm shift ($F(1, 19) = 88.50, p < .001$) condition. Listeners trained with no shift did not perform significantly different during the no shift and 3-mm shift test conditions.

A similar pattern was observed for the listeners trained on a 3-mm frequency shift. They performed the poorest on the 6.5-mm frequency shift. Keyword recognition scores obtained during the 6.5-mm-shift condition were significantly lower than those obtained during the no shift condition ($F(2, 18) = 11.57, p = .003$). This group of listeners also performed the best on the 3-mm shift, which was the frequency shift on which they received training. Keyword recognition scores obtained during the 3-mm-
shift condition were better than both the no shift and 6.5-mm frequency shift condition \(F(2, 18) = 118.63, p < .001\).

Finally, the listeners trained on a 6.5-mm frequency shift also performed best on the 3-mm shift during the generalization phase. Keyword recognition scores obtained for the 3-mm-shift condition were significantly higher than those obtained during testing with the two other frequency shifts \(F(2, 18) = 42.51, p < .001\). The difference in performance between the no shift and 6.5-mm-shift conditions only approached significance in this group \(F(1, 19) = 4.01, p = .059\).

A final analysis used to assess the amount of generalization of learning compared listeners' performance on the three frequency shift conditions to their average performance during the entire training period. Figure 3.8 displays each group's performance during generalization along with performance during training. If performance during the generalization phases is equal to or exceeds performance during the training period, it can be concluded that robust generalization to new tokens and/or new frequency shifts was observed. Thus, the magnitude of the difference between training and generalization scores can be used as one index of the generalization of learning.

![Figure 3.8](image)

**Figure 3.8.** Generalization of learning during the training period to sentences processed in an old shift and in novel shifts by listeners trained on either no shift, a 3-mm shift, or a 6.5-mm shift. Error bars represent the standard error of the mean.

A repeated-measures ANOVA using experimental phase (training, no shift, 3-mm shift, and 6.5-mm shift) as the within-subjects factor and training group as the between-subjects factor was run. Main effects were found for both experiment phase \(F(3, 57) = 79.70, p < .001\) and the amount of shift used during training \(F(6, 57) = 9.07, p < .001\). Post-hoc analysis indicated that the group of listeners trained with the 6.5-mm frequency shift performed worse than both the listeners trained with no shift \(p = .016\) and a 3-mm frequency shift \(p = .001\). The interaction effect between experiment phase and amount of shift used during training was significant \(F(2, 58) = 8.16, p = .001\).

This interaction effect was explored further by conducting separate within-subjects contrasts for each group. Listeners trained with no shift generalized their learning partially to new tokens and new shifts. Training and generalization performance to new sentences processed with no shift did not differ.
Similarly, training performance and generalization to new sentences processed with a 3-mm frequency shift did not differ. However, generalization performance on the 6.5-mm condition was significantly worse than performance during training and worse than generalization to the no shift and 3-mm-shift condition \( (F(2, 18) = 34.30, p < .001) \).

Listeners trained on the 3-mm frequency shift also showed evidence of robust generalization of learning to new stimuli processed with the familiar shift and new stimuli processed with no shift. Performance of the 3-mm group on the 3-mm-shift condition during generalization exceeded performance during training and generalization on the no shift and 6.5-mm-shift conditions \( (F(3, 17) = 188.02, p < .001) \). Generalization to the no shift condition also occurred. There was no difference between keyword recognition scores obtained during training and those obtained during generalization to the no-shift condition.

Like the no shift group, the group trained with a 3-mm frequency shift did not perform better on the 6.5-mm shift during generalization than on the training stimuli. In contrast, performance on the 6.5 mm-shift condition was significantly worse than performance during training \( (F(3, 17) = 17.90, p < .001) \). Thus, listeners trained on a small frequency shift were not able to adapt to the larger frequency shift. In addition, the listeners' identification of keywords in the 6.5 mm-shift condition was worse than their identification of keywords in the no-shift condition \( (F(3, 17) = 11.57, p = .003) \).

For the listeners trained on the 6.5-mm frequency shift, perceptual learning did appear to generalize to sentences processed using both the familiar and novel frequency shifts. Keyword recognition scores obtained during the no-shift \( (F(2, 17) = 8.73, p = .008) \) and 3-mm-shift condition \( (F(2, 17) = 50.37, p = .008) \) both exceeded keyword recognition scores obtained during training. Performance during training and on the 6.5-mm-shift condition during generalization did not differ significantly \( (F(3, 17) = .417, p = .526) \). This indicates that some generalization to new sentences processed in the trained shift did occur although performance was still only 30% correct. During generalization, performance on the 6.5 mm-shift condition was significantly worse than performance on the no shift \( (F(3, 17) = 8.73, p = .008) \) and 3-mm-shift condition \( (F(3, 17) = 50.37, p = .008) \).

**Discussion**

Several consistent findings emerged from the current study. Listeners were able to learn from a short period of training with speech processed through an acoustic simulation of a cochlear implant and were able to improve speech perception scores rapidly, regardless of the amount of frequency shift on which they were trained. However, the amount of learning that occurred during the training period was related to the amount of frequency shift used. Listeners trained on a 6.5-mm frequency shift showed the smallest gains during the training period than listeners in the two groups trained with no shift or with a 3-mm shift. This result indicates that large frequency mismatches can significantly affect the perceptual learning that occurs over a short training period.

**Generalization of Learning to Novel Sentences and Novel Frequency Shifts**

In addition to being rapid, the learning observed in this study was found to be robust as well. Evidence of generalization of learning to novel stimuli processed with the frequency shift used during training was observed in all the listener groups. The pattern of generalization occurring among the groups indicates that all listener groups can, to some degree, generalize what they learned during training to new sentences processed using the same frequency shift on which they were trained. However, generalization of learning to new sentences processed with a familiar shift was more robust in the group of listeners.
trained with a small 3-mm frequency shift than in listeners trained with no shift or with a 6.5-mm shift. During generalization, keyword recognition scores for a 3-mm frequency shift were significantly greater than keyword recognition scores obtained during training with a 3-mm frequency shift. In addition, generalization to new stimuli processed with the familiar 3-mm shift was greater than generalization to no frequency shift and a 6.5-mm frequency shift.

Generalization of perceptual learning to novel frequency shifts was also observed in the listeners in this study. Listeners trained with no frequency shift generalized learning to the smaller 3-mm frequency shift. Listeners trained with a frequency shift were also able to generalize learning to the less spectrally mismatched stimulus conditions or to the no shift condition. However, the listeners trained with some frequency shift did not perform any better on the no shift condition in the generalization phase than unexposed and untrained listeners who transcribed sentences with no frequency shift in a previous study. This result indicates that the listeners trained with a shift in this study may have performed just as well on the no shift condition even if they had received no training. Such a result also suggests that there may not be robust generalization of auditory perceptual learning of spectrally degraded speech in the present study. Rather, the results of the current study may be related to processes of normalization that listeners may use when perceiving degraded speech processed using similar signal processing strategies.

Other limitations to generalization of learning were observed in this study. Neither the listeners trained with no frequency shift nor with a 3-mm frequency shift were able to generalize learning to novel stimuli with a large 6.5-mm frequency shift. However both of these listener groups performed just as well on the 6.5-mm frequency shift during generalization as the listeners trained on the 6.5-mm shift. These results indicate that a large frequency shift may be too difficult for listeners to adapt to during a short experiment. In addition, this result suggests that even when listeners do receive training with a large frequency shift they do not benefit much from the training. This result is consistent with previous findings that have indicated that extended periods of training with a large frequency shift may be necessary for listeners to learn from it and to adapt to it (Rosen et al., 1999; Svirsky et al., 2003).

Some Evidence of Specificity of Learning

In addition to finding some evidence that may indicate generalization of learning or normalization to small frequency shifts, we also observed some evidence of specificity of perceptual learning. Listeners trained on the 3-mm frequency shift showed specificity to the 3-mm shift during generalization. Although a 3-mm frequency shift is relatively small, it is still produced a more degraded speech signal than no frequency shift. It is noteworthy that the listeners in the 3-mm group performed better on slightly more difficult speech stimuli during the generalization phase. This result suggests that specificity of learning to the training stimuli rather than the degree of signal degradation in the test stimuli influenced this listener group’s performance during the generalization phase. This result also supports one of the original hypotheses that encoding specificity would be observed in listeners trained with a small frequency shift.

However, specificity of learning was not apparent in the listeners trained with no shift or a 6.5-mm shift. Listeners trained with the large 6.5-mm frequency shift did not perform the best on the 6.5-mm shift during generalization. The lack of specificity of learning to the 6.5-mm condition is not surprising given how poorly listeners trained on these stimuli performed during the training period and how little they actually improved during training compared to the other two groups of listeners. However, lack of encoding specificity in the listeners trained with no shift was an unexpected result.
An additional surprising result was the finding that the listeners trained on the 6.5-mm frequency shift did not perform best on the easiest no-shift stimulus condition during generalization. Listeners trained with a 6.5-mm shift actually performed the best on the 3-mm shift during the generalization phase. Although the overall performance of the listeners trained on the 6.5-mm frequency shift does not provide support for specificity of learning to the exact signal they were trained on, this pattern of results suggests that learning may have been specific to the frequency mismatch in general. Thus, listeners may not simply be learning from the overall spectral shape of the signal but may also be learning something specific about spectral mismatch in these degraded signals.

Perhaps due in part to their lack of specificity of learning for their trained shift, the listeners trained with the 6.5-mm frequency shift generalized learning to the novel shift conditions. The amount of generalization demonstrated by the 6.5-mm group was also larger than the generalization demonstrated by the other two groups. Thus, the hypothesis that generalization to less degraded stimuli would be easiest for listeners trained on the most difficult stimuli was supported.

A lack of specificity of learning in the group trained with no frequency shift is of interest. Unlike the other two groups, the listeners trained with no shift showed no evidence of specificity of learning. These listeners performed slightly better on the 3-mm-shift condition than on the no-shift condition during generalization. However, keyword recognition scores for this group were lowest during the 6.5-mm frequency shift phase of generalization. Taken together, these results indicate that, although listeners trained with no frequency shift were able to adapt to a small degree of frequency shift, a large frequency shift proved to be too difficult to adapt to during a short period of time.

In addition, listeners initially trained on the 3-mm shift also failed to adapt to the larger 6.5-mm shift. This result does not support the proposed hypothesis that listeners would be able to adapt to stimuli that had small perceptual differences from the training stimuli. However, the inability of these listeners to adapt to the large frequency shift may have been related to the short training and generalization periods used in the current study. Future studies should use extended training periods to determine if the inability to generalize learning to large frequency shifts is due to insufficient training.

Conclusions

The major finding of this study is that listeners have difficulty perceiving and learning to understand speech processed with a large frequency shifts. However, the patterns of generalization and specificity of learning in these three listener groups are complicated. Specificity of auditory perceptual learning to the trained shift appeared to be confined to the listeners trained on the small 3-mm frequency shift. It is unclear why this same pattern of learning was not observed in the listeners trained with no frequency shift. The lack of specificity of learning in this group and the overall tendency of all listeners to perform best on the 3-mm-shift condition during generalization suggests that there may be another related explanation for the present results.

The tendency for listeners trained with a 3-mm or 6.5-mm frequency shift to perform best on the 3-mm frequency shift during the generalization phases may be evidence of specificity of learning to the trained frequency shift or to a frequency shift in general. However, given that listeners trained with no frequency shift also performed well on the 3-mm-shift condition, there may be an alternative explanation for why the other two listener groups performed the best on the intermediate shift. Superior performance on the 3-mm frequency shift during generalization by all listener groups may instead be evidence that listeners simply perform better on an intermediate level of frequency shift that is identical to or most similar to what they heard during training. If this alternative explanation is correct, it indicates that
listeners may learn more about the general qualities of the degraded speech than about the specific acoustic characteristics associated with a particular shift condition during training.

**Experiment II**

Due to the ambiguity in several of the findings in Experiment I, a replication study was conducted to determine more conclusively if auditory perceptual learning of spectrally degraded speech is specific to the signal degradation heard during training or whether learning tends to generalize best to novel stimuli processed with a level of signal degradation intermediate to the other two stimulus conditions listeners were exposed to. To reexamine this question, two additional groups of listeners were trained on the no-shift and 6.5-mm-shift conditions. Our goal was to replicate the main findings observed in the previous study.

A replication of the results obtained in Experiment I should reveal whether the learning is specific to the frequency shift listeners were trained or whether learning is not necessarily most generalizable to the least degraded and presumably easiest to stimuli to perceive. A replication of the previous results would indicate that listeners tend to be most capable of generalizing their learning to a frequency mismatch that is intermediate to the other frequency shifts they were exposed to and that they learn primarily from the basic spectro-temporal alterations in the signal. In contrast, failure to replicate the lack of specificity of learning in other listeners trained with no shift or a large frequency shift would suggest that listeners do learn from the fine structure of the acoustic signals on which they were trained.

**Method**

The methods, materials, and participant recruitment and payment were identical to those used in Experiment 1. Only two listener groups were used instead of three. Twenty listeners were trained using no frequency shift, and 20 were trained using the 6.5-mm shift.

**Results**

The descriptive statistics for each groups’ performance on each phase of the experiment appear in Table 3.3. To determine if both groups of listeners had roughly the same abilities to identify speech processed through an acoustic simulation of a cochlear implant prior to training, a paired-samples T-test was conducted on the pre-training scores. This test revealed no difference ($t(19) = .04, p = .970$) in the pre-training scores, indicating that the two groups were similar to one another prior to training.

**Pre- and Post-training Data**

Figure 3.9 shows the pre- and post-training scores of the groups. To assess the magnitude of improvement that occurred from pre- to post-training, a 2 x 2 repeated-measures ANOVA was conducted using pre- and post-test assessment scores as the two levels of the within-subjects factor. The between-subjects factor was the amount of frequency shift in the training stimuli. Results indicate a significant main effect of assessment session (pre-test or post-test) ($F(1, 38) = 278.06, p < .001$). Both groups improved from pre- to post-training. However, a significant interaction ($F(1, 38) = 7.59, p = .009$) indicated that the listeners trained on the 6.5-mm frequency shift did not improve as much after training as listeners trained using no frequency shift.
Figure 3.10. Performance of listener groups trained with no shift or a 6.5-mm shift before training, during five phases of training, and after training. Error bars represent the standard error of the mean.

(a) no frequency shift

(b) 6.5-mm frequency shift

Figure 3.11. Individual learning curves for listeners trained on (a) no frequency shift and (b) a 6.5-mm frequency shift before, during, and after training. Error bars represent the standard error of the mean.
Table 3.3. Means and standard deviations (SD) of the percent of key words correctly identified by listeners trained with no shift or a 6.5-mm frequency shift before, during, and after training (left three columns) and during the generalization phase (right three columns).

<table>
<thead>
<tr>
<th>Shift Condition</th>
<th>Pre-Training</th>
<th>Training</th>
<th>Pst-Training</th>
<th>No shift</th>
<th>3mm shift</th>
<th>6.5mm shift</th>
</tr>
</thead>
<tbody>
<tr>
<td>No shift</td>
<td>.42(.13)</td>
<td>.53(.10)</td>
<td>.63(.12)</td>
<td>.59(.10)</td>
<td>.56(.10)</td>
<td>.27(.09)</td>
</tr>
<tr>
<td>6.5 mm shift</td>
<td>.42(.11)</td>
<td>.31(.16)</td>
<td>.57(.10)</td>
<td>.47(.08)</td>
<td>.54(.09)</td>
<td>.33(.17)</td>
</tr>
</tbody>
</table>

Figure 3.9. Pre- and post-training performance by listener groups trained with no shift and 6.5-mm shift. The change in percentage points from pre-training to post-training is shown above the bars representing each group. Error bars represent the standard error of the mean.

Training and Learning Data

Because the amount and time course of learning occurring during the training period played a large role in pre- to post-training increases in performance, performance during training was determined and evaluated through several methods. Table 3.3 includes a summary of the descriptive statistics of the listeners’ performance during the training task. Figure 3.10 illustrates the learning curves of each group which include performance before, during, and after training. Figure 3.11 displays the individual learning curves of the 20 listeners in the no-frequency-shift and 6.5-mm-frequency-shift group.
A one-way ANOVA was conducted to determine the effect of the training stimuli's shift on keyword recognition scores during training. The main effect of frequency shift used during training was significant ($F(1, 39) = 18.42, p < .001$). Listeners trained on the 6.5-mm shift ($M = .31, SD = .16$) performed significantly worse than listeners who were trained with no shift ($M = .53, SD = .10$).

The training data were also analyzed to determine the amount of improvement during the training period. Figure 3.12 displays performance during the first, middle, and last phase of the entire sentence training period for the no-shift and 6.5-mm frequency shift groups. A repeated-measures ANOVA was conducted on the two groups' training performance using the three phases of training as the within-subjects factor and the amount of frequency shift used during training as the between-subjects factor. Results indicate a main effect of training phase ($F(2, 38) = 41.74, p < .001$). An interaction between training phase and frequency shift used during training was also significant ($F(2, 38) = 6.77, p = .002$). This interaction reveals that listeners trained with a 6.5-mm shift improved significantly less over the course of training than listeners trained with no shift. The main effect of frequency shift used during training was also significant ($F(2, 38) = 59.83, p < .001$). Listeners trained on a 6.5-mm frequency shift performed worse during the entire training period.

**Figure 3.12.** Performance of listener groups trained with no frequency shift and a 6.5-mm frequency shift on the first, middle, and last phase of training. The top panel shows the replication data. The bottom panel shows the previous data. The change in percentage points from the first to last phase of training is shown above the bars representing each group. Error bars represent the standard error of the mean.
Generalization Data

The three right-hand columns in Table 3.3 show the descriptive statistics of performance during the generalization phases. Figure 3.13 shows the groups’ performance on the three separate shift conditions administered during generalization. The generalization data were then analyzed to determine if there were order effects on the three groups’ abilities to identify speech processed using novel frequency shifts. Two repeated-measures ANOVAs were conducted on the groups’ generalization performance. Shift condition was the within-subjects factor, and order of presentation was the between-subjects factor. The results of these analyses were identical to Experiment 1. Significant effects of shift were found. However, the interaction between shift condition and order of presentation, and the main effect of order of presentation were not significant. Thus, listeners’ performance on particular shift conditions was not significantly related to the order in which each shift was presented.

![Replication](image1)

![Original Results](image2)

**Figure 3.13.** Generalization phase performance of listeners trained with no frequency shift or a 6.5-mm frequency shift. The top panel shows the replication data. The bottom panel shows the previous data. Error bars represent the standard error of the mean.
A repeated-measures ANOVA was run using frequency shift of the generalization stimuli as the within-subjects factor and frequency shift of the training stimuli as the between-subjects factor. The main effect of shift used during training was not significant. The frequency shift of the generalization stimuli had a significant effect on keyword recognition ($F(2, 38) = 56.18, p < .001$). Tests of within-subjects variables indicated that keyword recognition scores for stimuli processed with the 6.5-mm frequency shift were worse than keyword recognition scores for sentences processed with a 3-mm frequency shift ($F(1, 39) = 35.05, p < .001$) shift and no frequency shift ($F(1, 39) = 79.65, p < .001$).

The interaction effect between the frequency shift in test stimuli and frequency shift in training stimuli was significant ($F(2, 38) = 6.44, p = .003$). This interaction effect was explored further by examining within-subjects contrasts of each generalization phase in both groups. Listeners trained on a 6.5-mm frequency shift performed better on sentences processed with a 3-mm frequency shift than with no shift ($F(1, 18) = 5.40, p = .031$).

The generalization of learning extended to the 3-mm and 6.5-mm frequency shift conditions did not differ in listeners trained using no shift ($F(1, 18) = .801, p = .4333$). A paired-samples t-test on the keyword recognition scores for stimuli processed using no shift also indicated that listeners trained with no shift ($M = .59, SD = .10$) performed significantly better ($t(14) = 4.33, p < .001$) on novel sentences processed with no shift than listeners trained with a 6.5-mm shift ($M = .47, SD = .08$).

A final analysis was carried out to assess the amount of generalization of learning and compare listeners’ performance on the three frequency shift conditions to their performance during training. Figure 3.14 displays each group’s performance during generalization along with performance during training. If performance during the generalization phases is equal to or exceeds performance during the training period, it can be concluded that generalization to new tokens or new frequency shifts has occurred. Therefore, the magnitude of the difference between training and generalization scores is one measure of the robustness of learning.

A repeated-measures ANOVA using experimental phase (training, no shift, 3-mm shift, and 6.5-mm shift) as the within-subjects factor and training group as the between-subjects factor was run. Main effects were found for both experiment phase ($F(3, 38) = 36.52, p < .000$) and the amount of shift used during training ($F(1, 38) = 6.64, p = .014$). The interaction effect between experiment phase and amount of shift used during training was significant ($F(3, 38) = 10.57, p = .001$).

This interaction effect was explored further by examining within-subjects contrasts in each group. Listeners trained using no shift performed significantly better during training than during the generalization phase that used 6.5-mm frequency shift ($F(1, 19) = 30.56, p < .001$). Thus, they did not generalize learning to this large frequency shift. These listeners’ keyword recognition scores during the generalization period using no shift just significantly exceeded keyword recognition scores during training ($F(1, 19) = 4.85, p = .040$). No differences were found between the no-shift group’s training performance and performance during the phase of generalization using a 3-mm frequency shift ($F(1, 19) = 3.66, p = .071$). Thus, in these listeners, learning was generalized to stimuli processed using the same parameters as heard in training. However, listeners trained on no shift also partially generalized learning to novel sentences processed with a small 3-mm frequency shift.

Listeners trained with a 6.5-mm frequency shift did not perform differently during training and during the generalization phase which used a 6.5-mm frequency shift ($F(1, 19) = 2.12, p = .125$). However, their performances on sentences processed using no shift ($F(1, 19) = 11.28, p = .003$) and a 3-mm frequency shift ($F(1, 19) = 18.70, p < .001$) both exceeded performance during training. Thus,
listeners trained on stimuli processed with a 6.5-mm frequency shift do generalize their learning to new stimuli processed with a novel 3-mm frequency shift or no frequency shift at all.

(a) Replication

![Graph showing replication data]

(b) Original Results

![Graph showing original data]

**Figure 3.14.** Generalization of learning during the training period to sentences processed in an old shift and in novel shifts by listeners trained on either no shift or a 6.5-mm shift. The top panel shows the replication data. The bottom panel shows the previous data. Error bars represent the standard error of the mean.

**Discussion**

The results of this replication were, for the most part, similar to those found in Experiment 1. Several of the major findings observed in the previous study were replicated. Listeners trained on a 6.5-mm frequency shift showed less improvement during the training period than listeners trained with no shift. In addition, the listeners in this group failed to show robust generalization of their learning to new stimuli processed with the large frequency shift. Thus, listeners trained with a large 6.5-mm frequency shift
shift did not show specificity of learning to the stimuli on which they were trained. However, as in the first study, listeners trained on the 6.5-mm shift performed best on the small 3-mm frequency shift. This result supports the proposal that the listeners may have learned to alter their representations of speech based on the presence of spectral mismatch rather than on the specific fine structure of the training stimuli or on the simple spectro-temporal information.

Listeners trained with no frequency shift also failed to generalize what they learned during training to the larger 6.5-mm frequency shift. However, generalization to stimuli processed with a small frequency shift and no frequency shift was observed in both of these listener groups. Similar to the first study, listeners trained on the large 6.5-mm frequency shift appeared to show more evidence of generalization of learning. During the generalization phase, their performances on sentences processed using no shift and a 3-mm frequency shift both exceeded performance during training. This result provides additional support for the idea that it is easiest for listeners to generalize learning to less degraded stimuli.

Several other findings also emerged from this replication. First, the group of listeners trained with no frequency shift generalized their learning more robustly to novel sentences processed with no frequency shift than to sentences processed with any frequency shift. Listeners trained with no shift in Experiment 1 failed to show this specificity.

In the earlier study, we observed specificity of learning only in the group of listeners trained on the intermediate 3-mm frequency shift. This result was problematic because the implications of this finding were not clear. Listeners may have performed best on this condition either because learning was specific to the signals they were trained on or because the intermediate level of degradation was overall easier for listeners to generalize learning to. However, the results obtained with the present group of listeners provide some support to the idea that there is some specificity of learning for the training stimuli. Thus, listeners may learn something specific about the fine-structure of the degraded auditory signal that they hear rather than simply learn from the gross spectro-temporal information that is present in all the variations of the cochlear implant model that they heard.

A second piece of evidence supporting encoding specificity to trained frequency shifts was also obtained in this study. A significant interaction between the increase in performance from pre-training to post-training and training condition was found. Listeners trained with the large frequency shift did not improve as much on the post-training sentence set than listeners trained with no frequency shift. Listeners trained with no frequency shift likely performed better at post-training, because the stimuli used for this phase were also processed with the same parameters on which they were trained. However, lower post-training performances by the group trained with a 6.5-mm shift may also be partly due to the small amount of learning that they experienced during training.

General Discussion

Rapid and Robust Perceptual Learning with Acoustic Simulations of Cochlear Implants

Taken together, the two studies presented here have uncovered several theoretically and clinically important findings. One primary finding is that normal-hearing listeners can rapidly learn to perceive spectrally degraded speech that is both spectrally matched and slightly mismatched. In addition, listeners are able to generalize this perceptual learning to stimuli processed through novel signal processing parameters that generate different degrees of frequency shift. This result indicates that even short-term training can result in robust learning that generalizes to novel stimuli. In addition, evidence of
generalization to other frequency shifts suggests that listeners engage in broad learning of the spectrotemporal information rather than encoding of the fine structure specific to the training stimuli.

However, listeners’ abilities to generalize learning to novel frequency shifts were found to be dependent on the size of the novel frequency shift. The groups of listeners trained with no frequency shift or with a large 6.5-mm frequency shift were able to generalize learning to a small 3-mm frequency shift. In contrast to the training on the small frequency shift, training on the larger 6.5-mm frequency shift appeared to be too degraded for listeners to generalize their learning to after a short period of training.

Previous research has suggested that a 3-mm basalward frequency shift may result in an audible change that is comparable to changing the gender or age of a speaker (Fu & Shannon, 1999). Listeners are known to be able to normalize to these types of changes in the speech signal, thus, it is not surprising that listeners generalized what they had learned to this stimulus condition (Pisoni, 1996). However, the failure of listeners to generalize learning to the larger 6.5 mm frequency shift may have more interesting implications for the robustness of learning or for perceptual normalization of degraded auditory stimuli.

Theoretically, a lack of generalization to a large frequency shift suggests that there are fundamental sensory limitations both to the robustness of short-term auditory perceptual learning and to an auditory system that is confronted with highly degraded and unfamiliar speech stimuli. Clinically, this is an important finding because it suggests that cochlear implant users may not learn enough from short-term exposure to and training with tonotopically matched frequency to electrode placement to generalize learning to mapping strategies with a large amount of spectral mismatch. Thus, incremental training with multiple levels of frequency shift may indeed be one of the most efficient ways for cochlear implant users to adapt to large frequency shifts (Svirsky, 2003).

However, one drawback to the current study is that it was conducted using a short term of training and testing. Perhaps if more long-term training had been provided, more robust learning that was generalizable to a larger frequency shift would have been observed. Future directions of this research should include studies using more extended training and tests of generalization in normal-hearing listeners exposed to acoustic simulations of cochlear implants and in newly implanted cochlear implant patients.

It is important to consider how cochlear implant users generalize learning among frequency shifts because post-lingually deafened adults must cope with electrical stimulation that cannot accurately reproduce the normal tonotopy of the cochlea. Although clinicians working in the cochlear implant field are aware of these frequency shifts, no rehabilitative strategies have been formally designed or implemented to help patients adjust to them (McConkey-Robbins, 2000). Therefore, studying normal-hearing individuals’ abilities to generalize learning to frequency shifts in acoustic models of cochlear implants and studying cochlear implant patients’ abilities to generalize learning to new mapping strategies may provide valuable new information that is directly applicable to the development of treatments and intervention for cochlear patients.

Some Specificity of Learning with Acoustic Simulations of Cochlear Implants

Although the rapid pace and limitations to the robustness of auditory perceptual learning were clearly and consistently observed in the current studies, the degree of specificity that listeners developed for trained stimuli was less clear. In Experiment I, only listeners trained on a 3-mm frequency shift performed the best on their trained frequency shift during generalization. In addition, both of the other listener groups also performed the best on the 3-mm shift during generalization. Such a result could have
two plausible explanations. Some specificity of learning for the trained stimuli or for a frequency shift in general may have developed for the group of listeners trained on the 3-mm and 6.5-mm frequency shifts. Alternatively, the intermediate 3-mm frequency shift may have simply been the easiest for listeners to generalize learning to because of its close similarity to the training stimuli of all listeners.

In the replication study, evidence of specificity of learning was observed in the group of listeners trained with no frequency shift. Listeners trained with no frequency shift showed larger improvements on the trained stimuli than on the novel frequency shifts during the generalization phase. However, listeners trained on a 6.5-mm frequency shift still performed the best on the 3-mm frequency shift. The replication of this finding suggests that, although the listeners trained with a 6.5-mm frequency shift have no specificity for the stimuli on which they were trained, they may have learned something specific about the frequency shift in general.

Replicating a lack of specificity in the 6.5-mm training group also suggests that the difficulty of the training stimuli may impact the specificity of learning. The training data from the group of listeners trained on the 6.5-mm frequency shift supports this idea. In both experiments, listeners trained on the largest shift failed to learn as much throughout the training period as the other two groups of listeners did. It is possible that this learning was not sufficient to elicit the same specificity that was observed in listeners trained with a 3-mm frequency shift and in some listeners trained without a frequency shift.

The lack of specificity for a 6.5-mm frequency shift and superior performance on the 3-mm frequency shift in the listeners trained with a 6.5-mm shift suggests that these listeners may demonstrate an interesting pattern of learning. In listeners trained with the large 6.5-mm frequency shift, it appears that specificity of learning may be restricted to the general properties of a frequency shift or spectral mismatch rather than to the fine structure of the signal or specific signal processing used to generate the training stimuli. In addition, the performance of listeners trained on a large frequency shift suggests that the ability to encode fine structure in an auditory signal may depend on how degraded the training stimuli are when training begins.

However, even the minimal amount and ambiguous pattern of learning displayed by the listeners trained with the 6.5-mm frequency shift was robust enough to generalize to stimuli processed with a 3-mm frequency shift or no frequency shift. Thus, it appears that even short-term auditory training with speech processed with a large frequency shift can be beneficial to normal-hearing listeners who are later presented with similar but less degraded stimuli.

Taken together, the findings from both of these experiments suggest that, after short training and testing periods, auditory perceptual learning shows some specificity to the trained signal processing if it has no frequency shift or only a small frequency shift. However, despite the specificity of learning that was observed in listeners trained with no shift or a small shift, training with an acoustic simulation of a cochlear implant was also generalizable to novel frequency shifts. Thus, in the present study, the robustness of auditory perceptual learning was demonstrated in two ways.

Traditionally, robustness of learning is conceptualized as the ability to generalize learning to novel stimulus categories and signal processing. This ability was demonstrated by all groups of listeners in the current studies to some degree. However, learning that appears to specifically encode the trained signal processing after only brief training and that is retained over a short-term period could be considered to be "robust" as well. Several groups of listeners in the current studies also demonstrated some evidence of specificity which indicates short-term retention of encoding and learning.
Although the current results provide some evidence of generalization of learning to novel frequency shifts and suggest limited specificity of learning in several listener groups, it is possible that the performance of the listeners was simply due to normalization. Normalization is the process by which acoustically different stimuli are transformed and combined into a common representation that can be retained and accessed in memory (Pisoni, 1996). The present finding that listeners trained with no frequency shift generalize learning to a small frequency shift may actually be due to a normalization process that is no different than that used when listeners encounter speech spoken by talkers of different ages and genders (Fu & Shannon, 1999a). In addition, what appears to be generalization to a small frequency shift by the listeners trained on the large shift may instead be evidence that the listeners have simply converted the slightly different auditory signals into equivalent representations of spectrally mismatched speech.

Future directions of this research should aim to identify, separate, and analyze more conclusively the characteristics of the specificity and generalization of auditory perceptual learning. In addition, specific attempts should be made to distinguish true generalization and specificity from automatic perceptual normalization of degraded auditory signals. By conducting more detailed analyses of the specificity and generalization of perceptual learning and the perceptual normalization process, additional clues about what and how listeners are learning about the auditory signal may be uncovered. In addition, the time constraints related to robustness of auditory perceptual learning with acoustic simulations of cochlear implants should also be examined. Longitudinal examinations of auditory perceptual learning would be useful to determine if specificity of learning and the ability to generalize learning can be retained after prolonged periods without training or exposure to degraded auditory stimuli.

Answering such questions may have important theoretical implications for how listeners learn to adapt to degraded speech and for understanding the limitations of this form of adaptation over extended periods. Examining the long-term specificity and generalization of auditory perceptual learning and evaluating perceptual normalization processes in spectrally degraded speech may also have important clinical implications for cochlear implant patients. As cochlear implant technology and mapping and speech processing strategies evolve and become available to currently implanted patients, it becomes more important to understand how robust and specific their learning of previous auditory stimulation was and how this prior learning may impact their ability to learn to understand newer and better methods of electrical hearing. In addition, understanding if and how cochlear implant users normalize the speech stimuli that they hear could also provide valuable clues to understanding how their speech perception skills could be improved through training.
CHAPTER IV: TRANSFER OF LEARNING TO SPEECH AND NONSPEECH TASKS

Introduction

The ability to perceive speech in one’s native language is a robust skill for normal-hearing listeners even under highly degraded listening conditions such as sinewave, synthetic, noise-masked, and spectrally degraded speech (e.g. Burkholder, 2005; Davis, Johnsrude, Harvais-Adelman, Taylor, & McGettigan, 2005; Fenn, Nusbaum, & Margoliash, 2003; Pisoni, Manous, & Dedina, 1987; Remez, Rubin, Pisoni, & Carrell, 1981; Schwab, Nusbaum, & Pisoni, 1985; Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995; Sumby & Pollack, 1954). The abilities of normal-hearing listeners to perceive degraded speech appear to be robust, because they develop rapidly with brief exposure or training with degraded stimuli (e.g. Burkholder, 2005; Davis et al., 2005; Fenn et al., 2003; Remez et al., 1981). Rapid recognition and perceptual learning of degraded speech have been observed in studies using sinewave (Remez et al., 1981), synthetic (Fenn et al., 2003) and spectrally reduced speech (Burkholder, 2005; Davis et al., 2005) modeled after the input of cochlear implants.

The robustness of normal-hearing listeners’ abilities to learn to perceive highly degraded unnatural speech has also been demonstrated by measuring retention of learning. Retention of the ability to identify synthetic speech has been demonstrated over both short and extended periods of time. For example, Fenn and colleagues (2003) found that listeners trained to identify synthetic speech performed better on post-test sessions conducted 12 hours after training. In addition, they also found that retention of perceptual learning was most robust when the 12 hour separation between pre- and post-test was due to a sleep cycle rather than hours spent in an alert state (Fenn et al., 2003). This finding suggests that consolidation and retention of perceptual learning of synthetic speech is enhanced during sleep.

However, more recent data suggest that auditory perceptual learning can be retained over extended periods of wakefulness (Roth, Kishon-Rabin, Hildesheimer, & Karni, 2005). After being trained to identify consonant-vowel stimuli in noise, listeners were able to retain what they had learned after 12 hours in an awake-state. Based on these results, Roth and colleagues (2005) suggested that consolidation of perceptual learning may not even occur unless a sufficient amount of time is spent in an awake-state immediately after training. In the study by Roth and colleagues (2005), it appeared that significant gains in performance did not occur until at least fours had elapsed after training.

In addition to short-term and sleep-induced retention of perceptual learning, long-term retention of perceptual learning of synthetic speech has been explored in normal-hearing listeners. Schwab and colleagues (1985) found that listeners who were trained to identify synthetic speech showed evidence of retention of perceptual learning after an extended period of time. Retests with their trained listeners after a six month retention interval revealed that they had retained the knowledge enabling them to identify synthetic speech generated through a text-to-speech system. Taken together, these and other studies demonstrate the importance of assessing generalization and retention of learning to determine the robustness of auditory perceptual learning.

Listeners’ abilities to generalize learning to stimuli on which they received no training also provide another measure of the robustness of auditory perceptual learning (Karmarkar & Buonomano, 2003; Moore, Amitay, & Hawkey, 2003). Generalization of perceptual learning suggests that the benefits of training are often not restricted to the specific tokens on which listeners were trained. Rather, generalization indicates that perceptual learning was robust enough to extend to novel tokens that were similar to the tokens on which listeners were trained. For example, if listeners are able to identify novel
sentences after being trained only on other sentences, generalization to new training-stimulus tokens (i.e. sentences, words) is considered to have occurred (Burkholder, 2005; Fenn et al., 2003).

Generalization of learning can also be demonstrated on novel stimuli taken from a stimulus category that is different from the category on which listeners received training. If listeners are able to accurately identify isolated words after being trained only on sentences, generalization or "carry over" to a novel stimulus category is considered to have occurred (Greenspan, Nusbaum, & Pisoni, 1988; Hnath-Chisholm, 1997; Schwab et al., 1985). In addition to the word and sentence level, generalization of auditory perceptual learning can also occur at the phonemic level. Previous research has found that speech perception training on specific synthetic phonemes in one phonemic environment generalizes to the same phonemes in different phonemic environments (Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Logan, Lively, & Pisoni, 1991; Schwab et al., 1985). In addition, training listeners to discriminate between unfamiliar speech sounds with one place of articulation has been found to transfer to the discrimination of novel speech sounds with a different place of articulation (Tremblay, Kraus, Carrell, & McGee, 1997).

Generalization of auditory perceptual learning has also been examined by testing listeners' abilities to identify speech transformed using different signal processing techniques. In several studies using acoustic simulations of cochlear implants, generalization to different signal carriers and small frequency shifts or spectral mismatch has been found in normal-hearing listeners (Burkholder, 2005; Davis, Taylor, Johnsrude, & Carlyon, 2004; Fu & Shannon, 1999a). Davis and colleagues (2004) determined that training with spectrally degraded speech generated with noise-band carriers generalized to spectrally degraded speech created using pulse-train carriers. In addition, generalization of learning to small frequency shifts has been observed in normal-hearing listeners exposed to acoustic simulations of cochlear implants (Burkholder, 2005; Fu & Shannon, 1999a).

Generalization of auditory perceptual learning has also been found for environmental sound training and identification (Gygí, Kidd, & Watson, 2004). Gygí and colleagues (2004) found that normal-hearing listeners who were trained to identify 70 environmental sounds that were bandpass filtered were better at identifying the same sounds when they were processed through a single-channel noise vocoder. Taken together, these findings indicating that auditory perceptual learning for both speech and nonspeech sounds generalizes to different signal processing methods and to spectral mismatch suggest that listeners learn about the abstract spectro-temporal changes in the auditory signal rather than the complex acoustic-phonetic or fine spectral structure in the signals on which they were trained. In addition, studies of generalization and retention of learning have provided firm evidence that auditory perceptual learning is a robust and resilient process.

**Measuring the Robustness and Characteristics of Perceptual Learning through Transfer**

Taken together, the speed, generalizability, and retention of auditory perceptual learning in normal-hearing listeners all indicate that it is a very robust process. However, one additional and important benchmark of the robustness of auditory perceptual learning has gone relatively unexplored. The ability to transfer auditory perceptual learning to different tasks could arguably be one of the most accurate and important assessments of how robust and thorough auditory perceptual learning is. Transfer of learning suggests that listeners have flexibility enabling them to extend what they have learned in one task to entirely new tasks and/or test procedures on which they were not trained. Transfer of learning also suggests that training on one cognitive or perceptual task can facilitate performance on subsequent tasks that are related but inherently different.
Transfer of learning has been observed in the auditory (Delhommeau, Micheyl, Jouvent, & Collet, 2002; Nygaard & Pisoni, 1998; Nygaard, Sommers, & Pisoni, 1994), visual (Hunstad, 1985), and motor modalities (Murray, 1981; Teixeira, 2000) as well as in amodal cognitive tasks (Benson, Lovett, & Kroeber, 1997; Lovett, Borden, DeLuca, Lacerenza, Benson et al., 1994; Muramoto, 2001) in both humans and nonhuman species (Declay, 2001; Nakagawa, 1999; Nakagawa, 2000; Watanabe, 1986; Witte & Kipke, 2005). In the auditory domain, transfer of perceptual learning has been demonstrated in both simple psychophysical discrimination tasks (Delhommeau et al., 2002) and in more complex voice recognition and speech perception tasks (Nygaard & Pisoni, 1998; Nygaard et al., 1994).

In an early study of auditory perceptual learning, Nygaard and her colleagues (1994) trained listeners to identify novel voices by name. Using only listeners capable of identifying the voices correctly 70% of the time after nine days of training, Nygaard and her colleagues (1994) asked the listeners to identify novel noise-masked words spoken by either the familiar or unfamiliar voices. Listeners who heard novel words produced by familiar voices correctly identified significantly more words than the listeners who heard the words produced by unfamiliar voices (Nygaard et al., 1994). This result suggests that the information the trained listeners learned about the voices during voice identification training transferred to new words in a speech intelligibility task. Thus, even though the listeners were not instructed to identify words during the original voice training or to identify or remember anything specific about the quality or acoustic characteristics of the voices, they did learn something that enhanced their speech perception performance on a different task when the familiar voices were used again.

In a follow-up study, Nygaard and Pisoni (1998) used a control group of listeners who were simply exposed to the same voices as the trained listeners. These listeners did not identify noise-masked words spoken by the old speakers more accurately than listeners who heard words spoken by new speakers. This result replicated the earlier findings that listeners who heard words spoken by familiar speakers performed better than listeners who heard words spoken by unfamiliar speakers because of transfer of auditory perceptual learning rather than mere exposure to the voices.

Nygaard and Pisoni (1998) also found that training listeners to identify voices by name by using sentences transferred to the identification of individual words in sentences. However, training listeners to identify voices by name with sentences did not transfer to the identification of isolated words. This result indicates that there may be limits to the ability to transfer auditory perceptual learning from one task to another. Specifically, in circumstances in which both the transfer task and test stimuli change, listeners may be unable to fully transfer auditory perceptual learning. The difficulty in transferring learning under these circumstances may be due to the fact that generalization of learning to a new category of stimuli is also required.

Studying transfer of auditory perceptual learning in normal-hearing listeners is of theoretical interest for several reasons. First, the abilities of listeners to transfer learning to new tasks can provide new insights into how robust and generalizable perceptual learning really is. In addition, examining conditions in which transfer of auditory perceptual learning does or does not occur may help determine what properties of the auditory signal listeners become attuned to while being trained. For instance, Nygaard and her colleagues (1994; 1998) found evidence that listeners do learn about the specific characteristics and acoustic attributes of a speaker's voice, although they were not specifically trained to learn any of these things. This result also suggests that the encoding of vocal source characteristics and perceiving the linguistic content of speech are independent but closely intertwined (see Mullennix & Pisoni, 1990).
Clinical Implications for Transfer of Auditory Perceptual Learning

Studying the transfer of auditory perceptual learning can also have important clinical implications for certain populations of listeners. Hearing-impaired listeners are one population of individuals for which the ability to transfer auditory perceptual learning is important. Even with appropriate amplification from hearing aids or treatment with a cochlear implant, many hearing-impaired listeners still lack sufficient speech perception, spoken word recognition, and speech production skills (Kirk, Diefendorf, Riley, & Osberger, 1995; Miyamoto, Osberger, Robbins, Myres, Kessler et al., 1991; Osberger, Maso, & Sam, 1993; Pisoni, Cleary, Geers, & Tobey, 2000). Because of the difficulties many hearing-impaired listeners have in developing robust and highly adaptive speech perception and production skills, methods of rehabilitation and training appropriate for these listeners have been widely explored (Bode & Oyer, 1970; Deguchi, Kagami, & Hiki, 1981; Hutchinson, 1990; Kennedy & Weener, 1973; Massaro & Light, 2004).

Most early research on rehabilitative strategies in hearing-impaired listeners focused exclusively on hearing aid users (Bode & Oyer, 1970; Deguchi et al., 1981; Kennedy & Weener, 1973). One goal of this research was to identify auditory training paradigms that had the broadest amount of benefit for speech perception and production and language related skills. Thus, the most desired rehabilitative and auditory training programs were considered to be those that resulted in learning that transferred to multiple auditory conditions outside of the clinical setting.

Transfer of auditory perceptual learning has been observed using a variety of different treatment methods and stimulus materials with hearing-impaired adults and children. Boyd & Oyer (1970) found that training listeners to identify spoken words in both varied and constant speech-to-noise ratios resulted in improvements to speech discrimination. In addition, both open- and closed-set word recognition training resulted in improved speech discrimination skills (Boyd & Oyer, 1970).

Providing training on auditory discrimination skills has also been found to result in learning transferable to speech skills. Thus, it appears that for some auditory and speech skills, the transfer of learning between each other may be symmetric. Deguchi and colleagues (1981) found that training hearing-impaired adolescents to discriminate between falling and rising fundamental frequencies of complex tones resulted in an improvement in the hearing-impaired adolescents’ own vocal pitch as they spoke. Simple consonant recognition training has also been found to improve adults’ spoken word recognition (Rubinstein & Boothroyd, 1987).

Taken together, previous research indicates that transfer of learning from clinical training by hearing-impaired listeners is widespread and diverse. However, this prior interest in transfer of learning was directed exclusively at hearing-impaired individuals fitted with hearing aids. Currently, many deaf individuals who derive little to no benefit from hearing aids are being fit with cochlear implants instead. Cochlear implants are distinctly different from hearing aids because they do not amplify sound. Rather, cochlear implants provide electrical stimulation to auditory nerve fibers in order to elicit auditory percepts at the cortical level (Dorman & Wilson, 2004; House, Berliner, & Eisenberg, 1980).

The auditory input generated through cochlear implants is far more degraded and distorted than the input signals provided by hearing aids. In addition, some important clinical factors that influence the performance outcomes with a cochlear implant (e.g. active electrodes, electrode placement, survival of ganglion cells) are distinctly different than those influencing performance with a hearing aid. Taken together, the numerous and substantial differences between cochlear implants and hearing aids and the patients who use them make it difficult to draw any firm conclusions about the training effects, learning,
and transfer of learning that may be involved in using cochlear implants based on research examining learning and training effects in hearing aid users. Thus, an important and relatively new area of research has emerged to investigate auditory perceptual learning, training effects, and transfer of learning in cochlear implant users.

**Associations Among Speech and Nonspeech Tasks: Implications for Transfer of Learning**

The primary goal of most patients who receive cochlear implants is to improve their perception of speech. However, for some cochlear implant recipients who have been profoundly deaf for extended periods of time or who were implanted with single-channel devices, robust speech perception may be an unattainable goal, especially if visual cues from lip reading are unavailable. Rather, patients who fall into these categories may only be able to expect an awareness of the sounds within their environment (Teoh et al., 2004). In contrast, cochlear implant users already skilled at perceiving sound and speech may set more ambitious goals such as talking on a telephone or understanding and enjoying music. Because cochlear implant users are confronted with a wide range of auditory environments and activities in their daily life, it is important to understand how experience or training in one area may influence or interact with other processing domains.

Several recent studies have examined how cochlear implant users’ basic speech perception skills relate to other auditory skills such as environmental sound identification (Reed & Delhorne, 2005), music perception (Gfeller, Witt, Stordahl, Mehr, & Woodworth, 2000), speaker-gender identification (Fu, Chiachilla, & Galvin, 2004), and talker discrimination (Cleary, 2003; Cleary & Pisoni, 2002; Cleary, Pisoni, & Kirk, 2005). Not surprisingly, it has been found that basic speech perception and spoken word recognition skills are associated with other auditory capabilities in cochlear implant users.

In pediatric cochlear implant users, Cleary and Pisoni (2002) found that both open- and closed-set word recognition and open-set sentence recognition scores were strongly related to the ability to discriminate between speakers. This result was replicated by Cleary and her colleagues (2005) using an adaptive procedure to systematically vary voice similarity in pairs of test stimuli. The pediatric cochlear implant users who were best at discriminating similar voices also performed more accurately in a keyword identification task. Pediatric cochlear implant users’ abilities to identify the gender of a speaker have also been found to be related to their spoken word recognition skills (Osberger, Miyamoto, Zimmerman-Phillips, Kemink, Stroer et al., 1991; Staller, Dowell, Beiter, & Brimacombe, 1991).

Although adult cochlear implant users’ abilities to identify speaker-gender have been examined along with basic speech perception skills, the specific relationship between these two abilities has not been conclusively determined in these listeners (Fu et al., 2004). In their study of six adults with cochlear implants and normal-hearing, Fu and colleagues (2004) found that both vowel recognition and identification of speaker-gender improved when more spectral information was made available using an acoustic simulation of a cochlear implant by adding additional channels. However, only speaker-gender identification improved when more temporal information was made available by altering the filter cutoff frequencies in the acoustic model.

Because of the small sample size used in this study, no correlational or regression analyses could be conducted to assess the relationship between vowel recognition and identification of speaker-gender in adult cochlear implant users. In addition, no published studies appear to have been conducted on the talker discrimination abilities of adult cochlear implant users. A limited amount of research on this topic has revealed that adult cochlear implant users find it very difficult to discriminate between speakers, especially when the speakers are of the same gender and when there is linguistic variability within the
pairs of test stimuli (see Kirk, Houston, Pisoni, Sprunger, & Kim-Lee, 2002; McDonald, Kirk, Krueger, & Houston, 2003). In addition, McDonald and his colleagues (2003) found that adult cochlear implant users had difficulties identifying speakers’ voices from a closed-set of response alternatives. Except for these studies, very little research on gender and speaker identification has been carried out in cochlear implant users.

Thus, the sensory and neural mechanisms adult cochlear implant users use to learn to discriminate gender and identity of speakers has not been explored in any depth. Similarly, it is also unknown whether auditory perceptual learning for other speech related skills is transferable to identifying and discriminating indexical properties of speech or whether learning to identify and discriminate these properties transfers to speech perception and spoken word recognition. Measuring the transfer of learning with these auditory tasks may have important implications for assessing what training and rehabilitative strategies would be most robust and successful in cochlear implant users.

Although examinations of cochlear implant users’ abilities to identify speaker- gender and to discriminate among speakers have been sparse, several investigations of cochlear implant users’ abilities to identify and understand nonspeech stimuli have been conducted. Both music perception and environmental sound identification have recently been examined in adult cochlear implant users (Gfeller, Knutson, Woodworth, Witt, & DeBus, 1998; Gfeller et al., 2002; Kong, Cruz, Jones, & Zeng, 2004; Reed & Delhomme, 2005). In one study, Gfeller and colleagues (2002) found that adult cochlear implant users’ closed-set identification of musical instruments was related to listeners’ accuracy in identifying vowels in the clear and in noise. Musical instrument identification was also related to listeners’ abilities to identify words and sentences in noise. In contrast, appreciation ratings assigned to musical instruments by adult cochlear implant users did not relate to the listeners’ speech perception scores in this study. These results suggest an interesting relationship between speech perception and music perception that appears to be dependant on whether the assessment of music is objective and based on the accurate identification of instruments or musical tunes or subjective and purely based on the amount of enjoyment cochlear implant users derive from listening to music.

Recent research has also examined environmental sound identification in adult cochlear implant users. Using a closed-set and categorically organized environmental sound identification task, Reed and Delhorne (2005) found that adult cochlear implant users’ identification of environmental sounds was related to their ability to identify monosyllabic words in isolation. The authors used perceptual confusion data and a hierarchical clustering analysis to determine the acoustic characteristics of sounds that were the most perceptually salient to the listeners.

Reed and Delhorne (2005) found that all cochlear implant users in their study could accurately identify sounds that had highly distinctive temporal-envelope characteristics (e.g. footsteps, slamming door). In contrast, sounds whose spectral properties were more distinctive than temporal properties were harder for the cochlear implant users to identify (e.g. air conditioner, dishwasher). However, cochlear implant users who were considered high performers, according to their open-set word recognition abilities, identified these more difficult sounds better than the lower performing users. Their results suggest that the perception of speech and nonspeech sounds may be closely linked in cochlear implant users and may draw on common processing resources.

Although it has been determined that speech perception skills relate both to speech related skills that don’t require accurate encoding of the spoken message (e.g. speaker-gender identification, speaker discrimination) and to nonspeech auditory identification and discrimination skills (e.g. music appreciation, sound identification), it is unclear whether auditory training designed to improve one skill
is transferable to other skills. The current study was designed to examine the transfer of auditory perceptual learning in normal-hearing adults trained to identify speech processed by an acoustic simulation of a cochlear implant. Specifically, the ability of listeners to transfer learning from a sentence transcription task to an environmental sound identification test, speaker-gender identification, and speaker discrimination was examined.

**Studying Non-speech Auditory Processing Skills using Cochlear Implant Simulations**

Acoustic simulations of cochlear implants have been useful tools to study many aspects of speech perception (i.e. Dorman, Loizou, & Rainey, 1997; Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995; Shannon, Fu, & Galvin, 2004), perceptual learning (Burkholder, 2005; Davis et al., 2005; Fu & Galvin, 2003), and adaptation (Fu & Shannon, 1999a; Rosen, Faulkner, & Wilkinson, 1999) in normal-hearing listeners. Recently, research with acoustic simulations of cochlear implants has been broadened to include examinations of auditory skills that do not involve speech recognition or explicit identification of a spoken message. Music perception (Burns, Sanborn, Shannon, & Fu, 2001; Kong et al., 2004; Shannon, Fu, & Galvin, 2004; Smith, Delgutte, & Oxenham, 2002), speaker-gender identification, and speaker discrimination (Gonzalez & Oliver, 2005) have recently been examined in studies using acoustic simulations of cochlear implants in normal-hearing listeners.

**Music Perception.** Perceiving music is one nonspeech task that has proven to be very difficult for normal-hearing listeners exposed to acoustic simulations of cochlear implants. Burns and colleagues (2001) found that listeners needed more than 15 channels to accurately identify simple and familiar isochronous (lacking rhythm cues) melodies processed through an acoustic simulation of a cochlear implant. In listening conditions in which target melodies were masked with competing music, normal-hearing listeners needed between 32 and 48 channels to accurately identify the melodies (Smith et al., 2002). Even with 48 channels of spectral information available, normal-hearing listeners reported that the music sounded very poor. Thus, more spectral channels may be needed for normal-hearing listeners to enjoy music that they hear through acoustic simulations of cochlear implants, especially if the music lacks rhythmic cues based on temporal variation.

When listening to music with rhythm cues through an acoustic simulation of a cochlear implant, normal-hearing listeners perform much better and even outperform cochlear implant users. Kong and colleagues (2004) found that, even with just one channel of spectral information available, normal-hearing listeners correctly identified rhythmic melodies 80% of the time. However, cochlear implant patients were only able to correctly identify 50% of the same melodies when they received only one channel of spectral information through their devices.

Kong’s and colleagues’ (2004) results confirm previous findings that music perception is one of the most difficult auditory tasks for cochlear implant patients even when temporal or rhythmic cues are available. Because music perception is hard for cochlear implant users but is still a skill that most listeners desire, an important direction in this area of research is determining how music perception and other nonspeech auditory skills can be efficiently and effectively improved. More thorough examinations of the transfer of auditory perceptual learning among auditory skills could be useful to determine efficient and successful methods of training for music perception and other nonspeech auditory skills.

**Talker-gender Identification and Talker Discrimination.** Recently, normal-hearing listeners' abilities to perceive indexical properties of speech while listening to acoustic simulations have also been examined (Fu et al., 2004; Gonzalez & Oliver, 2005). Identifying indexical properties of speakers such as age, gender, dialect, and emotional state does not specifically require that the spoken message be
understood. However, the ability to identify and use indexical information about speakers is an important part of spoken communication that could potentially aid listeners in identifying speech (Cleary et al., 2005).

Gender is one important indexical property that normal-hearing adults and children have no difficulty discriminating in clear listening conditions (Cleary & Pisoni, 2002; Wu & Childers, 2004). However, when normal-hearing adults are asked to identify the gender of speakers while listening to acoustic simulations of cochlear implants, their performance varies widely. The ability to identify the gender of speakers when listening to an acoustic simulation of a cochlear implant has been found to be dependant on the carrier signal used in the acoustic model, the number of channels, and the frequency of the envelope cutoff used in signal processing (Fu et al., 2004; Gonzalez & Oliver, 2005).

When only a small number (< 6) of channels are used in acoustic simulations of cochlear implants, gender identification is far worse than in clear listening conditions (Fu et al., 2004; Gonzalez & Oliver, 2005). This result indicates that spectral information is vital to determining the gender of speakers. Fu and colleagues (2004) also suggest that temporal information plays an important role in determining the gender of speakers. They varied the frequency of the envelope cutoff in order to modify temporal cues in their acoustic simulation of a cochlear implant. Listeners could more accurately identify gender when the envelope was cutoff at higher frequencies (160 Hz, 320 Hz) rather than at lower frequencies (20 Hz, 40 Hz, 80 Hz).

In addition to the spectral and temporal information available to listeners, the carrier signal used in acoustic simulations of cochlear implants also influences the ability to identify the gender of speakers. Fu and colleagues (2004) used a sinewave carrier in their acoustic model based on preliminary data indicating that, when only a few noise-band modulated channels were used, normal-hearing listeners performed near chance in a speaker-gender identification task. Chance performance in a speaker-gender identification task by normal-hearing listeners exposed to an acoustic simulation of a cochlear implant with multiple channels is far worse than the performance of cochlear implant patients with just one channel (Fu et al., 2004). Thus, cochlear implant simulations using sinewave carriers may more accurately represent how spectral degradation affects the ability to identify the gender of speakers.

Gonzalez and Oliver (2005) recently confirmed that normal-hearing listeners are better at identifying speaker-gender when listening to acoustic simulations using sinewave carriers rather than noise-band carriers. Using acoustic simulations of cochlear implants that ranged from 3 to 16 channels, they found that speaker-gender identification scores were consistently better when sinewave carriers were used. The benefit of sinewave acoustic simulations of cochlear implants was most evident when less than 8 channels were available. At 8 channels and above, speaker-gender identification was only slightly better when sinewave carriers were used to generate the acoustic simulation of the cochlear implant.

Aside from examinations about how different acoustic models of cochlear implants influence speaker-gender identification and speaker discrimination, there have been no other attempts to examine this skill or uncover its relationship with other speech perception skills in normal-hearing listeners exposed to spectrally degraded speech or in cochlear implant patients. Thus, it is currently unknown whether experience or training with speech processed through acoustic simulations of cochlear implants would result in learning that would transfer to speaker-gender identification or speaker discrimination tasks. The present study was designed to examine normal-hearing listeners' abilities to transfer auditory perceptual learning of spectrally degraded speech modeled after a cochlear implant to speaker-gender identification and speaker discrimination tasks.
Environmental Sound Identification. In addition, this study also examined transfer of auditory perceptual learning to environmental sound identification. Like speaker-gender identification and speaker discrimination, environmental sound identification is another auditory task that has not been fully examined in cochlear implant users or in normal-hearing listeners exposed to acoustic simulations of cochlear implants. Using noise-vocoded speech similar to that generated through acoustic simulations of cochlear implants, Gygi and colleagues (2004) assessed the importance of spectral-temporal factors in the identification of 70 environmental sounds in both experienced and naïve normal-hearing listeners. Their results indicate that the listeners were able to identify nearly half of the sounds based on temporal information alone. This finding is consistent with results found in cochlear implant patients (Reed & Delhorne, 2005). Thus, temporal information appears to be the most critical component to identifying environmental sounds in spectrally degraded conditions for both normal-hearing and hearing-impaired listeners.

Although Gygi and his collaborators (2004) suggested that environmental sounds are identified in a manner similar to speech sounds when both are spectrally degraded, they collected no speech perception data from the listeners in their study. Assessing speech perception in listeners trained to identify environmental sounds would be a useful way to assess transfer and the robustness of auditory perceptual learning. Similarly, assessing environmental sound identification in listeners trained to identify speech processed through an acoustic simulation of a cochlear implant could also provide valuable information about the transfer of training and specific characteristics of auditory perceptual learning.

Transfer of Auditory Learning and its Implications for Speech Specific Domains

Studying transfer of speech perception training to environmental sound identification and other nonspeech auditory tasks such as speaker-gender identification and speaker discrimination is both theoretically and clinically important. Theoretically, it is important to know if auditory perceptual learning from sentence transcription training will transfer to nonspeech tasks, because transfer is one measure of the robustness of perceptual learning. In addition, transfer of learning is an indication that learning was not specific to the training stimuli. Demonstrating robustness and lack of specificity and transfer of auditory perceptual learning in a sentence transcription task is significant, because it suggests that auditory perceptual learning in a speech processing task is not necessarily specific to the domain of speech.

Domain specificity and modularity in perception and cognition have historically been popular but controversial (Appelbaum, 1998; Fodor, 1983; 1989). A wide range of cognitive processes have been suggested to be driven by domain specific processing mechanisms rather than more global and general purpose mechanisms. For example, theory of mind (Scholl & Leslie, 1999), numerical knowledge (Gelman, 1998), spatial reorientation (Hermer & Spelke, 1996) and causality of movement (Gelman, 1990) have all been suggested to operate in a domain-specific manner. In addition, a nearly infinite amount of individually determined non-core domain specific skills such as making sushi and playing chess have also been proposed (Gelman, 1998).

However, among the many suggested domain specific cognitive skills, none seems to have drawn as much attention as language and speech processing. (Liberman, Cooper, Shankweiler, Studdert-Kennedy, 1967; Liberman & Mattingly, 1989; Lillo-Martin, 1999; Tager-Flusberg, Boshart, & Baron-Cohen, 1998; Trout, 2003). In fact, in Fodor’s original and subsequent discussion of modularity of the mind (1983; 1989), he suggested that speech fits many of the necessary requirements to be considered a
modular system. That is, he suggested that speech processing was fast, automatic, domain specific, characterized by information encapsulation, and localized within the brain.

The complex and variable nature of speech signals was one of the first characteristics that prompted speech researchers to propose that there was a special mechanism used for speech perception. Liberman and colleagues (1967) speculated that, in most cases, speech is articulated at rates too fast for the auditory system to process it effectively using typical general-purpose auditory processing capabilities. Therefore, they proposed that a special speech decoding mode was required to accommodate its complexity.

A “phonetic module” was specifically suggested by Liberman and Mattingly (1989) to increase the rate of information processing by the auditory system and to decode speech. This phonetic module was proposed to more clearly specify the special speech processor forming the foundation of the previously constructed “motor theory” of speech perception (Liberman et al., 1967; Liberman & Mattingly, 1985). The motor theory of speech perception suggests that the gestures accompanying the sounds of speech are processed in a special and innate mode similar to the phonetic module. Evidence for the phonetic module was provided by Liberman and Mattingly’s (1989) report of the existence of “duplex perception” that indicated that “speech and nonspeech percepts are different kinds of representations” (p 490). One primary difference suggested between speech and nonspeech perception is that speech stimuli are processed independently of pitch, loudness, and timbre, which are all proposed to be resolved within their own separate processing modules.

However, one complication for the proposal that speech required a special encoding module is that the auditory system has since been found to be capable of processing nonspeech signals that are modeled after the complex and rapid presentation of speech (Moore, 1997). Such findings suggest that the mechanisms used for the perception of speech and nonspeech sounds may be more similar and more closely linked than previously thought. An additional way to gauge the similarity or links between the perception of speech and nonspeech sounds is to determine if training listeners to perceive degraded speech stimuli will transfer to the perception of nonspeech sounds such as those heard in the environment. Such a study may provide further insights into whether auditory perceptual learning of speech has some domain specific properties or if it is a general processing mechanism extendable to other auditory stimuli.

In the present study, the transfer of auditory perceptual learning of spectrally degraded speech to spectrally degraded environmental sounds was assessed. Based on previous research suggesting that the perception of speech and nonspeech sounds may not be entirely distinct in either clear (Moore, 1997) or degraded auditory conditions (Gygi et al., 2004), it is hypothesized that transfer of learning from a speech transcription task to environmental sound identification will occur. This result was predicted because both speech and environmental sound perception have been found to be relatively preserved when spectral information is reduced as long as sufficient temporal cues are still available (Gygi et al., 2004; Reed & Delhorne, 2005; Shannon et al., 1995).

However, it is less likely that transfer of learning of spectrally degraded speech will be extended to speech related tasks that require fine-grained perception and discrimination of frequency cues. Previous studies of perceptual learning with acoustic simulations of cochlear implants suggest that listeners learn primarily about the abstract spectro-temporal information within the speech signal rather than the fine acoustic-phonetic structure (Burkholder, 2005; Davis et al., 2005). However, speaker-gender identification and speaker discrimination both rely more on the ability to resolve finer acoustic cues such as frequency and formant information rather than temporal information. Thus, in the present study, we
did not expect that transfer of auditory perceptual learning would extend to speaker-gender identification and speaker discrimination even though they are speech related tasks.

Transfer of Auditory Learning and its Clinical Implications for Cochlear Implant Users

The results of the present study will not only be of theoretical interest. It is also clinically important to determine if normal-hearing listeners can transfer what they learn about speech processed through an acoustic simulation of a cochlear implant to environmental sounds. Speech perception and environmental sound identification are both important skills for deaf cochlear implant users. However, for some cochlear implant users, perceiving sounds and speech is rather difficult (Reed & Delhorne, 2005). Thus, determining ways in which both these skills can be improved in cochlear implant patients is an important clinical goal.

Any evidence of transfer of learning from a speech perception task to environmental sound identification by normal-hearing listeners exposed to an acoustic simulation of a cochlear implant would suggest that separate training programs for each skill may not be necessary. Rather, such results would suggest that speech-based training that is transferable to other speech and nonspeech auditory skills may be the most efficient method of developing the hearing abilities of deaf cochlear implant users.

It is likely that most adults who receive cochlear implants without any formal training unconsciously focus their efforts on improving their speech perception skills. Thus, due to their own instinctive efforts, adults may gradually become better at perceiving both speech and sound. However, in rehabilitative settings with pediatric cochlear implant users, it is not uncommon for clinicians first to teach children the explicit awareness of auditory percepts through sounds in the environment instead of speech. Children may spend time on “listening walks” that alert them to the sound of birds, cars, and whistles (Robbins, 1998). While the awareness of sounds and the ability to identify them may certainly be important for the children’s safety and curiosity, the ability to communicate effectively using speech and language could arguably be the most important for deaf children to develop in a timely manner.

Thus, immerging young infants in speech and explicit training later-implanted children to perceive speech immediately after cochlear implant activation may have the most benefit on their oral communication skills. In addition, if speech perception training transfers or carries over to other skills such as identifying sounds, it could undoubtedly be one of the most efficient methods of developing cochlear implant users’ general auditory skills as well. The goal of the present study was to determine if normal-hearing adult listeners given speech transcription training can transfer learning to several speech and nonspeech tasks. Transfer of learning to environmental sound identification, speaker-gender identification, and speaker discrimination was specifically tested. Although the present study examines transfer of auditory perceptual learning in normal-hearing adults listening to an acoustic simulation of a cochlear implant, the results of this study should have important implications for the post-implant rehabilitative methods used with both adult and pediatric cochlear implant users.

Method

Experimental Design

Figure 4.1 shows a detailed schematic representation of the experimental design used. The experimental procedure was divided into five phases: (1) Familiarization, (2) Pre-training, (3) Training, (4) Post-training, and (5) Transfer. The specific procedures carried out during these five phases will be described in detail in the Experimental Procedures section. Two groups of participants were used. The groups differed according to whether or not they were trained to identify speech processed through an
acoustic simulation of a cochlear implant. One group of listeners received speech transcription training on sentences processed through the acoustic simulation of the cochlear implant. The control group listeners simply transcribed the same sentences in their clear, unprocessed form.

<table>
<thead>
<tr>
<th>Passive Listening</th>
<th>Sentence transcription</th>
<th>(1) Experimental Group</th>
<th>Sentence transcription</th>
<th>Transfer Tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nursery rhymes with text</td>
<td>20 meaningful sentences</td>
<td>Sentence transcription training</td>
<td>100 Harvard sentences</td>
<td>(1) Sound Identification</td>
</tr>
<tr>
<td>NO feedback</td>
<td>NO feedback</td>
<td>Processed stim + text given as feedback</td>
<td>NO feedback</td>
<td>4 categories of 10 sounds with 3 tokens apiece</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(2) Control Group</th>
<th>Sentence transcription</th>
<th>(2) Gender Identification</th>
<th>50 different Male Harvards</th>
</tr>
</thead>
<tbody>
<tr>
<td>100 Harvard sentences in the clear</td>
<td></td>
<td></td>
<td>5 speakers x 10 tokens</td>
</tr>
<tr>
<td>NO Exposure</td>
<td></td>
<td></td>
<td>50 different new Female Harvards</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5 speakers x 10 tokens</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(3) Talker Discrimination</th>
<th>50 “same”</th>
<th>50 “different”</th>
</tr>
</thead>
<tbody>
<tr>
<td>15 male1 – male1</td>
<td>15 male2 – male2</td>
<td>15 male1 – male1</td>
</tr>
<tr>
<td>15 female1 – female1</td>
<td>15 female2 – female2</td>
<td>15 female1 – female2</td>
</tr>
<tr>
<td>15 female2 – female1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NO feedback provided

**Figure 4.1.** Schematic representation of the experimental procedure used in the current study.

**Participants.** Twenty-four young adult participants were included in each of the two groups. The participants were recruited from a database containing email addresses of undergraduate students who had previously indicated in an online Psychology survey that they were interested in participating in paid studies. All participants were monolingual native speakers of American English and reported that they were free of any speech, hearing, language, and attentional disorders. Participants were paid $12 for their participation in this study.

**Stimuli and Materials**

**Stimuli.** Well known nursery rhymes (e.g. Jack and Jill; Humpty Dumpty) were used for a brief period of familiarization with the acoustic simulation. The stimuli used before, during, and after training were taken from the Harvard sentence set (IEEE, 1969). Harvard sentences are between seven and 11 words long and contained five key content words that were scored during data analysis. One set of 20 Harvard sentences was used in the pre- and post-training phases. 100 different Harvard sentences were used during training. The sentences used in the current study are listed in Appendix C.
All the familiarization stimuli and the Harvard sentences were recorded by the author, who is a female speaker with a Midwestern dialect. Recordings were made in a sound-attenuated booth (IAC Audiometric Testing Room, Model 402A) using a Shure head-mounted microphone (SM98) placed approximately two inches from the mouth. Recordings were digitized online (16-bit analog-to-digital converter (DSC Model 240)) at 22,050 Hz and stored directly on a personal computer in .wav format using an individualized version of a speech acquisition program (Dedina, 1987). Three tokens of each stimulus were recorded. The repetition judged to be of the highest quality by the experimenter was used.

Using sound editing software (Adobe Audition) all usable sound files were segmented to the zero crossing closest to the beginning and end of the waveform. The root mean square (RMS) amplitude for all the files was then obtained. Based on the RMS amplitude measure, all stimuli were equated to an amplitude of 65 db(A) using the Level16 software program (Tice & Carrell, 1998).

In a pilot study, ten listeners were asked to transcribe 100 sentences spoken by the speaker used in this study to establish speech intelligibility. Results showed that nearly 95% ($M = 94.72, SD = 1.23$) of key words were identified correctly when listeners heard sentences produced by this speaker in the clear. This analysis indicates that the speaker has intelligible speech. This is an important finding, because it indicates that any difficulties that listeners may have in identifying the speech when it is processed by the acoustic simulation of a cochlear implant are due to the nature of the signal degradation techniques and not due to initially poor speaker intelligibility.

For each transfer task, a different set of previously compiled stimuli was used. Environmental sound identification was measured using a set of 120 tokens of 40 different sounds compiled by Reed and Delhorne (2003). The 40 sounds were divided into four different categories according to the environment in which the sounds are usually heard. The four categories used were: General Home, Kitchen, Office, and Outdoors. Table 4.1 lists the sounds included in each sound category and identifies those that have distinct transient or temporal properties according to Reed and Delhorne (2005).

**Table 4.1.** List of the sound categories and sounds used in the environmental sound identification task. Sounds classified as having transient or distinct temporal properties appear in italics (Reed & Delhorne, 2005).

<table>
<thead>
<tr>
<th>General Home</th>
<th>Kitchen</th>
<th>Office</th>
<th>Outside</th>
</tr>
</thead>
<tbody>
<tr>
<td>Air conditioner</td>
<td>Smoke alarm</td>
<td>Paper rustling</td>
<td>Birds singing</td>
</tr>
<tr>
<td>Baby crying</td>
<td>Cat meowing</td>
<td>File-door slamming</td>
<td>Keys jangling</td>
</tr>
<tr>
<td>Dog barking</td>
<td><strong>Cupboard-door slamming</strong></td>
<td>Typing on keyboard</td>
<td>Dog barking</td>
</tr>
<tr>
<td>Male voice</td>
<td>Dishes clanging</td>
<td>Telephone</td>
<td>Helicopter</td>
</tr>
<tr>
<td>Music</td>
<td>Dishwasher</td>
<td>Photocopier</td>
<td>Airplane</td>
</tr>
<tr>
<td>Toilet flushing</td>
<td>Doorbell</td>
<td>Water cooler</td>
<td>Fire siren</td>
</tr>
<tr>
<td>Rain</td>
<td><strong>Footsteps</strong></td>
<td>Smoke alarm</td>
<td>Car horn</td>
</tr>
<tr>
<td><strong>Knock on door</strong></td>
<td>Oven timer</td>
<td>Crowd talking</td>
<td>Thunder</td>
</tr>
<tr>
<td>Vacuum cleaner</td>
<td>Telephone</td>
<td><strong>Footsteps</strong></td>
<td>Babbling brook</td>
</tr>
<tr>
<td><strong>Door closing</strong></td>
<td>Running water</td>
<td>Fan</td>
<td>Car starting</td>
</tr>
</tbody>
</table>

The sounds that were used in each category were originally chosen because they were judged to be sounds that are commonly heard in daily life and that represent a broad range of types of sounds classified by Schafer (1977) and Ballas (1993). Ten different sounds were presented in each category. Several sounds (i.e., footsteps; dog barking) were used in multiple categories. For each of the ten sounds within a category, three different tokens of the sound were played. To include different tokens of each

791
sound, sounds were obtained from different sources or separate sound files were created by segmenting one sound file at several different places in the recording (see Reed & Delhorne, 2003).

The results reported by Reed and Delhorne (2003) demonstrated that normal-hearing listeners were able to identify the current set of sounds with near-perfect accuracy when they were presented in the clear. According to the procedures of Reed and Delhorne (2003), the sound stimuli were not leveled for amplitude in the current study, because intensity levels may provide valuable information when listeners are required to identify sounds.

The stimuli used in the gender identification and speaker discrimination task were taken from the Indiana Multitalker Database (Karl & Pisoni, 1994) which is a collection of 20 male and 20 female speakers each uttering the same 100 sentences taken from the Harvard sentence list. Each speaker in this database was a young adult and a monolingual, native speaker of American English who had resided in Indiana for his or her entire life. The intelligibility of each speaker’s speech also was evaluated by Bradlow, Torretta, and Pisoni (1996) who used normal-hearing listeners to transcribe the sentences uttered by each speaker.

The five most intelligible male and female speakers were selected for use in the gender identification task. Ten utterances from each of the ten speakers were used which resulted in a set of 100 sentences. Fifty sentences were spoken by male speakers, and 50 sentences were spoken by female speakers. Each of the sentences was a novel utterance. All utterances used for this task were leveled to the same amplitude as the generalization, pre- and post-training, and training stimuli, based on the root-mean-square (RMS) amplitude values of all stimuli.

The two most intelligible male and female speakers from the Indiana Multitalker Database (Karl & Pisoni, 1994) were used for the speaker discrimination task. Thus, the speakers used in the speaker discrimination task were also used in the gender identification task. However, the same token was never used within or between the two tasks. Two-hundred-forty novel sentences where combined into 120 stimulus pairs for the speaker discrimination task. All utterances used for this task were leveled to the same amplitude as the generalization, pre- and post-training, training, and gender identification stimuli based on the root-mean-square (RMS) amplitude values of all stimuli.

**Simulation Strategy.** All auditory stimuli were processed offline using a personal computer. The signal processing procedure used for the cochlear implant simulation was adapted from the real-time signal processing methods designed by Kaiser and Svirsky (2000). The signal was lowpass filtered with a cutoff frequency of 12,000 Hz. A bank of eight bandpass analysis filters was then used to divide the auditory signal into its corresponding frequency components. The temporal envelope was extracted from the output of the filters and then used to modulate noise bands created through a bank of eight synthesis filters with the same cutoff frequencies as the original analysis filters. The cutoff frequencies for the analysis and synthesis filter banks are listed in Table 4.2.

Because the frequency of the noise bands and the output of the filters were matched, this acoustic model most accurately represents the input transmitted to cochlear implant patients who hypothetically have deep enough electrode insertions to enable well matched tonotopic frequency and electrode placement. Such a placement would presumably result in little to no basalward frequency shift. In other circumstances, a frequency shift may result from cochlear implantation if electrodes are unable to be placed far enough into the apical end of the cochlea which is tonotopically tuned for low frequency sounds. When electrodes cannot be implanted this deep into the apex and are instead placed more basalward within the cochlea, electrical stimulation for low frequency sounds is directed to neural fibers
ordinarily tuned for somewhat higher frequency sounds. This frequency mismatch results in sounds that are perceived to be higher frequency than normal.

Table 4.2. Frequency boundaries of the analysis and synthesis filters used in the acoustic model of the cochlear implant.

<table>
<thead>
<tr>
<th>Frequency Boundaries</th>
</tr>
</thead>
<tbody>
<tr>
<td>854 – 1467 Hz</td>
</tr>
<tr>
<td>1467 – 2032 Hz</td>
</tr>
<tr>
<td>2032 – 2732 Hz</td>
</tr>
<tr>
<td>2732 – 3800 Hz</td>
</tr>
<tr>
<td>3800 – 5150 Hz</td>
</tr>
<tr>
<td>5150 – 6622 Hz</td>
</tr>
<tr>
<td>6622 – 9568 Hz</td>
</tr>
<tr>
<td>9568- 11000 Hz</td>
</tr>
</tbody>
</table>

The choice of an eight channel acoustic model was based on studies showing that adult cochlear implant users achieve asymptotic speech perception scores when they have this number of active channels (Fishman, Shannon, & Slattery, 1997; Friesen, Shannon, Baskent, & Wang, 2001). Additionally, the best adult cochlear implant patients achieve word identification scores similar to those obtained by normal-hearing listeners exposed to an 8-channel acoustic simulation of a cochlear implant, when the acoustic model is implemented without any frequency shift (Dorman, Loizou, Fitzke, & Tu, 2000).

Experimental Procedures

Participants were tested in individual perceptual testing stations in the Speech Research Laboratory at Indiana University. Each booth was equipped with a Gateway PC (P5-133) using a 15" monitor (Vivitron15). Auditory stimuli were presented over calibrated headphones (Beyer Dynamics DT100) at approximately 70dB(A) SPL as measured by a standard sound-level meter. Stimuli were always presented in random order within blocks to the participants.

Familiarization phase. Testing began with a brief period of familiarization in which participants simply listened to and read silently along with six popular nursery rhymes. The text of these nursery rhymes appeared on the top half of the monitor in a 30pt. sans serif font 500ms prior to the start of each utterance. The text of the nursery rhymes remained visible for the entire duration of the utterance. Listeners did not have the option to replay any of the familiarization stimuli.

Pre- and post-training phase. During pre- and post-training phases, listeners transcribed 20 Harvard sentences by typing their responses onto the computer’s keyboard. Responses could be entered immediately after each utterance ended and when a prompt and cursor appeared on the monitor. The same set of 20 sentences was used for both pre- and post-training. Participants did not receive any feedback during these two phases and did not have the option to repeat the sentence after it was played.
**Training phase.** During the training period, one group of listeners heard 100 Harvard sentences in their processed form. Listeners entered their responses following the same procedures used in the pre- and post-training phases. Feedback was presented to these listeners by playing the processed sentence again after the written text of the sentence appeared on the screen. The text of the sentence appeared 250ms prior to the onset of each utterance and remained visible until each utterance stopped playing. Listeners did not have the option of replaying the feedback. The decision to use this type of feedback was based on the findings of an earlier study which indicated that this form of feedback results in the largest learning effects when compared to auditory-alone feedback using several control groups receiving either no feedback or no exposure to the degraded speech (Burkholder, 2005).

A control group was also used in this study. These listeners heard the same 100 Harvard sentences in their unprocessed form. They were required to transcribe what they heard in each sentence following the same procedures used in the pre- and post-training phases. No feedback was provided to these listeners.

**Transfer tasks.** Three separate tasks: (1) environmental sound identification (2) speaker-gender identification, and (3) speaker discrimination were used to measure transfer of learning. Environmental sound identification was the first transfer task presented to the listeners. The presentation of the speaker-gender identification and speaker discrimination task was counterbalanced among listeners. Half of the listeners completed the gender identification task first. Half of the listeners completed speaker discrimination task first. No feedback was provided to the listeners during any of the tasks.

The environmental sound identification task was conducted using procedures developed by Reed and Delhorne (2005). A 10-alternative forced-choice procedure was used. The 10 sound alternatives appeared on the computer terminal in numbered order prior to the start of the first sound and remained displayed until all 30 tokens of the sounds from the current sound category had been played.

Listeners provided responses after each sound had been completely played and a cue and cursor appeared at the bottom of the computer screen. Listeners indicated the sound they heard by choosing the number that corresponded to the selected sound. To enter their response, they simply pressed the keypad number that was paired with the sound. Each answer was recorded after the listener pressed the carriage return to initiate the next trial. The four sound categories were presented in counterbalanced order to the listeners. Listeners could not replay any of the sounds at any time.

In the speaker-gender identification task, listeners heard 50 sentences spoken by five different male speakers and 50 sentences spoken by five different female speakers. Each sentence was played to the listeners once. Listeners provided responses after each sentence had been completely played and a cue and cursor appeared at the bottom of the computer screen. Listeners indicated the gender of the speaker by selecting the number that was paired with the gender. The gender-number pairings remained visible on the computer monitor throughout the entire task. Each answer was recorded after the listener pressed the carriage return to initiate the next trial. Listeners could not replay any of the sentences.

During the talker discrimination task, pairs of utterances taken from two male and two female talkers were played for the listeners. There were 60 pairs in which both tokens were spoken by the same talker. Each of the four talker’s voice was used to make 15 same-talker pairs. There were also 60 tokens spoken by different talkers of the same sex. The order of the talkers was counterbalanced in the pairs so that each talker’s utterances were presented in the first and second position an equal number of times.
After the first sentence of each pair was played, there was a 1000 ms pause before the second sentence played. Upon completion of the second sentence in the pair a cue and cursor appeared on the bottom of the computer screen signaling listeners to respond. Listeners indicated whether they thought the sentence-pair they heard was spoken by the same talker or by two different talkers by choosing the number that was assigned to that response. The response-number pairings remained visible throughout the entire task on the computer monitor. Each answer was recorded after the listener pressed the carriage return to initiate the next trial. Listeners could not replay individual sentences or sentence pairs at any time.

Data Analysis

The data output files containing the correct responses and each participant’s responses for each experimental phase were imported into data spreadsheets for scoring. For the pre-training, post-training, and training phases, each sentence was scored by determining the percentage of key words that were identified correctly. Keywords containing obvious spelling and typographical errors were scored as correct. Errors in spelling were scored as correct only if the response did not resemble a different word or combination of words. Similarly, typing errors were only scored as correct if the response did not resemble a different word.

A typing error was scored as correct if a target letter was substituted by any letter immediately surrounding it or if letters immediately surrounding the target letter were inserted into the response. Responses in which the correct letters were transposed were also considered typographical errors and scored correctly. Insertions of unnecessary spaces within words were also scored as correct if the response did not result in a different word. Homophones (i.e. “rain” for “reign”) of key words were also accepted as correct. However, key words reported in an incorrect tense (i.e. “place” for “placed”) or with other incorrect affixes (i.e. “shoes” for “shoe”) were scored as incorrect.

Scoring was completed by the experimenter and a trained undergraduate research assistant. Inter-rater reliability was established by computing the Cronbach’s alpha value of scores assigned to all the responses given by the first ten listeners in the study (Bland & Altman, 1997). The Cronbach’s alpha value obtained was .97, indicating that inter-rater reliability for scoring key words correct in this study was very high. Any ambiguous scoring decisions were resolved at the experimenter’s discretion.

The data from the three transfer tasks were scored automatically in spreadsheet software using a simple algorithm that compared the correct numerical responses with the listeners’ responses. Matches between the correct and recorded responses were scored as correct. Mismatches between the correct and recorded responses were scored as incorrect. The percentage of correct responses was the final score calculated for each listener in the environmental sound and gender identification and talker discrimination tasks.

Results

The descriptive statistics for each groups’ performance on each phase of the experiment appear in Table 4.3. To determine if both groups of listeners had roughly the same abilities to identify speech processed through an acoustic simulation of a cochlear implant prior to training, an independent-samples T-test was conducted on the pre-training scores. This test revealed no significant difference ($t(46) = -0.81$, $p = .422$) in the pre-training scores, indicating that the two groups were similar to one another prior to training.
Table 4.3. Means and standard deviations of trained and untrained listeners’ performance before, during, and after training and on three transfer tasks.

<table>
<thead>
<tr>
<th>Group</th>
<th>Pre-Training</th>
<th>Training</th>
<th>Pst-Training</th>
<th>Sound ID</th>
<th>Gender ID</th>
<th>Talker Discrimination</th>
</tr>
</thead>
<tbody>
<tr>
<td>Untrained</td>
<td>.37 (.13)</td>
<td>.97 (.02)</td>
<td>.48 (.14)</td>
<td>.51 (.11)</td>
<td>.76 (.10)</td>
<td>.57 (.09)</td>
</tr>
<tr>
<td>Trained</td>
<td>.40 (.09)</td>
<td>.52 (.16)</td>
<td>.61 (.09)</td>
<td>.57 (.06)</td>
<td>.77 (.09)</td>
<td>.56 (.07)</td>
</tr>
</tbody>
</table>

Pre- and Post-training Data

Figure 4.2 shows the pre- and post-training scores of the two groups of 24 listeners. The scores reflect the percentage of keywords reported correctly. To assess the magnitude of improvement that occurred from pre- to post-training, a repeated-measures ANOVA was conducted using pre- and post-test transcription keyword scores as the two levels of the within-subjects factor. The between-subjects factor was the listener group. Results indicate a significant main effect of session (pre-test vs. post-test) \( F(1, 46) = 433.59, p < .001 \) and group \( F(1, 46) = 5.79, p = .020 \). The interaction was also significant \( F(1, 46) = 45.08, p < .001 \) indicating that listeners who received training improved more from pre- to post-training than the listeners who received no exposure during the training period.

Figure 4.2. Pre- and Post-training performance of untrained and trained listeners. The change in percentage points from pre- to post-training is shown above the bars representing each group. Error bars represent the standard error of the mean.

Training and Learning Data

Because the amount and time course of learning occurring during the training period played a large role in pre- to post-training increases in performance, performance during training and learning was evaluated. Figure 4.3 illustrates the individual learning curves of the 24 listeners within the trained group which includes performance before, during, and after training. Values for performance during each phase of training were obtained by splitting the 100 training sentences into fifths that each contained 20 sentences.
Figure 4.3. Individual learning curves for the 24 trained listeners. Error bars represent the standard error of the mean.

The training data of the trained listeners were also analyzed to determine the amount of improvement during the training period. The overall average percent correct for each phase of training was then determined. A repeated-measures ANOVA was conducted on the trained group's performance using the first, middle, and final phases of training as the repeated measures. Results indicate a main effect of training phase ($F(2, 46) = 7.81, p = .001$).

Although no training data were obtained from the untrained group, their responses to the unprocessed stimuli were still scored. Listeners in the untrained group reported keywords in the unprocessed sentences with nearly 100% accuracy ($M = .97, SD = .02$). This result indicates that this set of stimuli was highly intelligible prior to being processed through the acoustic simulation of the cochlear implant.

Transfer tasks

Table 4.3 provides a summary of the two groups' performance on the three transfer tasks. Figure 4.4 illustrates group performance on the environmental sound identification, talker-gender identification, and talker discrimination tasks. Independent-samples T-tests were used to determine the effect of speech transcription training on listeners' performance on these tasks.

Figure 4.4. Performance of trained and untrained group on closed-set environmental sound identification, gender identification, and speaker discrimination. Error bars represent the standard error of the mean. Significant group differences are marked with an asterisk.
**Environmental sound identification.** The group difference in the sound identification task was significant ($t(46) = -2.05, p = .046$). Listeners who received training ($M = .57, SD = .06$) performed better on the environmental sound identification task than listeners who received no training to the spectrally degraded speech ($M = .51, SD = .11$).

Figure 4.5 illustrates group performance on each separate environmental sound category. The listeners' abilities to identify environmental sounds were analyzed further by using repeated-measures ANOVAs and independent-samples T-tests to compare the two groups' performance on individual environmental sound categories. This analysis was conducted to determine if the differences between the two groups' abilities to identify environmental sounds may have been category specific.

![Diagram showing percent correct for trained and untrained listeners in different sound categories.](image)

**Figure 4.5.** Closed-set environmental sound identification scores of untrained and trained listeners according to sound category. Error bars represent the standard error of the mean. Significant group differences are marked by an asterisk.

The repeated-measures ANOVA indicated a main effect of sound category ($F(3, 138) = 13.47, p < .001$). The interaction between sound category and group was also significant ($F(3, 138) = 2.81, p = .042$). Significant group differences were observed in only one of the four sound categories. Listeners who received training ($M = .64, SD = .10$) identified outdoor sounds ($t(46) = -2.94, p = .005$) sounds significantly better than listeners who received no training ($M = .54, SD = .14$). Trained listeners ($M = .61, SD = .14$) performed better than untrained listeners ($M = .52, SD = .18$) on office sounds as well ($t(46) = -1.83, p = .074$). However, this difference did not meet the conventional level of significance. No group differences were found for sounds classified into the home ($t(46) = .57, p = .58$) or kitchen ($t(46) = -1.05, p = .301$) category.

Figures 4.6-4.9 illustrate group performance on the individual sounds included in each sound category. Repeated-measures ANOVAs were run to find the overall effects of sound type and training groups. Independent-samples T-tests were run on each sound to determine group differences. The Bonferroni correction method was used to adjust the significance level of these tests because more than four multiple comparisons were made (Hochberg, 1988). The adjusted significance level was determined by dividing the traditional significance ($p = .05$) by the number of comparisons ($c = 10$). Thus, the adjusted $p$-value was .005.

Figure 4.6 illustrates group means for the ten outdoor sounds. Results of the repeated-measures ANOVA indicated a main effect of sound ($F(9, 414) = 41.23, p < .001$). The interaction was not significant ($F(9, 414) = .44, p = .92$). Although listeners who received training had higher scores on
almost all of the outdoor sounds, there were no significant group differences among any of the individual sounds. Differences in performance on the dog bark (t(30.47) = -1.92, p = .061) and airplane sounds (t(46) = -1.81, p = .076) approached significance.

**Figure 4.6.** Performance on individual sounds from the outdoor category by untrained and trained listeners. Error bars represent the standard error of the mean. Group differences that approached significance are marked with an asterisk.

Figure 4.7 illustrates group means for the ten office sounds. Results of the repeated-measures ANOVA indicated a main effect of sound (F(9, 414) = 19.72, p < .001). The interaction between sound type and listener group approached significance (F(9, 414) = 1.64, p = .111). Listeners who received training performed significantly better on the keyboard (t(46) = -2.77, p = .008) sound. Trained listeners’ performance on the fan sound was better than untrained listeners’ performance (t(46) = -2.17, p = .035). However, this difference did not reach the conventional level of significance. The group difference for sounds made by a smoke detector approached significance (t(46) = -1.51, p = .139). The two most accurately identified sounds for both the trained and untrained listeners were the file door slam and footsteps. Both sounds were classified as having distinct transient and temporal properties.

**Figure 4.7.** Performance on individual sounds from the office category by untrained and trained listeners. Error bars represent the standard error of the mean. Significant group differences are marked by a double asterisk. Group differences approaching significance are marked with a single asterisk.
Figure 4.8 illustrates group means for the home sounds. Results of the repeated-measures ANOVA indicated a main effect of sound ($F(9, 414) = 40.18, p < .001$). The interaction between sound type and listener group was not significant ($F(9, 414) = .72, p = .688$). Surprisingly, untrained listeners performed better on the air conditioner sound than trained listeners ($t(31.56) = 2.57, p = .015$). However, this difference did not meet the adjusted level of significance. The transient sounds in this category (knock, closing door) were both identified with a high degree of accuracy by both groups of listeners.

![Graph showing performance on individual home sounds](image)

**Figure 4.8.** Performance on individual sounds from the home category by untrained and trained listeners. Error bars represent the standard error of the mean. Group differences that approached significance are marked by an asterisk.

Figure 4.9 illustrates group means for the kitchen sounds. Results of the repeated-measures ANOVA indicated a main effect of sound ($F(9, 414) = 430.52, p < .001$). The interaction between sound type and listener group approached significance ($F(9, 414) = 1.90, p = .051$). The group differences for the ringing phone ($t(46) = -1.98, p = .054$) and clanging dishes ($t(46) = 2.57, p = .030$) approached significance. The two most accurately identified sounds were footsteps and the cupboard door slam. Both of these easily identified sounds had distinct transient and temporal properties.

![Graph showing performance on individual kitchen sounds](image)

**Figure 4.9.** Performance on individual sounds from the Kitchen category by untrained and trained listeners. Error bars represent the standard error of the mean. Group differences that approached significance are marked by an asterisk.
Correlates of environmental sound identification

To further specify the relationship between speech transcription training and environmental sound identification, simple correlational analyses were conducted. The Bonferroni method of adjustment was used to obtain the significance level appropriate for the number of correlations run. The left panel in Figure 4.10 illustrates the correlation between trained listeners’ performance during training and their ability to identify all environmental sounds. This correlation did not reach statistical significance \( r = .272, p = .198 \).

![Graph](image)

**Figure 4.10.** Correlations between trained listeners’ training performance and (a) identification of all environmental sounds and (b) office sounds.

Additional correlations were conducted to determine if training performance was related to trained listeners’ performance on specific sound categories. The right panel of Figure 4.10 illustrates the correlation between trained listeners’ performance during training and their ability to identify office sounds. These analyses indicated that speech transcription training was significantly related to listeners’ abilities to identify office sounds \( r = .561, p = .004 \).

Figure 4.11 displays the scatterplots of the relationship between training scores and performance on the three other categories of sound. The relationship between speech transcription training and identification of kitchen sounds approached significance \( r = .333, p = .112 \). However, speech transcription training was not related to trained listeners’ abilities to identify any other category of environmental sounds.

**Perceptual confusions and hierarchical clustering of environmental sounds**

More detailed information on environmental sound identification performance of each group is shown in Tables 4.4-4.7. Tables 4.4-4.7 include stimulus-response confusion matrices of the trained and untrained listeners’ performance on each environmental sound category. Each matrix displays the probabilities of each different response alternative given a specific stimulus.
Figure 4.11. Scatterplots of relationship between training scores and identification scores for a) home b) kitchen, and c) outdoor sounds.
Table 4.4. Stimulus-response confusion matrices for outside stimulus set for untrained (top panel) and trained listeners (bottom panel). Cell contents represent the probabilities of a response given a specific stimulus.

### Untrained Listeners

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>Plane</th>
<th>Siren</th>
<th>Thunder</th>
<th>Horn</th>
<th>Birds</th>
<th>Brook</th>
<th>Keys</th>
<th>Dog</th>
<th>Car</th>
<th>Chopper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plane</td>
<td>.12</td>
<td>.12</td>
<td>.24</td>
<td>0</td>
<td>.04</td>
<td>.21</td>
<td>.02</td>
<td>0</td>
<td>.18</td>
<td>.07</td>
</tr>
<tr>
<td>Siren</td>
<td>.40</td>
<td>.19</td>
<td>.08</td>
<td>0</td>
<td>.04</td>
<td>.05</td>
<td>0</td>
<td>0</td>
<td>.06</td>
<td>.17</td>
</tr>
<tr>
<td>Thunder</td>
<td>.04</td>
<td>.01</td>
<td>.33</td>
<td>.01</td>
<td>0</td>
<td>.47</td>
<td>.03</td>
<td>.06</td>
<td>.05</td>
<td>0</td>
</tr>
<tr>
<td>Horn</td>
<td>.10</td>
<td>0</td>
<td>.15</td>
<td>.42</td>
<td>0</td>
<td>.11</td>
<td>.02</td>
<td>.15</td>
<td>.05</td>
<td>0</td>
</tr>
<tr>
<td>Birds</td>
<td>.03</td>
<td>.04</td>
<td>.04</td>
<td>.55</td>
<td>.05</td>
<td>.10</td>
<td>.03</td>
<td>.0</td>
<td>.12</td>
<td></td>
</tr>
<tr>
<td>Brook</td>
<td>.06</td>
<td>.02</td>
<td>.07</td>
<td>0</td>
<td>.07</td>
<td>.58</td>
<td>.02</td>
<td>0</td>
<td>.06</td>
<td>.12</td>
</tr>
<tr>
<td>Keys</td>
<td>.02</td>
<td>0</td>
<td>.02</td>
<td>0</td>
<td>.18</td>
<td>.05</td>
<td>.63</td>
<td>0</td>
<td>.05</td>
<td>.05</td>
</tr>
<tr>
<td>Dog</td>
<td>0</td>
<td>.02</td>
<td>0</td>
<td>.08</td>
<td>0</td>
<td>.02</td>
<td>.03</td>
<td>.76</td>
<td>.02</td>
<td>.07</td>
</tr>
<tr>
<td>Car</td>
<td>.01</td>
<td>0</td>
<td>.06</td>
<td>0</td>
<td>.10</td>
<td>0</td>
<td>0</td>
<td>.02</td>
<td>.81</td>
<td>0</td>
</tr>
<tr>
<td>Chopper</td>
<td>.01</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>.01</td>
<td>.03</td>
<td>0</td>
<td>.02</td>
<td>.93</td>
</tr>
</tbody>
</table>

### Trained Listeners

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>Plane</th>
<th>Siren</th>
<th>Thunder</th>
<th>Horn</th>
<th>Birds</th>
<th>Brook</th>
<th>Keys</th>
<th>Dog</th>
<th>Car</th>
<th>Chopper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plane</td>
<td>.25</td>
<td>.07</td>
<td>.17</td>
<td>.04</td>
<td>0</td>
<td>.15</td>
<td>.03</td>
<td>0</td>
<td>.18</td>
<td>.11</td>
</tr>
<tr>
<td>Siren</td>
<td>.35</td>
<td>.32</td>
<td>.03</td>
<td>0</td>
<td>.05</td>
<td>.10</td>
<td>0</td>
<td>.01</td>
<td>.01</td>
<td>.13</td>
</tr>
<tr>
<td>Thunder</td>
<td>.03</td>
<td>0</td>
<td>.38</td>
<td>.01</td>
<td>.01</td>
<td>.40</td>
<td>.07</td>
<td>.03</td>
<td>.07</td>
<td>0</td>
</tr>
<tr>
<td>Horn</td>
<td>.11</td>
<td>.04</td>
<td>.10</td>
<td>.40</td>
<td>0</td>
<td>.03</td>
<td>.01</td>
<td>.19</td>
<td>.10</td>
<td>.01</td>
</tr>
<tr>
<td>Birds</td>
<td>.02</td>
<td>.02</td>
<td>0</td>
<td>.03</td>
<td>.65</td>
<td>0</td>
<td>.11</td>
<td>.01</td>
<td>.05</td>
<td>.10</td>
</tr>
<tr>
<td>Brook</td>
<td>.03</td>
<td>0</td>
<td>.03</td>
<td>.03</td>
<td>.68</td>
<td>.04</td>
<td>.06</td>
<td>.04</td>
<td>.03</td>
<td></td>
</tr>
<tr>
<td>Keys</td>
<td>.01</td>
<td>.04</td>
<td>.01</td>
<td>.03</td>
<td>.10</td>
<td>.07</td>
<td>.67</td>
<td>0</td>
<td>.04</td>
<td>.03</td>
</tr>
<tr>
<td>Dog</td>
<td>0</td>
<td>0</td>
<td>.04</td>
<td>.02</td>
<td>0</td>
<td>.01</td>
<td>.01</td>
<td>.90</td>
<td>0</td>
<td>.03</td>
</tr>
<tr>
<td>Car</td>
<td>.01</td>
<td>0</td>
<td>.01</td>
<td>.03</td>
<td>.01</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>.92</td>
<td>.01</td>
</tr>
<tr>
<td>Chopper</td>
<td>.02</td>
<td>0</td>
<td>.02</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>.03</td>
<td>.93</td>
</tr>
</tbody>
</table>

Several interesting patterns of perceptual confusions were made by both the trained and untrained listeners. In the outdoor sound category, both groups of listeners identified the siren sound as plane. Thunder was also frequently confused with the babbling brook sound by both groups of listeners.

In the office sound category, both groups of listeners frequently reported hearing the water cooler sound instead of the target sounds. The three sounds frequently reported as being the water cooler were the fan, copy machine, and papers rustling. However, the frequency of confusions for two of these sound pairs differed between the trained and untrained groups. Trained listeners more frequently confused the copier and water cooler. Untrained listeners more frequently confused the rustling papers and water cooler.
Table 4.5. Stimulus-response confusion matrices for office stimulus set for untrained (top panel) and trained listeners (bottom panel). Cell contents represent the probabilities of a response given a specific stimulus.

### Untrained Listeners

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>Fan</th>
<th>Copier</th>
<th>Papers</th>
<th>Phone</th>
<th>Cooler</th>
<th>Typing</th>
<th>Crowd</th>
<th>Smoke Det.</th>
<th>Steps</th>
<th>File Door</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fan</td>
<td>.17</td>
<td>.11</td>
<td>.20</td>
<td>.11</td>
<td>.31</td>
<td>.03</td>
<td>.02</td>
<td>.03</td>
<td>.02</td>
<td>0</td>
</tr>
<tr>
<td>Copier</td>
<td>.09</td>
<td>.30</td>
<td>.27</td>
<td>.06</td>
<td>.12</td>
<td>.02</td>
<td>.04</td>
<td>.04</td>
<td>.04</td>
<td>.06</td>
</tr>
<tr>
<td>Papers</td>
<td>0</td>
<td>.04</td>
<td>.35</td>
<td>.05</td>
<td>.35</td>
<td>.09</td>
<td>.07</td>
<td>0</td>
<td>0</td>
<td>.04</td>
</tr>
<tr>
<td>Phone</td>
<td>.23</td>
<td>.12</td>
<td>0</td>
<td>.40</td>
<td>.10</td>
<td>.03</td>
<td>.04</td>
<td>.08</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Cooler</td>
<td>.17</td>
<td>0</td>
<td>.04</td>
<td>.06</td>
<td>.63</td>
<td>.06</td>
<td>.03</td>
<td>.02</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Typing</td>
<td>.25</td>
<td>.09</td>
<td>.09</td>
<td>.05</td>
<td>.04</td>
<td>.48</td>
<td>.02</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Crowd</td>
<td>.11</td>
<td>.05</td>
<td>.05</td>
<td>0</td>
<td>.12</td>
<td>0</td>
<td>.64</td>
<td>0</td>
<td>0</td>
<td>.03</td>
</tr>
<tr>
<td>SmokeDet.</td>
<td>.13</td>
<td>.12</td>
<td>0</td>
<td>.10</td>
<td>0</td>
<td>.01</td>
<td>0</td>
<td>.60</td>
<td>0</td>
<td>.04</td>
</tr>
<tr>
<td>Steps</td>
<td>.03</td>
<td>.03</td>
<td>0</td>
<td>.03</td>
<td>.06</td>
<td>.08</td>
<td>0</td>
<td>0</td>
<td>.78</td>
<td>0</td>
</tr>
<tr>
<td>File Door</td>
<td>.02</td>
<td>.05</td>
<td>.09</td>
<td>0</td>
<td>0</td>
<td>.03</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>.82</td>
</tr>
</tbody>
</table>

### Trained Listeners

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>Fan</th>
<th>Copier</th>
<th>Papers</th>
<th>Phone</th>
<th>Cooler</th>
<th>Typing</th>
<th>Crowd</th>
<th>Smoke Det.</th>
<th>Steps</th>
<th>File Door</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fan</td>
<td>.36</td>
<td>.09</td>
<td>.15</td>
<td>.04</td>
<td>.30</td>
<td>0</td>
<td>0</td>
<td>.05</td>
<td>.01</td>
<td>0</td>
</tr>
<tr>
<td>Copier</td>
<td>.11</td>
<td>.21</td>
<td>.15</td>
<td>.04</td>
<td>.32</td>
<td>.05</td>
<td>0</td>
<td>.01</td>
<td>.01</td>
<td>.07</td>
</tr>
<tr>
<td>Papers</td>
<td>.07</td>
<td>.20</td>
<td>.44</td>
<td>.01</td>
<td>.10</td>
<td>.10</td>
<td>.03</td>
<td>0</td>
<td>.04</td>
<td>.03</td>
</tr>
<tr>
<td>Phone</td>
<td>.28</td>
<td>.11</td>
<td>.04</td>
<td>.44</td>
<td>.05</td>
<td>0</td>
<td>0</td>
<td>.06</td>
<td>.02</td>
<td>0</td>
</tr>
<tr>
<td>Cooler</td>
<td>.14</td>
<td>.10</td>
<td>.05</td>
<td>.10</td>
<td>.57</td>
<td>0</td>
<td>.02</td>
<td>.02</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Typing</td>
<td>.04</td>
<td>.03</td>
<td>.10</td>
<td>.04</td>
<td>.02</td>
<td>.76</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Crowd</td>
<td>.03</td>
<td>0</td>
<td>.08</td>
<td>0</td>
<td>.17</td>
<td>0</td>
<td>.70</td>
<td>0</td>
<td>.02</td>
<td>0</td>
</tr>
<tr>
<td>SmokeDet.</td>
<td>.10</td>
<td>.05</td>
<td>0</td>
<td>.04</td>
<td>0</td>
<td>.01</td>
<td>0</td>
<td>.76</td>
<td>.03</td>
<td>0</td>
</tr>
<tr>
<td>Steps</td>
<td>.02</td>
<td>.01</td>
<td>0</td>
<td>.02</td>
<td>0</td>
<td>.10</td>
<td>.03</td>
<td>0</td>
<td>.81</td>
<td>.01</td>
</tr>
<tr>
<td>File Door</td>
<td>.01</td>
<td>.04</td>
<td>.04</td>
<td>.03</td>
<td>0</td>
<td>.03</td>
<td>.01</td>
<td>.01</td>
<td>0</td>
<td>.82</td>
</tr>
</tbody>
</table>

Several interesting patterns of confusion were also obtained in the home sounds category. Both groups of listeners identified the air conditioner most frequently as rain. Music was also primarily identified as a toilet flush. Untrained listeners frequently identified the vacuum as an air conditioner. However, trained listeners reported that the vacuum sound was rain. In both groups of listeners, the male's voice and baby crying were confused with one another.

Fewer systematic patterns of perceptual confusions were made in the kitchen category. Both groups of listeners frequently reported that they heard running water when nearly half the sounds were played. The oven timer, doorbell, cat, and dishwasher were all more commonly identified as running water by many of the listeners.
Table 4.6. Stimulus-response confusion matrices for home stimulus set for untrained (top panel) and trained listeners (bottom panel). Cell contents represent the probabilities of a response given a specific stimulus.

### Untrained Listeners

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>Air</th>
<th>Music</th>
<th>Vacuum</th>
<th>Baby</th>
<th>Voice</th>
<th>Flush</th>
<th>Knock</th>
<th>Rain</th>
<th>Door</th>
<th>Dog</th>
</tr>
</thead>
<tbody>
<tr>
<td>Air</td>
<td>.08</td>
<td>.04</td>
<td>.07</td>
<td>.01</td>
<td>0</td>
<td>.13</td>
<td>0</td>
<td>.66</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Music</td>
<td>.03</td>
<td>.22</td>
<td>.13</td>
<td>.03</td>
<td>.14</td>
<td>.30</td>
<td>.04</td>
<td>.06</td>
<td>0</td>
<td>.05</td>
</tr>
<tr>
<td>Vacuum</td>
<td>.17</td>
<td>.06</td>
<td>.31</td>
<td>.03</td>
<td>.01</td>
<td>.12</td>
<td>0</td>
<td>.30</td>
<td>.01</td>
<td>0</td>
</tr>
<tr>
<td>Baby</td>
<td>.06</td>
<td>.13</td>
<td>.24</td>
<td>.24</td>
<td>.10</td>
<td>.10</td>
<td>.04</td>
<td>.03</td>
<td>.01</td>
<td>.06</td>
</tr>
<tr>
<td>Voice</td>
<td>.12</td>
<td>.18</td>
<td>.01</td>
<td>.19</td>
<td>.37</td>
<td>.01</td>
<td>.02</td>
<td>.01</td>
<td>.04</td>
<td>.03</td>
</tr>
<tr>
<td>Flush</td>
<td>.07</td>
<td>.04</td>
<td>.04</td>
<td>.03</td>
<td>.01</td>
<td>.44</td>
<td>0</td>
<td>.26</td>
<td>.11</td>
<td>0</td>
</tr>
<tr>
<td>Knock</td>
<td>.12</td>
<td>0</td>
<td>.03</td>
<td>0</td>
<td>.01</td>
<td>.06</td>
<td>.68</td>
<td>.07</td>
<td>.02</td>
<td>.02</td>
</tr>
<tr>
<td>Rain</td>
<td>.11</td>
<td>0</td>
<td>.04</td>
<td>0</td>
<td>0</td>
<td>.08</td>
<td>0</td>
<td>.76</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Door</td>
<td>0</td>
<td>.01</td>
<td>0</td>
<td>.01</td>
<td>0</td>
<td>.02</td>
<td>.17</td>
<td>.01</td>
<td>.76</td>
<td>.01</td>
</tr>
<tr>
<td>Dog</td>
<td>0</td>
<td>0</td>
<td>.02</td>
<td>0</td>
<td>.06</td>
<td>0</td>
<td>.03</td>
<td>0</td>
<td>.02</td>
<td>.86</td>
</tr>
</tbody>
</table>

### Trained Listeners

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>Air</th>
<th>Music</th>
<th>Vacuum</th>
<th>Baby</th>
<th>Voice</th>
<th>Flush</th>
<th>Knock</th>
<th>Rain</th>
<th>Door</th>
<th>Dog</th>
</tr>
</thead>
<tbody>
<tr>
<td>Air</td>
<td>.25</td>
<td>.03</td>
<td>.07</td>
<td>0</td>
<td>0</td>
<td>.15</td>
<td>0</td>
<td>.49</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Music</td>
<td>.07</td>
<td>.16</td>
<td>.09</td>
<td>.01</td>
<td>.12</td>
<td>.30</td>
<td>.08</td>
<td>.14</td>
<td>0</td>
<td>.04</td>
</tr>
<tr>
<td>Vacuum</td>
<td>.24</td>
<td>.09</td>
<td>.27</td>
<td>.14</td>
<td>0</td>
<td>.21</td>
<td>0</td>
<td>.05</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Baby</td>
<td>.02</td>
<td>.10</td>
<td>.18</td>
<td>.31</td>
<td>.18</td>
<td>.11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>.09</td>
</tr>
<tr>
<td>Voice</td>
<td>.18</td>
<td>.10</td>
<td>.05</td>
<td>.12</td>
<td>.38</td>
<td>.07</td>
<td>.02</td>
<td>.02</td>
<td>.03</td>
<td>.04</td>
</tr>
<tr>
<td>Flush</td>
<td>.05</td>
<td>.09</td>
<td>.06</td>
<td>.09</td>
<td>.02</td>
<td>.44</td>
<td>0</td>
<td>.20</td>
<td>.05</td>
<td>0</td>
</tr>
<tr>
<td>Knock</td>
<td>.09</td>
<td>.06</td>
<td>0</td>
<td>0</td>
<td>.01</td>
<td>.03</td>
<td>.71</td>
<td>.07</td>
<td>.03</td>
<td>0</td>
</tr>
<tr>
<td>Rain</td>
<td>.04</td>
<td>.01</td>
<td>.04</td>
<td>0</td>
<td>0</td>
<td>.07</td>
<td>0</td>
<td>.82</td>
<td>0</td>
<td>.01</td>
</tr>
<tr>
<td>Door</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>.17</td>
<td>0</td>
<td>.79</td>
<td>.04</td>
<td>0</td>
</tr>
<tr>
<td>Dog</td>
<td>0</td>
<td>.01</td>
<td>0</td>
<td>0</td>
<td>.01</td>
<td>.01</td>
<td>.07</td>
<td>.01</td>
<td>.07</td>
<td>.81</td>
</tr>
</tbody>
</table>

The confusion matrices for each set of stimuli were then submitted to four separate hierarchical cluster analyses (see Everitt, 1980; Lorr, 1983). Figures 4.12-4.15 shows the dendograms of the clustering solutions obtained in each analysis. The clustering solutions were obtained by identifying the clustering patterns with the smallest within-cluster distances and largest between-cluster distances. In these dendograms, distance is shown horizontally on the x-axis. The perceptual similarity of the sounds is inversely related to horizontal distance. Vertical distance is irrelevant to the interpretation of the figures. The clustering solutions for each sound category were generally quite similar for the two groups of listeners.
Table 4.7. Stimulus-response confusion matrices for kitchen stimulus set for untrained (top panel) and trained listeners (bottom panel). Cell contents represent the probabilities of a response given a specific stimulus.

### Untrained Listeners

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>Timer</th>
<th>Doorbell</th>
<th>Cat</th>
<th>Dish Wash</th>
<th>Phone</th>
<th>H2O</th>
<th>Smoke Det</th>
<th>Dishes</th>
<th>Cupboard</th>
<th>Steps</th>
</tr>
</thead>
<tbody>
<tr>
<td>Timer</td>
<td>.07</td>
<td>.01</td>
<td>0</td>
<td>.10</td>
<td>.04</td>
<td>.71</td>
<td>.01</td>
<td>.03</td>
<td>.01</td>
<td>0</td>
</tr>
<tr>
<td>Doorbell</td>
<td>.09</td>
<td>.13</td>
<td>.06</td>
<td>.26</td>
<td>.03</td>
<td>.12</td>
<td>.05</td>
<td>.06</td>
<td>.18</td>
<td>.03</td>
</tr>
<tr>
<td>Cat</td>
<td>.03</td>
<td>.07</td>
<td>.14</td>
<td>.12</td>
<td>.05</td>
<td>.35</td>
<td>0</td>
<td>.07</td>
<td>.14</td>
<td>.03</td>
</tr>
<tr>
<td>Dish Wash</td>
<td>.03</td>
<td>.03</td>
<td>.02</td>
<td>.10</td>
<td>.04</td>
<td>.75</td>
<td>0</td>
<td>.01</td>
<td>.01</td>
<td>0</td>
</tr>
<tr>
<td>Phone</td>
<td>.15</td>
<td>.03</td>
<td>.05</td>
<td>.12</td>
<td>.26</td>
<td>.20</td>
<td>.16</td>
<td>0</td>
<td>.01</td>
<td>.01</td>
</tr>
<tr>
<td>H2O</td>
<td>.10</td>
<td>.02</td>
<td>0</td>
<td>.13</td>
<td>.04</td>
<td>.61</td>
<td>.07</td>
<td>.02</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SmokeDet.</td>
<td>.20</td>
<td>0</td>
<td>.02</td>
<td>.10</td>
<td>.07</td>
<td>.03</td>
<td>0</td>
<td>0</td>
<td>.74</td>
<td>.05</td>
</tr>
<tr>
<td>Dishes</td>
<td>.01</td>
<td>.04</td>
<td>.02</td>
<td>.07</td>
<td>.03</td>
<td>0</td>
<td>.3</td>
<td>0</td>
<td>.83</td>
<td>.03</td>
</tr>
<tr>
<td>Cupboard</td>
<td>.04</td>
<td>.01</td>
<td>0</td>
<td>.03</td>
<td>.01</td>
<td>0</td>
<td>.04</td>
<td>.83</td>
<td>.03</td>
<td>0</td>
</tr>
<tr>
<td>Steps</td>
<td>.01</td>
<td>.04</td>
<td>.03</td>
<td>.04</td>
<td>.04</td>
<td>0</td>
<td>.03</td>
<td>0</td>
<td>.81</td>
<td>0</td>
</tr>
</tbody>
</table>

### Trained Listeners

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>Timer</th>
<th>Doorbell</th>
<th>Cat</th>
<th>Dish Wash</th>
<th>Phone</th>
<th>H2O</th>
<th>Smoke Det</th>
<th>Dishes</th>
<th>Cupboard</th>
<th>Steps</th>
</tr>
</thead>
<tbody>
<tr>
<td>Timer</td>
<td>.17</td>
<td>.06</td>
<td>0</td>
<td>.11</td>
<td>.8</td>
<td>.56</td>
<td>.01</td>
<td>.01</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Doorbell</td>
<td>.07</td>
<td>.17</td>
<td>.08</td>
<td>.10</td>
<td>.10</td>
<td>.22</td>
<td>.03</td>
<td>.13</td>
<td>.10</td>
<td>0</td>
</tr>
<tr>
<td>Cat</td>
<td>.03</td>
<td>.05</td>
<td>.18</td>
<td>.16</td>
<td>0</td>
<td>.25</td>
<td>.03</td>
<td>.14</td>
<td>.13</td>
<td>.04</td>
</tr>
<tr>
<td>Dishwasher</td>
<td>0</td>
<td>.01</td>
<td>.01</td>
<td>.19</td>
<td>.01</td>
<td>.68</td>
<td>.03</td>
<td>.03</td>
<td>.01</td>
<td>.03</td>
</tr>
<tr>
<td>Telephone</td>
<td>.10</td>
<td>.03</td>
<td>.07</td>
<td>.03</td>
<td>.46</td>
<td>.20</td>
<td>.05</td>
<td>.02</td>
<td>.01</td>
<td>.04</td>
</tr>
<tr>
<td>Water</td>
<td>.06</td>
<td>.07</td>
<td>.03</td>
<td>.20</td>
<td>.06</td>
<td>.53</td>
<td>.01</td>
<td>.02</td>
<td>.01</td>
<td>0</td>
</tr>
<tr>
<td>SmokeDet.</td>
<td>.12</td>
<td>.09</td>
<td>.03</td>
<td>.05</td>
<td>.04</td>
<td>0</td>
<td>.65</td>
<td>.03</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Dishes</td>
<td>.05</td>
<td>.04</td>
<td>.01</td>
<td>.08</td>
<td>.03</td>
<td>0</td>
<td>.54</td>
<td>.10</td>
<td>.76</td>
<td>0</td>
</tr>
<tr>
<td>Cupboard</td>
<td>.06</td>
<td>.03</td>
<td>.02</td>
<td>.02</td>
<td>.02</td>
<td>0</td>
<td>.10</td>
<td>.76</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Steps</td>
<td>0</td>
<td>0</td>
<td>.02</td>
<td>.04</td>
<td>.06</td>
<td>0</td>
<td>.02</td>
<td>0</td>
<td>.86</td>
<td>0</td>
</tr>
</tbody>
</table>

Figure 4.12 illustrates the clusters of untrained and trained listeners' responses to outdoor sounds. The results of this cluster analysis were similar for both groups. Both trained and untrained listeners appeared to cluster the ten outdoor sounds into five clusters. For both groups of listeners, the largest cluster was comprised of five sounds: *airplane, thunder, brook, siren, horn*. In each group, keys and bird sounds also formed a cluster. Finally, individual clusters were obtained for three sounds: *barking dog, car, helicopter*.

Figure 4.13 illustrates the clusters of untrained and trained listeners' responses to office sounds. Untrained listeners clustered these sounds into five groups, while trained listeners formed six distinct groups. The largest cluster for untrained listeners included the sounds: *fan, papers rustling, photocopier, water cooler, phone*, and *typing*. The largest cluster of the trained listeners included all the sounds of the untrained listeners' cluster except for *typing*. Trained listeners clustered typing sounds independently.
Both groups of listeners also clustered four other sounds (*crowd talking, smoke detector, file door slamming, footsteps*) independently.

Untrained Listeners

![Rescaled Distance Measure for Untrained Listeners]

Trained Listeners

![Rescaled Distance Measure for Trained Listeners]

**Figure 4.12.** Dendograms of the hierarchical-clustering analysis of the outdoor sounds for the untrained (top panel) and trained (bottom panel) listeners.
Figure 4.13. Dendograms of the hierarchical-clustering analysis of the office sounds for the untrained (top panel) and trained (bottom panel) listeners.

Figure 4.14 illustrates the clusters of untrained and trained listeners’ responses to home sounds. The results of this cluster analysis were similar for both groups. Both trained and untrained listeners appeared to cluster the ten outdoor sounds into five clusters. For both groups of listeners, the largest
cluster was comprised of five sounds: baby crying, male voice, music, toilet flush, vacuum. In each group, air conditioning and rain sounds also formed a cluster. Finally, individual clusters were obtained for three sounds: knock on door, door closing, dog barking.

**Untrained Listeners**

![Rescaled Distance Measure Chart for Untrained Listeners]

**Trained Listeners**

![Rescaled Distance Measure Chart for Trained Listeners]

*Figure 4.14. Dendograms of the hierarchical-clustering analysis of the home sounds for the untrained (top panel) and trained (bottom panel) listeners.*
Figure 4.15 illustrates the clusters of untrained and trained listeners' responses to kitchen sounds. The results of this cluster analysis were similar for both groups. Both trained and untrained listeners appeared to cluster the ten outdoor sounds into six clusters. Although the number of clusters formed by each group was equal, the composition of several of the clusters differed slightly. Untrained listeners had two clusters each containing three sounds. One cluster included the sounds: *oven timer, dishwasher, running water*. A second cluster formed by untrained listeners included the sounds: *cat meowing, doorbell, phone*. Untrained listeners clustered *smoke detector, dishes clanging, cupboard closing, and footsteps* independently.

Untrained Listeners

![Dendrogram for untrained listeners]

Trained Listeners

![Dendrogram for trained listeners]

Figure 4.15. Dendograms of the hierarchical-clustering analysis of the kitchen sounds for the untrained (top panel) and trained (bottom panel) listeners.
The largest cluster of kitchen sounds formed by trained listeners included the sounds: *dishwasher, running water, cat meowing*, and *doorbell*. *Oven timer* and *phone* were also clustered together. Like the untrained listeners, trained listeners also clustered *smoke detector, dishes clanging, cupboard closing*, and *footsteps* independently.

**Talker-gender identification.** Figure 4.4 displays group means for talker-gender identification scores. An independent-samples T-test revealed no significant difference between trained and untrained listeners' abilities' to identify the gender of talkers ($t(46) = .47, p = .64$). Both the trained ($t(23) = 12.60, p < .001$) and untrained ($t(23) = 10.94, p < .001$) listeners performed significantly better than chance on the talker-gender identification task but no differences were observed between the two groups.

Figure 4.16 displays group means for the talker-gender identification task separately based on the gender of the talkers. A repeated-measures ANOVA was conducted to determine the effect of talker-gender on identification accuracy. A main effect of talker-gender was found ($F(1, 46) = 18.64, p < .001$). Gender-identification scores were higher for male voices ($M = 81.35, SD = 14.87$) than for female voices ($M = 69.79, SD = 14.52$). The interaction between talker-gender and group assignment was not significant ($F(1, 46) = .23, p = .637$). This result suggests that listeners may have simply had a response bias towards male voices or that some aspect of the signal processing was more favorable for the identification of male voices. One possibility is that there was more availability of low frequency information in the male voices due to the signal processing used.

![Figure 4.16](image)

**Figure 4.16.** Performance on speaker-gender identification task by trained and untrained listeners according to the gender of the speaker. Error bars represent the standard error of the mean.

**Talker discrimination.** Figure 4.4 displays group means for the talker discrimination task. An independent-samples T-test indicated that there was no significant difference between trained and untrained listeners' abilities to discriminate between talkers ($t(46) = .52, p = .60$). Both the trained ($t(23) = 6.22, p < .001$) and untrained ($t(23) = 5.49, p < .001$) listeners performed significantly better than chance on the talker discrimination task.

**Correlations among transfer tasks**

To determine if there were any notable relationships between the untrained and trained listeners' performance on the three transfer tasks, simple bivariate correlations were conducted. The Bonferroni method of adjustment was used to obtain the significance level appropriate for the number of correlations run. The adjusted significance level used for these bivariate correlations was .01. Figures 4.17-4.19
show the scatterplots of the relationships between the three transfer tasks in both the untrained and trained listeners.

A positive correlation was found between gender and environmental sound identification in the untrained listeners ($r = .632$, $p = .001$). The relationship between environmental sound identification and talker discrimination approached significance in the untrained listeners ($r = .441$, $p = .031$). The correlation between gender and talker discrimination also approached significance in these listeners ($r = .505$, $p = .012$).

![Scatterplot of environmental sound identification and gender identification in untrained listeners](image1)

![Scatterplot of environmental sound identification and gender identification in trained listeners](image2)

**Figure 4.17.** Scatterplots of the relationships between environmental sound identification and gender identification in a) untrained listeners and b) trained listeners.

In trained listeners, weaker intercorrelations were observed among the transfer tasks. The relationship between gender and talker discrimination approached significance in the trained listeners ($r = .385$, $p = .064$). However, neither gender or talker discrimination were related to environmental sound identification in the trained listeners.

![Scatterplot of environmental sound identification and speaker discrimination in untrained listeners](image3)

![Scatterplot of environmental sound identification and speaker discrimination in trained listeners](image4)

**Figure 4.18.** Scatterplots of the relationships between environmental sound identification and speaker discrimination in a) untrained listeners and b) trained listeners.
Figure 4.19. Scatterplots of the relationships between gender identification and speaker discrimination in a) untrained listeners and b) trained listeners.

Discussion

Several novel findings on auditory perceptual learning and the perception of speech and nonspeech stimuli were uncovered in the present study. First, on average, normal-hearing listeners were fairly accurate in identifying spectrally degraded environmental sounds whether or not they were trained to identify spectrally degraded speech. However, sounds with distinct transient or temporal properties appeared to be the easiest for listeners to identify. These results are consistent with the earlier findings using adult cochlear implant patients (Reed & Delhorne, 2005) and normal-hearing listeners trained to identify spectrally degraded environmental sounds (Gygi et al., 2004). Taken together, the results of these studies show that, under degraded listening conditions, both normal-hearing and hearing-impaired listeners rely on temporal information to accurately identify environmental sounds.

However, for tasks in which temporal information alone cannot be reliably used to attain the correct result, listeners’ performances were much more variable. Listeners’ performances on a talker-gender identification and talker discrimination task in spectrally degraded conditions were much worse than would have been expected if the auditory conditions were normal. Under clear listening conditions, normal-hearing listeners can identify talker-gender with very little context or acoustic information. For instance, studies have shown that listeners can identify talker-gender based only on isolated words, vowels, or fricatives (Lass, Hughes, Bowyer, Waters, & Bourne, 1976; Perry, Ohde, & Ashmead, 2001; Schwartz, 1968). However, in the present study, even when entire sentences were played to the listeners, they achieved less than 80% accuracy on a talker-gender identification task.

Similarly, the listeners’ performance on a talker discrimination task was much worse than would be expected if the auditory conditions were clear. Although both groups of listeners performed significantly better than chance, their overall performance was less than 60%. However, the difficulty of completing the talker discrimination task may have been related to the experimental procedures used in this study.
Studies on talker discrimination in cochlear implant users have indicated that it is much more difficult for listeners to accurately discriminate talkers when the linguistic information in each talker’s utterance is varied instead of held constant (Cleary, 2003; Cleary & Pisoni, 2002; McDonald et al., 2003). However, in the present study, listeners were asked to discriminate between talkers who spoke sentences with different linguistic content. Although performance would have conceivably been better if the same sentences were used in training pairs, this method was chosen because it is presumed to be more clinically applicable and externally valid than presenting identical sentences to listeners in a talker discrimination task.

Although identifying the gender of a talker or discriminating between talkers is a relatively easy task in clear auditory conditions, when the speech signal is spectrally degraded, it is harder to identify talker-gender and to discriminate between talkers. It may be harder to discriminate the gender and identity of talkers while listening to an acoustic simulation of a cochlear implant, because some of the signature cues of talker-gender and identity are disrupted in spectrally degraded speech. Determining the gender of a talker’s voice relies on several acoustic factors. The most important factors include fundamental frequency, formant frequencies, and breathiness of the voice (Klatt & Klatt, 1990).

Discriminating between two different voices also relies on specific cues. Cues that may be used to discriminate talkers include formant range, pitch variability, intensity, and other spectral qualities (Remez, Fellowes, & Rubin, 1997). With spectrally degraded speech, the cues important to discriminating talker-gender and identity are not only attenuated but also distorted. Thus, it may be difficult to accurately use these cues when trying to identify talker-gender and discriminate between talkers when listening to spectrally degraded speech. In addition, the results of this study suggest that it is also difficult to learn about cues for gender and talker identity in a short period of sentence transcription training. The trained listeners in this study performed no better on the gender and talker discrimination task than listeners who received no exposure to and training with the acoustic simulation of the cochlear implant.

The results of this study indicate that the ability to transfer learning from a speech transcription training task to other speech and non-speech tasks may be somewhat limited in normal-hearing listeners trained with an acoustic simulation of a cochlear implant. Transfer of learning to a talker-gender identification task and a talker discrimination task was not observed in this study. Listeners who received a short period of speech-transcription training performed no better on the transfer tasks than listeners who received no training or exposure to the spectrally degraded speech. Transfer of learning may have failed to occur because the signal processing, training, and feedback that the trained individuals received did not provide them with reliable information needed to easily identify the gender of a talker or to discriminate between talkers.

Based on previous studies that have shown generalization of learning with spectrally degraded speech to small frequency shifts (Burkholder, 2005; Fu & Galvin, 2003; Fu & Shannon, 1999a), different simulation signal carriers (Davis, Taylor, Johnsruude, & Carlyon, 2004), and different analysis filters (Davis, Hervais-Adelman, Taylor, Carlyon, & Johnsruude, 2005), it appears that listeners learn primarily from the gross spectro-temporal information rather than from encoding the fine structure of the speech signal when being trained with acoustic simulations of cochlear implants. However, talker and gender identification rely heavily on the resolution and encoding of fine acoustic cues such as frequency and rapid formant movements rather than temporal information. Thus, whatever attributes listeners did learn during the short training period in the current study could not be transferred to the talker and gender discrimination tasks.
In addition, the results of this study suggest that both the type of learning and the method in which it was acquired were not transfer appropriate to these different speech tasks. Having transfer-appropriate learning could be one crucial factor determining whether or not training effects from one task will carry over to another task (Benson, Lovett, & Kroeber, 1997; Franks, Bilbrey, Lien, McNamara, 2000; Hunstad, 1985; Lockhart, 2002; Lovett, Borden, DeLuca, Lacerenza, Benson, & Brackstone, 1994; Morris, Bransford, & Franks, 1977; Roediger, Buckner, & McDermott, 1999). The theory of transfer-appropriate processing predicts that learning will only transfer to a new task if encoding and processing that occurred during initial learning in the training task is the same as the encoding and processing required in the new task.

Speech transcription training may not have transferred to gender and talker discrimination tasks because both the encoding and processing operations used in the tasks were different. During sentence transcription training with speech processed through an acoustic simulation of a cochlear implant, listeners were required to encode the spectro-temporal and linguistic information in the signal to recognize the words. However, to discriminate between talker-gender and talkers, encoding of frequency, formant, and harmonic information is required. Thus, the underlying processes used in a sentence transcription task and any discrimination task are also fundamentally different. Sentence transcription requires that responses be generated from an infinite open-set. However, identifying a voice as male or female or discriminating between two talkers is a closed-set, two-alternative forced-choice task. These differences may have made it difficult to transfer learning from a sentence transcription task to talker-gender and talker discrimination tasks.

Similarly, assessing transfer of learning to the environmental sound identification task may have been limited because the encoding and underlying processes used in the training and test task differed. Unlike talker-gender and talker discrimination, tests of environmental sound identification are not necessarily restricted to a closed-set formant. However, the listeners in the current study would likely have performed at floor if environmental sound identification had been measured through an open-set. To make the two tasks more comparable, speech perception training could have been conducted using a closed-set format in order to preserve similarity between the training and test tasks. The mismatched test formats of the training and transfer tasks may limit the evaluation of transfer of learning in the current study.

Despite the fact that speech transcription training and environmental sound identification required different processing and response strategies, some limited evidence of transfer of learning from speech training to sound identification was observed. Environmental sound identification was significantly easier for listeners who received speech transcription training than for listeners who received no training or exposure to the stimuli. Evidence of transfer of learning to an environmental sound identification task was observed in the trained listeners. Trained listeners performed significantly better overall on environmental sound identification than listeners who received no training or exposure with the spectrally degraded speech. However, transfer of learning to environmental sound identification did not appear to be particularly robust across all of the different environmental sound categories.

The differences observed between the trained and untrained listeners were specific to only one of the four categories of sounds. Trained listeners only performed significantly better than untrained listeners on outdoor sounds. This is an interesting finding, because the outdoor sound category differed slightly from the other three sound categories. The outdoor sound category had the least amount of transient or temporally distinct sounds. In contrast, nearly all the sounds in the outdoor sound category lacked distinct temporal qualities.
In the other three sound categories, the untrained listeners identified most transient sounds with a rate of accuracy similar to the trained listeners. Thus, for most transient or temporally distinct sounds, the trained and untrained listeners did not differ. However, evidence that the trained listeners performed significantly better on the sound category comprised of sounds lacking distinct temporal-envelope characteristics suggests that they were able to make finer judgments about sounds based on gross spectral cues. This result is similar to findings observed in adult cochlear implant users. Reed and Delhorne (2005) found that, while high and low performing cochlear implant users performed similarly on transient sounds, high performing users were much better at identifying sounds without distinct temporal cues than low performing users.

The trained listeners’ abilities to make finer judgments based on gross spectral cues were probably the result of the speech transcription training that they received. Previous research has suggested that listeners who receive speech transcription training learn primarily about the overall spectro-temporal qualities of the degraded speech (Burkholder, 2005; Davis et al., 2005). The results of the present transfer of training study support these previous findings and also suggest that the learning acquired in speech transcription training may transfer to an environmental sound identification task if the sounds are characterized by fine spectral details rather than simple temporal cues.

The current evidence of transfer of learning from a speech task to environmental sound identification is both theoretically and clinically interesting. Theoretically, transfer of learning is an important finding, because it suggests that the learning observed in the trained listeners was not necessarily specific to speech. Rather, learning in the speech task was partially extendable to a nonspeech task.

The current finding suggests that the auditory perceptual learning that occurs during a speech transcription task is not exclusively limited to speech or other auditory stimuli that may be used for training. Instead, auditory perceptual learning of speech appears to be a more domain-general process capable of generating knowledge that can be used to modify more than one auditory capability. This result complements previous findings suggesting that speech processing does not simply involve the analysis of phonetic properties or phonetic information but also involves the encoding of talker-specific information and the identification of talkers (Nygaard et al., 1994; Nygaard & Pisoni, 1998; Remez et al., 1997). Taken together, it appears that both auditory perceptual learning of speech and speech perception are multifaceted processes that may impact other important auditory skills.

The results of this study are also clinically important. Although being able to identify speech is the primary goal for most cochlear implant users, being able to identify everyday sounds is undoubtedly also important for these listeners. This study indicates that both speech and sound identification may be achieved through speech transcription training. Thus, specific training on sounds in the environment may not be necessary if cochlear implant users are first able to learn to identify speech. Overall, the results of this study suggest that the primary focus of rehabilitation in cochlear implant users should be on speech.

However, in the current study, speech transcription training did not transfer to talker-gender identification and talker discrimination. A future direction of this research should include investigations of how these skills can be improved in cochlear implant users. It is currently unclear whether task-specific training would be required to improve gender and talker discrimination or if training on other skills such as frequency or amplitude modulation discrimination would enhance listeners’ abilities to identify the gender of talkers and discriminate between talkers. Future studies on this topic could be valuable to cochlear implant users, because having the ability to correctly determine the gender and identity of talkers is an important part of effective spoken communication.
Being able to encode gender and talker-specific attributes while listening to speech is important because this information facilitates not only speech perception (Nygaard et al., 1994; Nygaard & Pisoni, 1998) and spoken word recognition (Goldinger, 1992) but other related cognitive processes as well. Lightfoot (1989) found that familiarity with talkers' voices can facilitate the serial recall of spoken words. Similarly, Goldinger (1996) found that implicit recognition memory of spoken words was influenced by talker repetition. Taken together, the results of these studies indicate that talker-specific information is encoded and used by listeners and not just a source of "noise" that needs to be normalized to accurately perceive speech (Pisoni, 1993; Pisoni, 1996). Thus, encoding and interpreting speaker-specific information in speech is a cognitive process separate from encoding the linguistic content of the message.

The neural basis of both linguistic and voice encoding processes have recently been explored. The results of several studies have suggested that two different processing streams originate in the auditory cortex and are functionally analogous to processing streams in the visual system. For more than 20 years, visual scientists have known that two different processing streams extend dorsally and ventrally from visual cortex to prefrontal cortex. The function of these pathways is to indicate where and what a visual percept is respectively (Ungerleider & Mishkin, 1982).

Similar ventral and dorsal processing streams have been hypothesized to exist in the auditory system in order to help listeners identify "where" a sound is coming from and "what" it is (Belin & Zatorre, 2000; Crinion, Lambon-Ralph, Warburton, Howard, & Wise, 2003; Romanski, 1999; Scott & Johnsrude, 2003; Scott, Blank, Rosen, & Wise, 2000). In addition, substreams of these primary processing pathways have also been suggested. One of the most notable of these substreams, associated with the "what" pathway, is suggested to process voice-specific information and identify speakers or to function as a "who" pathway (Belin & Zatorre, 2000). The identification of this neural pathway supports the numerous pieces of behavioral evidence suggesting that speaker identification is an independent and important process in speech perception that should be further examined in both normal-hearing listeners and cochlear implant patients.

Conclusions

In summary, the results of the present study suggest that training normal-hearing listeners to identify speech processed through an acoustic simulation of a cochlear implant transfers only to a limited extent to the perception of nonspeech environmental sounds using a closed-set, forced-choice identification task. The benefits of this transfer of learning were observed only for sounds lacking distinct temporal cues. This result suggests that the trained listeners learned to make fine distinctions of gross spectral cues during speech perception training and that this learning transferred to a limited extent to an environmental sound identification task. In contrast, the trained listeners' learning was not appropriate or applicable to speaker-gender identification and speaker discrimination tasks. Perceptual learning with spectrally degraded speech may not transfer to speaker-gender identification and speaker discrimination because these tasks rely extensively on early sensory encoding of frequency and formant information in speech rather than on the gross spectro-temporal cues that remained relatively intact after signal degradation.

Overall, the results of this study suggest that some aspects of auditory perceptual learning of spectrally degraded speech may not be specific to speech processing. Thus, speech perception training may be characterized by more domain-general auditory processing and perceptual learning rather than speech-specific processes. Clinically, the findings of the present study suggest that speech-based training programs for cochlear implant users may transfer in limited ways to environmental sound identification. However, improving cochlear implant users' abilities to discriminate gender and speakers may be a much
more difficult task requiring explicit gender and speaker discrimination training or a more diverse rehabilitative program that includes psychophysical training in addition to speech perception training and task-specific training. Future studies should be conducted in this clinical population to determine what training methods could be used to improve speaker-gender and speaker discrimination skills. In addition, new research and development efforts should explore how signal processing methods could be modified to provide cochlear implant users with more detailed frequency information to use in discriminating and categorizing talker-gender and identifying talkers.
CHAPTER V: SUMMARY AND CONCLUDING REMARKS

The current project examined several important aspects of auditory perceptual learning in normal-hearing adults exposed to an acoustic simulation of a cochlear implant. Three separate experiments were conducted to determine what type of training, feedback, and training stimuli facilitate perceptual learning of highly degraded speech processed by an acoustic simulation of a cochlear implant. In addition, the current set of experiments examined generalization and transfer of auditory perceptual learning to new stimuli and auditory tasks. Taken together, these studies should be of theoretical interest to researchers studying speech perception, auditory perceptual learning, and the development of other auditory skills. The results of the current project should also be clinically important to cochlear implant users and the researchers and clinicians who develop and utilize rehabilitative strategies with these individuals.

Experiment I was designed primarily to investigate the role of feedback on auditory perceptual learning and to determine which form of feedback was most successful in helping the listeners learn to identify sentences processed through an 8-channel acoustic simulation of a cochlear implant. The types of feedback used in this study were clear auditory feedback and written feedback presented simultaneously with the spectrally degraded stimulus. In addition, control listeners who received either no feedback or exposure to the spectrally degraded speech were also included in this study.

The primary finding concerning the effects of feedback in this first study was that listeners who received feedback that paired the written text of a sentence with the spectrally degraded token learned more than listeners who received clear auditory feedback, no feedback, or no exposure to the spectrally degraded speech. Finding that this form of feedback produced the most learning was important because this type of feedback may be the most appropriate to use with cochlear implant users. Unlike the normal-hearing listeners in this study, cochlear implant users are unable to hear training tokens in an unprocessed form because of their hearing loss.

Several explanations for why feedback that pairs degraded speech with written text was the most successful in this study were considered. First, it has been suggested that knowing the identity of a training token while rehearsing it in its spectrally degraded form induces a “pop out” effect (Davis et al., 2005). “Pop out” is presumably effective because it allows listeners to learn to match their ordinary auditory representations of speech to the spectrally degraded signal that they are hearing more efficiently and accurately. In Experiment I, feedback that simultaneously paired auditory and written forms of training tokens was effective because it allowed listeners to match their auditory representations with the degraded speech signal online with little load on memory.

An additional explanation for why the listeners trained with degraded auditory and written feedback performed the best is based on the transfer-appropriate processing theory of learning and memory (TAP; Lockhart, 2002; Morris et al., 1977; Neath & Surprenant, 2003; Rajaram et al., 1998; Roediger et al., 1999). TAP suggests that training and feedback will be most effective when they are appropriate for the task and stimuli being learned. In a perceptual learning task using degraded speech, it may be most appropriate to provide feedback that allows the listeners to match their preexisting auditory representations to the degraded acoustic-phonetic accurately and easily.

Feedback that simultaneously provides the degraded auditory signal and written text of a sentence may also produce gains in learning because it reinstates the conditions of the listeners’ original learning. While listening to each spectrally degraded speech token for the first time, prior to even hearing feedback, listeners are undoubtedly learning something new about how the acoustic simulation of the
cochlear implant transforms speech. By providing a form of feedback that pairs the written identity of the degraded sentences with the original spectrally degraded signal, the original conditions of learning are reinstated. The reinstatement of auditory perceptual learning during the degraded auditory and written feedback phase may be a critical factor in demonstrating the success of this form of feedback.

In addition to determining what type of feedback resulted in the most auditory perceptual learning in normal-hearing adults exposed to an acoustic simulation of a cochlear implant, Experiment I also investigated the role of sentence context and top-down influences on perceptual learning. Several listener groups were trained with semantically anomalous sentences instead of meaningful sentences. Not surprisingly, transcription of semantically anomalous sentences during the training period was more difficult than transcription of meaningful sentences. However, post-test scores indicated that listeners trained with anomalous sentences nearly performed significantly better after training than listeners trained with meaningful sentences. Results of a linear regression analysis suggested that the training materials predicted post-test performance. According to the regression analysis, training using semantically anomalous sentences predicted better post-test scores than training using meaningful sentences.

Finding that listeners trained on anomalous sentences performed better after training than listeners trained on meaningful sentences suggests that removing contextual support from the training materials may force listeners to rely more on bottom-up processing to identify speech. When ordinary semantic support is stripped from the training stimuli, listeners may become more attuned to the fine acoustic-phonetic detail and variability of the degraded auditory signal than to the linguistic content embedded in the signal. The current findings suggest that cochlear implant patients may benefit from speech transcription training that uses anomalous sentence or nonword stimuli.

In Experiments II and III, the robustness of auditory perceptual learning with an acoustic simulation of a cochlear implant was explored in greater depth. Based on the results of the first experiment, the listeners in these subsequent studies were trained using feedback that paired the degraded auditory sentences and the written text of the sentences. In Experiment II, the robustness of learning was evaluated by testing generalization of learning or adaptation to novel frequency shifts. This is an important topic to explore in cochlear implant research because post-lingually deafened adult cochlear implant users frequently experience a frequency shift due to a baselward mismatch of the electrodes and the characteristic frequency of the neurons that these electrodes stimulate.

Several new findings about auditory perceptual learning and generalization of learning were uncovered in Experiment II. First, the quality and quantity of auditory perceptual learning with an acoustic simulation of a cochlear implant is influenced by the amount of spectral shift listeners are exposed to during training. Listeners trained with a large 6.5-mm frequency shift learned significantly less during training than listeners trained with no shift or a small 3-mm shift. It is possible that minimal learning effects were observed in this group of listeners because it was too difficult to encode speech processed with a large spectral mismatch even when feedback was available. In contrast, encoding and learning to identify speech with minimal spectral mismatch did not appear to be a problem for most of the listeners in this study. This result is not surprising given that a 3-mm frequency may only result in an audible change that is comparable to a shift in the gender or age of a talker (Fu & Shannon, 1999).

A more interesting pattern of results was found when we examined listeners' abilities to generalize auditory perceptual learning to novel frequency shifts. Results of the spectral shift study suggest that listeners are able to generalize auditory perceptual learning to either small upward or small downward shifts. This result suggests that listeners may be learning more about the general and abstract
properties of the spectrally degraded speech rather than about the specific fine-structure associated with the frequency shift on which they were trained.

However, a limit on listeners' abilities to generalize learning to novel frequency shifts was observed. Listeners were unable to generalize learning to large 6.5-mm frequency shifts even if they were trained on a small shift. This result suggests that there are fundamental limitations to the auditory system's ability to accommodate large frequency mismatches after a limited amount of training in laboratory experiments using acoustic simulations of cochlear implants.

Although some generalization was observed in this study, evidence of encoding specificity was also observed. Listeners who were trained with the 3-mm frequency shift performed significantly better on the 3-mm shift condition than the two novel shift conditions during generalization. This result indicates encoding specificity for the trained stimulus. However, the listeners trained with a 3-mm frequency shift did not perform any worse during the no-shift condition than during training. According to the operational definition of generalization in this study, this result suggests that listeners trained with a 3-mm frequency shift are able to generalize learning to spectrally degraded speech that has no frequency shift. Taken together, the results observed in this group of listeners suggest that specificity and generalization of auditory perceptual learning are not necessarily mutually exclusive.

However, the data of listeners trained with a 3-mm frequency shift were unique in this spectral shift study. Encoding specificity was not observed in the listeners trained with a 6.5-mm shift or in the listeners trained with no frequency shift. To investigate this asymmetry in encoding specificity further, a replication was conducted. Results of the replication confirmed that listeners trained with a 6.5-mm frequency shift or with no frequency shift failed to show encoding specificity to the signal processing on which they were trained. It is not surprising that the listeners trained with the 6.5-mm frequency shift failed to demonstrate encoding specificity because they exhibited very little learning in the training phase. However, the absence of encoding specificity in listeners trained with no frequency shift was unexpected.

A final result in the spectral shift study was the pattern of generalization demonstrated by the listeners trained with the 6.5-mm frequency shift. Listeners trained with a 6.5-mm frequency shift demonstrated more robust generalization to the 3-mm frequency shift than to the no-shift condition. This result suggests that, even though these listeners did not exhibit encoding specificity to the 6.5-mm frequency shift, they may have learned something about spectral mismatch in general. This result could also simply be a demonstration that listeners are more capable of generalizing auditory perceptual learning to stimuli that have the smallest perceptual distance or difference from the training stimuli than to stimuli that are just perceptually easier.

Taken together, although some of the current findings from the spectral shift study may not be accommodated by one consistent theoretical explanation, they may have important clinical implications for cochlear implant users. One of the primary purposes of the spectral shift study was to determine if training with spectrally degraded but matched speech would generalize to speech with a large basalward frequency shift. This question was explored in order to determine if there was a more efficient method of training normal-hearing listeners to adapt to large frequency shifts than has previously been determined (Rosen et al., 1999; Svirsky et al., 2003).

It is clinically important to determine how listeners adapt to frequency shifts because cochlear implant users may be exposed to them as a result of shallow electrode placement on the surface of the cochlea. In addition, cochlear implant users have been found to have difficulty adapting to large
frequency shifts (Fu et al., 2002). Failure of generalization to a large basalward shift after training with no shift indicates that cochlear implant users may derive little benefit from brief exposure to spectrally matched speech prior to being remapped with a frequency shift. Thus, gradual adaptation (Svirsky et al., 2003) or prolonged and interactive continuous discourse training (Rosen et al., 1999) programs are likely to be better methods of training cochlear implant users to understand speech with a large spectral mismatch.

The third and final study in the current investigation examined the ability of normal-hearing listeners to transfer perceptual learning from a speech transcription task to environmental sound and speaker-gender identification and to speaker discrimination. Transfer of training is an important part of auditory perceptual learning to examine because it may be one of the best indicators of how perceptually robust the learning process is. Examining transfer of training between speech and nonspeech tasks is one way to determine the robustness or specificity of auditory perceptual learning. In addition, examining transfer of learning between speech and nonspeech tasks may have important clinical implications for cochlear implant users.

The results of Experiment III indicated that normal-hearing listeners may be able to transfer what they learned in a speech transcription task to the identification of environmental sounds. Transfer of training occurred because trained listeners performed better on the environmental sound identification task than untrained listeners. This is a significant finding because it suggests that auditory perceptual learning using an acoustic simulation of a cochlear implant may not necessarily be specific to speech.

However, transfer of learning to the environmental sound identification task was not robust based on the group differences observed. The trained listeners only performed significantly better than the untrained listeners on one category of environmental sounds. This observed pattern of transfer in the trained listeners revealed details about how and what they learned during training. The category of sounds that the trained listeners performed better on was different from all other sound categories. Very few of the sounds in this category had distinct temporal or transient qualities which made many sounds of the other categories easily identifiable to both the trained and untrained listeners.

Most sounds in the outdoor sound category were only identifiable when finer distinctions among gross spectral cues could be made by listeners. Thus, it appears that the trained listeners’ environmental sound identification abilities benefited from specific speech transcription training because they learned something about the changes in the fine spectral detail of the acoustic signal. In addition, the trained listeners’ perceptual learning may have showed some evidence of transfer because what they learned about the degraded auditory signal was not necessarily specific to speech or processes used for speech intelligibility.

Several additional findings from Experiment III revealed limits on listeners’ abilities to transfer auditory perceptual learning. Speech transcription training did not transfer to speaker-gender identification or speaker discrimination tasks. These two tasks may not have been transfer-appropriate because what the listeners learned during the initial training task which emphasized word recognition was insufficient to support the discrimination of speakers’ indexical properties. Discriminating between genders and speakers when the auditory signal is spectrally degraded is a difficult task because these skills rely more on frequency and formant pattern information than on gross spectro-temporal cues. Like the current generation of cochlear implant processing strategies, the signal processing used in the current simulation did not adequately preserve and encode frequency and formant information important in speech. Thus, during speech transcription training listeners did not have the opportunity to learn about the acoustic cues that are important to determining speaker-gender and discriminating between speakers.
The results of the third study are of some clinical interest because they suggest that cochlear implant patients may not need any specific training to identify sounds in their environment if they receive speech perception training or if they are already good at perceiving speech. However, the inability of listeners to transfer learning to gender and speaker discrimination tasks suggests that these skills may take more time and effort for cochlear implant users to develop after receiving their implant. Alternatively, cochlear implant users may never fully learn to identify speaker-gender and to discriminate between speakers if the acoustic cues needed for these tasks are not adequately coded and transmitted to them through their speech processors. Future studies should explore more thoroughly the types of training regimens that could successfully improve cochlear implant users’ gender and speaker discrimination skills. In addition, current research efforts aimed at improving speech processing systems should be expanded to help determine how speaker-gender and speaker-specific characteristics could be more accurately represented through electrical stimulation.

Taken together, the current set of studies provides an initial picture of the auditory perceptual learning process in normal-hearing listeners trained to identify speech processed through an acoustic simulation of a cochlear implant. Overall, perceptual learning occurs rapidly in a short period of time and generalizes to new stimulus categories and some new forms of signal processing. In addition, perceptual learning with an acoustic simulation of a cochlear implant may transfer to nonspeech environmental sounds in closed-set identification tasks. These results have theoretical implications for how normal-hearing listeners learn to identify speech under highly degraded and unfamiliar listening conditions.

In addition, this series of perceptual learning studies was also designed to explore how cochlear implant users learn to understand speech through their devices and how rehabilitative training programs could be developed to best suit the needs of this clinical population. In the current set of studies, acoustic simulations of a cochlear implant were used in normal-hearing listeners to gain perspective into these issues. Acoustic models of cochlear implants have recently emerged as a valuable tool to simulate the auditory experience of cochlear implant patients in normal-hearing listeners and to study a variety of basic speech perception and auditory skills. However, there are several caveats involved in using this method of research and drawing any general conclusions about actual cochlear implant users based on its findings.

First, acoustic simulations of cochlear implants are only approximate models of what real patients may hear through their devices. In addition, acoustic models of cochlear implants often do not even sound the same to the normal-hearing listeners that are used across studies in this field of research. A variety of different signal processing methods are currently being used in cochlear implant simulation studies. These differences make it difficult to draw any firm conclusions about how both normal-hearing listeners and cochlear implant patients learn to identify spectrally degraded speech. In order for acoustic simulation studies to give a more accurate and consistent picture of how normal-hearing listeners’ speech perception and perceptual learning skills are affected by spectrally degraded speech, it may be useful to determine if there is one “best” acoustic model of cochlear implants or if the method of modeling used should vary based on the listening conditions and research question under investigation.

The amount of training and exposure that normal-hearing listeners receive in acoustic simulations studies may also limit their applicability to actual cochlear implant patients. Using naïve normal-hearing participants as listeners in acoustic simulation studies may be an appropriate way to draw inferences about how cochlear implant patients will perceive speech immediately after activation. However, many cochlear implant patients still struggle to understand speech long after activation and an extended period of auditory exposure and use under real-world listening conditions. Thus, acoustic
simulation studies using naïve normal-hearing listeners may not provide reliable insights about which training and rehabilitative methods should be used with experienced cochlear implant patients who still have problems understanding speech.

Finally, normal-hearing listeners exposed to acoustic simulations of cochlear implants may not provide an accurate representation of cochlear implant users implanted after a period of prolonged deafness. Important differences exist between normal-hearing listeners exposed to acoustic simulations of cochlear implants and real cochlear implant users at both peripheral and cortical levels. For instance, normal-hearing listeners have intact cochlear mechanics that operate nonlinearly in order to compress sound (Moore, 1997). However, the nonlinearity of an intact cochlea cannot be accurately modeled using current signal processing strategies in cochlear implants (Wilson, Lawson, Muller, Tyler, & Kiefer, 2003). Thus, the processing and interpretation of spectrally degraded speech and other environmental sounds may be fundamentally different in normal-hearing listeners the instance it arrives at the inner ear.

In addition, prolonged periods of deafness also result in numerous cortical changes that will impact the amount of success patients have with their device. Neural changes may influence the auditory perceptual learning processes that cochlear implant patients use to understand speech and recognize environmental sounds. Unfortunately, deafness induced neural plasticity cannot be modeled through any method of signal processing used with normal-hearing listeners.

In conclusion, although using acoustic simulations of cochlear implants in normal-hearing listeners is proving to be an extremely valuable and informative supplement to cochlear implant research with clinical populations, there are several important limitations to this experimental methodology. Given the current limitations and the continuous changes in cochlear implant technology and patient characteristics, it has become increasingly important to focus research efforts directly to cochlear implant patients if the clinical population is readily available. One important focus of future research with both adult and pediatric cochlear implant users should be on fundamental perceptual learning processes. Perceptual learning is a foundational process to understand in this clinical population because it could be a factor that contributes to the success cochlear implant users have with their devices. In addition, perceptual learning and other cognitive process variables such as memory and attention may account for the enormous individual differences in outcome and benefit that are frequently observed but currently unexplained in many patients who have received cochlear implants.
References


132


APPENDIX A: STIMULUS MATERIALS

Harvard Meaningful Sentences

Acid burns holes in wool cloth.
A cloud of dust stung his tender eyes.
Act on these orders with great speed.
A cup of sugar makes sweet fudge.
Adding fast leads to wrong sums.
Add the column and put the sum here.
Add the stores account to the last cent.
Add the sum to the product of these three.
A filing case is now hard to buy.
A joy to every child is the swan boat.
A king ruled the state in the early days.
A lame back kept his score low.
A large size in stockings is hard to sell.
All sat frozen and watched the screen.
An abrupt start does not win the prize.
A pot of tea helps to pass the evening.
A rag will soak up spilled water.
A ridge on a smooth surface is a bump or flaw.
A rod is used to catch pink salmon.
A salt pickle tastes fine with ham.
A saw is a tool used for making boards.
A shower of dirt fell from the hot pipes.
A small creek cut across the field.
A speedy man can beat this track mark.
A tame squirrel makes a nice pet.
At that high level the air is pure.
A yacht slid around the point into the bay.
A young child should not suffer fright.
Bail the boat to stop it from sinking.
Both lost their lives in the raging storm.
Burn peat after the logs give out.
Cars and busses stalled in snow drifts.
Cats and dogs hate each other.
Drop the two when you add the figures.
Eight miles of woodland burned to waste.
Fairy tales should be fun to write.
Four hours of steady work paced us.
Glue the sheet to the dark blue background.
Guess the results from the first scores.
He broke a new shoelace that day.
He carved a head from the round block of marble.
Hedge apples may stain your hands green.
He knew the skill of the great young actress.
Help the woman get back to her feet.
Hemp is a weed found in parts of the tropics.
He ordered peach pie with ice cream.
He ran halfway to the hardware store.
Her purse was full of useless trash.
Host the load to your left shoulder.
Hop over the fence and plunge in.
It's easy to tell the depth of a well.
It snowed, rained, and hailed that same morning.
Kick the ball straight and follow through.
Leaves turn brown and yellow in the fall.
Lift the square stone over the fence.
March the soldiers past the next hill.
Mend the coat before you go out.
Move the vat over the hot fire.
Mud was splattered on the front of his white shirt.
Note closely the size of the gas tank.
Oak is strong and also gives shade.
Open the crate but don't break the glass.
Place a rosebush near the porch steps.
Press the pants and sew a button on the vest.
Pure bread puddles have curls.
Quench your thirst then eat the crackers.
Read verse out loud for pleasure.
Rice is often served in round bowls.
Shape the clay gently into block form.
She has a smart way of wearing clothes.
She sewed the torn coat quite neatly.
Sickness kept him home the third week.
Smokey fires lack flame and heat.
Split the log with a quick sharp blow.
Steam hissed from the broken valve.
Sunday is the best part of the week.
Take the winding road to reach the lake.
The two met while playing on the sand.
The third act was dull and tired the players.
The bark of the pine tree was shiny and dark.
The beauty of the view stunned the young boy.
The birch canoe slid on the smooth planks.
The box was thrown beside the parked truck.
The boy was there when the sun rose.
The child almost hurt the small dog.
The cigar burned a hole in the desktop.
The clock struck to mark the third period.
The coffee stand is too high for the couch.
The colt reared and threw the tall rider.
The crooked maze failed to fool the mouse.
The doctor cured him with these pills.
The dune rose from the edge of the water.
The empty flask stood on the tin tray.
The fish twisted and turned on the bent hook.
The friendly gang left the drugstore.
The frosty air passed through the coat.
The fruit peel was cut into thick slices.
The girl at the booth sold fifty bonds.
The heart beat strongly and with firm strokes.
The hog crawled under the high fence.
The hogs were fed chopped corn and garbage.
The horn of the car woke the sleeping cop.
The ink stain dried on the finished page.
The jacket hung on the back of the wide chair.
The just claim got the right verdict.
The lazy cow lay in the cool grass.
The lease ran out in sixteen weeks.
The meal was cooked before the bell rang.
The new girl was fired today at noon.
The office paint was a dull sad tan.
The pearl was worn in a thin silver ring.
The pencils have all been used.
The pennant waved when the wind blew.
The pipe began to rust while new.
The pirate seized the crew of the lost ship.
There the flood mark is ten inches.
There was a sound of dry leaves outside.
The ripe taste of cheese improves with age.
The rope will bind the seven books at once.
The salt breeze came across from the sea.
These days a chicken leg is a rare dish.
These thistles bend in a high wind.
The set of china hit the floor with a crash.
The shelves were bare of both jam and crackers.
The ship was torn apart by the sharp reef.
The show was a flop from the very start.
The sky that morning was clear and bright blue.
The small pup gnawed a hole in the sock.
The sofa cushion is red and of light weight.
The soft cushion broke the man's fall.
The source of the huge river is the clear spring.
The spot on the blotter was made by green ink.
The stray cat gave birth to kittens.

The swan dive was far short of perfect.
The term ended in late June that year.
The treetop waved in a graceful way.
The urge to write short stories is rare.
The wagon moved on well-oiled wheels.
The walled town was seized without a fight.
The wide road shimmered in the hot sun.
The wrist was badly strained and hung limp.
The felt gay when the ship arrived at port.
The young girl gave no clear response.
The young kid jumped the rusty gate.
The thieves who rob friends deserve jail.
This is a grand season for hikes on the road.
Those words were the cue for the actor to leave.
Tight curls get limp on rainy days.
To reach the end he needs much courage.
Torn scraps littered the stone floor.
Two blue fish swan in the tank.
Use a pencil to write the first draft.
We admire and love a good cook.
Weave the carpet on the right hand side.
We find joy in the simplest things.
We talked of the sideshow in the circus.
We tried to replace the coin but failed.
What joy there is in living.
Wipe the grease off his dirty face.
Wood is best for making toys and blocks.
Harvard Nonsense Sentences

A cone is no whole sheep on a sun.
A rude screen muffled his thirst limp.
A sour alarm is now good to read.
A thick screen can save this wild rack.
At that fine level the pedal is banned.
A vast mob does not fail the road.
A winding dinner lasts fine with pockets.
Batches came into raise the working lease.
Breathe the swords flood to the public lamp.
Brothers spill corner in the sharpest ducks.
Carry fans after the ruins finish out.
Cloth and floor like each snapper.
Coax the house but don’t sun the ads.
Crackers reach gray and rude in the paint.
Draw the pants form the restless coins.
Dunk your mail to a cat at a heavy gain.
Dust is best for stretching trinkets and clowns.
Eight stores of funds jangled to waste.
Find the shelves with a clean big button.
Gold facts should be sweet to happen.
Grass is the best weight of the wall.
Green ice can be used to sip a slab.
Green stories drilled the soft hammer.
Heave a new crowd to the council you light.
Heavy cork names have pins.
He played a new box that day.
He pressed the bid of the funny ripe bench.
He trotted gold thirst with tasty fun.
He wrote the long tar thirty seeds.
It gathered its shallow person in pink wit.
Juice is a better drip with a clear thief.
Metal can sew the most dull switch.
Pearl is accord used in flavors of the hero.
Piles and penny are early to eastern fever.
Plead the fake silk without shares.
Press the two when you say the tent.
Relax the idea of the thin graceful code.
Ribbons who work buyers reach salt.
Serve your logs to the red thaw.
Steam was twisted on the front of his dry grace.
She drove the fence quite deeply.
Slash the start to the pencil of these islands.
Soak the dust on the brisk high flaw.
Soap and sky is less then lamb.
Soap moves kits in green rocks.
Take your best towel to the third treadmill.
The blue chart is young and of thick tea.
The bowl stood and served its pages.
The brass spice is full of red leaves.
The clean chair flew on the old walnut.
The deep buckled walked the old crowd.
The dots lay beside the extra slate.
The draft on the dime was struck by thirty sheep.
The dust of the tan laugh was zestful and sharp.
The early trail was round and scared the dishes.
The empty stripe leaned off her news.
The fence began to float while soon.
The fail marsh got the cold wax.
The healthier he floated the less he got dropped.
The lawn wore a knife in the paper cup.
The light cord was seen today at noon.
The little orchid was a pleasant square spin.
The mule pierced him with these friends.
The oats helped the kite of the clear sheet.
The package almost circled the crooked gun.
The paper bag is too bright for the phone.
There was a vat of dense clams outdoors.
The round lathe for scarce morning is case.
The rush rented on the fast hostess.
The sail paved in the winds of the pleasant look.
These dice bend in a hot desk.
The soft birch of wires takes with map.
The sparks have all been told.
The stew was on the stone of the dusty crate.
The straw thought carved in a felt hat.
The streets fall with the hard hair of faults.
The swan lined the gem with a brass chorus.
The tea of a stuffed-chair is brass shaped.
The time could be met at the neater luck.
The tube that Sunday was deep and white silk.
The urge to send priceless glasses is old.
The yard stole when the train stung.
They could scoot although they were told.
They felt stately when the toad flickered in maple.
Trout is straight and also writes brass.
Try on these taps with blue cement.
We go when seeds wash a gold hip.
We tried to end the doll but failed.
Write the corn before the bright Tuesday.
Boys Town Meaningful Sentences

Apes swing from trees.
Bad clowns can't smile.
Big cars are loud.
Birds like long worms.
Blue planes fly far.
Boats sail at sea.
Brown cakes tastes rich.
Bug bites will itch.
Cats catch slow mice.
Clocks tick on time.
Close all three doors.
Cooks make hot food.
Cows give sweet milk.
Dad buys new shirts.
Dull paint won't shine.
Dump trucks fill holes.
Fall nights turn cool.
Feed your dogs meat.
Find bright pink shoes.
Fresh bread smells great.
Get back to bed.
Girls drink with straws.
Green cheese may stink.
He poked my eye.
It was her gum.
John wrecked his bike.
Knock these blocks down.
Large ducks splash hard.
Leave that trash out.
Most boys play ball.
Now draw four chairs.
Our socks look strange.
Pick up this room.
Pour me more tea.
Quick tails could wag.
Please don't throw rice.
Sam's tooth is loose.
Sharp nails might hurt.
Small kids need help.
Stay off the hill.
Tall men jump high.
Tell Mom those jokes.
The she drank pop.
They want some corn.
Tough guys sound mean.
Toy trains move fast.
True kings have crowns.
Warm sun feels good.

Boys Town Anomalous Sentences

We can be friends.
What a fun ride.
All boys are paint.
Bad dogs sail up.
Big apes grab sun.
Blocks can't run sharp.
Blue chairs draw well.
Brown night stinks too.
Cats get good boats.
Clocks catch old cows.
Cold worms have toys.
Cups give fat ducks.
Drums pour tall pats.
Dull socks wag off.
Feed my down tooth.
Feet drink hot bread.
Find girls these clouds.
Four rats close warm.
Green hands don't fall.
Guys tell loud meat.
Hard corn feels mean.
Help Dad tick out.
High bears move holes.
John bites at sound.
Jokes sleep on fields.
Large food jumps loose.
Late milk drank shirts.
Leave them cool fun.
Lend this ball home.
Long kids stay back.
Most birds knock tea.
Now straws need cheese.
Pink chalk is fresh.
Planes smile with mice.
Please shine some crowns.
Quick books look bright
Sad cars want chills.
Sam's hill could bike.
Slow bags itch far.
Smart shoes will play.
Snow smells more tough.
Strange nails taste dark.
Take splash from her.
Then pick what sea.
Throw his crown ice.
Time me and tails.
Tin cake may fill.
APPENDIX B: STIMULUS MATERIALS

A blue crane is a tall wading bird.
Acid burns holes in wool cloth.
A cloud of dust stung his tender eyes.
A cramp is no small danger on a swim.
Act on these orders with great speed.
A cup of sugar makes sweet fudge.
A dash of pepper spoils beef stew.
Adding fast leads to wrong sums.
Add the column and put the sum here.
Add the stores account to the last cent.
Add the sum to the product of these three.
A filing case is now hard to buy.
A fresh start will work such wonders.
After the dance they went straight home.
A joy to every child is the swan boat.
A king ruled the state in the early days.
A lame back kept his score low.
A large size in stockings is hard to sell.
All sat frozen and watched the screen.
Always close the barn door tight.
An abrupt start does not win the prize.
A pencil with black lead writes best.
A pot of tea helps to pass the evening.
A pound of sugar costs more than eggs.
A rag will soak up spilled water.
A ridge on a smooth surface is a bump or flaw.
A rod is used to catch pink salmon.
A salt pickle tastes fine with ham.
A saw is a tool used for making boards.
A shower of dirt fell from the hot pipes.
A small creek cut across the field.
A speedy man can beat this track mark.
A tame squirrel makes a nice pet.
At that high level the air is pure.
A waxed floor makes us lose balance.
A wisp of cloud hung in the blue air.
A yacht slid around the point into the bay.
A young child should not suffer fright.
A zestful food is the hot-crossed bun.
Bail the boat to stop it from sinking.
Both brothers were the same size.
Both lost their lives in the raging storm.
Bring your best compass to the third class.
Bring your problems to the wise chief.
Burn peat after the logs give out.
Canned pears lack full flavor.
Carry the pail to the wall and spill it there.
Cars and busses stalled in snow drifts.

Cats and dogs hate each other.
Clothes and lodging are free to new men.
Coax a young calf to drink from the bucket.
Corn cobs can be used to kindle a fire.
Cut the pie into large parts.
Do that with a wooden stick.
Drop the two when you add the figures.
Ducks fly north but lack a compass.
Eight miles of woodland burned to waste.
Even the worst will best his low score.
Fairy tales should be fun to write.
Farmers came into the oat crop.
Feel the heat of the weak dying flame.
Float the soap on top of the bath water.
Four hours of steady word paced us.
Fruit flavors are used in fizzy drinks.
Glue the sheet to the dark blue background.
Gray paint stretched for miles around.
Guess the results from the first scores.
He broke a new shoelace that day.
He broke his ties with former groups of friends.
He carved a head from the round block of marble.
Hedge apples may stain your hands green.
He knew the skill of the great young actress.
He lay prone and hardly moved a limb.
Help the woman get back to her feet.
Hemp is a weed found in parts of the tropics.
He ordered peach pie with ice cream.
He ran halfway to the hardware store.
Her purse was full of useless trash.
He said the same phrase thirty times.
He wrote his last novel there at the inn.
High seats are best for football fans.
Hoist the load to your left shoulder.
Hop over the fence and plunge in.
In some form or other we need fun.
It caught its hind paw in a rusty trap.
It’s easy to tell the depth of a well.
It’s hard to erase blue or red ink.
It snowed, rained, and hailed that same morning.
Jazz and swing fans like fast music.
Jerk the rope and the bell rings weakly.
Jump the fence and hurry up the bank.
Kick the ball straight and follow through.
Leaves turn brown and yellow in the fall.
Lift the square stone over the fence.
Live wires should be kept covered.
Madam this is the best brand of corn.
March the soldiers past the next hill.
Mend the coat before you go out.
Men strive but seldom get rich.
Move the vat over the hot fire.
Mud was splattered on the front of his white shirt.
Note closely the size of the gas tank.
Oak is strong and also gives shade.
Open the crate but don’t break the glass.
Paste can cleanse the most dirty brass.
Place a rosebush near the porch steps.
Pluck the bright rose without leaves.
Port is a strong wine with a smoky taste.
Press the pants and sew a button on the vest.
Pure bred poodles have curls.
Quench your thirst then eat the crackers.
Rake the rubbish up and then burn it.
Read verse out loud for pleasure.
Rice is often served in round bowls.
Schools for ladies teach charm and grace.
See the cat glaring at the scared mouse.
Sell your gift to a buyer at a good gain.
Shape the clay gently into block form.
She has a smart way of wearing clothes.
She sewed the torn coat quite neatly.
Sickness kept him home the third week.
Slash the gold cloth into fine ribbons.
Smoggy fires lack flame and heat.
Some ads serve to cheat buyers.
Split the log with a quick sharp blow.
Steam hissed from the broken valve.
Sunday is the best part of the week.
Take the winding road to reach the lake.
Tea served from the brown jug is tasty.
The ancient coin was quite dull and worn.
The bark of the pine tree was shiny and dark.
The beauty of the view stunned the young boy.
The birch canoe slid on the smooth planks.
The boss rand the show with a watchful eye.
The box was held with a bright red snapper.
The box was thrown beside the parked truck.
The boy was there when the sun rose.
The brown house was on fire to the attic.
The cement had dried when he moved it.
The child almost hurt the small dog.
The cigar burned a hole in the desktop.
The clock struck to mark the third period.
The club rented the rink for the fifth night.
The coffee stand is too high for the couch.
The colt reared and threw the tall rider.
The crooked maze failed to fool the mouse.
The cup cracked and spilled its contents.
The dark pot hung in the front closet.
The doctor cured him with these pills.
The dune rose from the edge of the water.
The empty flask stood on the tin tray.
The fin was sharp and cut the clear water.
The first worm gets snapped early.
The fish twisted and turned on the bent hook.
The friendly gang left the drugstore.
The frosty air passed through the coat.
The fruit of a fig tree is apple-shaped.
The fruit peel was cut into thick slices.
The girl at the booth sold fifty bonds.
The glow deepened in the eyes of the sweet girl.
The grass curled around the fence post.
The harder he tried the less he got done.
The hat brim was wide and too droopy.
The heart beat strongly and with firm strokes.
The hog crawled under the high fence.
The hogs were fed chopped corn and garbage.
The horn of the car woke the sleeping cop.
The hostess taught the new maid to serve.
The houses are built of red clay bricks.
The ink stain dried on the finished page.
The jacket hung on the back of the wide chair.
The just claim got the right verdict.
The lamp shone with a steady green flame.
The large house had hot water taps.
The lawyer tried to lose his case.
The lazy cow lay in the cool grass.
The lease ran out in sixteen weeks.
The loss of the second ship was hard to take.
The lure is used to catch trout and flounder.
The map had an X that meant nothing.
The meal was cooked before the bell rang.
The navy attacked the big task force.
The new girl was fired today at noon.
The office paint was a dull sad tan.
The paper box is full of thumb tacks.
The pearl was worn in a thin silver ring.
The pencils have all been used.
The pennant waved when the wind blew.
The petals fall with the next puff of wind.
The pipe began to rust while new.
The pirates seized the crew of the lost ship.
The play began as soon as we sat down.
The play seemed dull and quite stupid.
The prince ordered his head chopped off.
There are more than two factors here.
There the flood mark is ten inches.
There was a sound of dry leaves outside.
The ripe taste of cheese improves with age.
The rope will bind the seven books at once.
The rude laugh filled the empty room.
The rush for funds reached it's peak Tuesday.
The salt breeze came across from the sea.
These days a chicken leg is a rare dish.
These pills do less good than others.
These thistles bend in a high wind.
The set of china hit the floor with a crash.
The shaky barn fell with a loud crash.
The shelves were bare of both jam and crackers.
The ship was torn apart by the sharp reef.
The show was a flop from the very start.
The sky that morning was clear and bright blue.
The slang word for raw whisky is booze.
The slush lay deep along the street.
The small pup gnawed a hole in the sock.
The sofa cushion is red and of light weight.
The soft cushion broke the man's fall.
The source of the huge river is the clear spring.
The spot on the blotter was made by green ink.
The stray cat gave birth to kittens.
The swim dive was far short of perfect.
The term ended in late June that year.
The third act was dull and tired the players.
The tiny girl took off her hat.
The tongs lay beside the ice pail.
The train brought our hero to the big town.
The treetop waved in a graceful way.
The two met while playing on the sand.
The urge to write short stories is rare.
The wagon moved on well-oiled wheels.
The walled town was seized without a fight.
The warf could be seen at the farther shore.
The wide road shimmered in the hot sun.
The wreck occurred by the bank on main street.
The wrist was badly strained and hung limp.
They are men who walk in the middle of the road.
They are pushed back each time they attack.
They could laugh although they were sad.
They felt gay when the ship arrived at port.
They floated on the raft to sun their white backs.
They took the axe and the saw to the forest.
The young girl gave no clear response.
The young kid jumped the rusty gate.
Thieves who rob friends deserve jail.
This is a grand season for hikes on the road.
This will lead the world to more sound and fury.
Those words were the cue for the actor to leave.

Tight curls get limp on rainy days.
To make pure ice you freeze water.
To reach the end he needs much courage.
Torn scraps littered the stone floor.
Try to have the court decide the case.
Two blue fish swan in the tank.
Two plus seven is less than ten.
Type out three lists of orders.
Use a pencil to write the first draft.
We admire and love a good cook.
We are sure that one war is enough.
Weave the carpet on the right hand side.
We find joy in the simplest things.
We frown when events take a bad turn.
We talked of the sideshow in the circus.
We tried to replace the coin but failed.
What joy there is in living.
Where were we when the noise started?
Whitings are small fish caught in nets.
Wipe the grease off his dirty face.
Wood is best for making toys and blocks.
Write a fond note to the friend you cherish.
Write at once or you may forget it.
Yell and clap as the curtains slide back.
APPENDIX C: STIMULUS MATERIALS

A cup of sugar makes sweet fudge.
Adding fast leads to wrong sums.
A king ruled the state in the early days.
A large size in stockings is hard to sell.
Always close the barn door tight.
A pot of tea helps to pass the evening.
A pound of sugar costs more than eggs.
A rag will soak up spilled water.
A rod is used to catch pink salmon.
A saw is a tool used for making boards.
A shower of dirt fell from the hot pipes.
A small creek cut across the field.
A tame squirrel makes a nice pet.
A tusk is used to make costly gifts.
A wisp of cloud hung in the blue air.
A yacht slid around the point into the bay.
Bail the boat to stop it from sinking.
Both lost their lives in the raging storm.
Cars and busses stalled in snow drifts.
Cut the pie into large parts.
Four hours of steady work paced us.
Glue the sheet to the dark blue background.
He lay prone and hardly moved a limb.
He ran halfway to the hardware store.
Help the woman get back to her feet.
Her purse was full of useless trash.
Hoist the load to your left shoulder.
Hop over the fence and plunge in.
It snowed, rained, and hailed that same morning.
It’s easy to tell the depth of a well.
Kick the ball straight and follow through.
Lift the square stone over the fence.
March the soldiers past the next hill.
Men strive but seldom get rich.
Mend the coat before you go out.
Mesh wire keeps chicks inside.
Note closely the size of the gas tank.
Place a rosebush near the porch steps.
Press the pants and sew a button on the vest.
Read verse out loud for pleasure.
Rice is often served in round bowls.
See the cat glaring at the scared mouse.
Sickness kept him home the third week.
Take the winding road to reach the lake.
Ten pins were set in order.
The beauty of the view stunned the young boy.
The bill was paid every third week.

The birch canoe slid on the smooth planks.
The box was thrown beside the parked truck.
The boy was there when the sun rose.
The clock struck to mark the third period.
The colt reared and threw the tall rider.
The crooked maze failed to fool the mouse.
The cup cracked and spilled its contents.
The dune rose from the edge of the water.
The fin was sharp and cut the clear water.
The fish twisted and turned on the bent hook.
The friendly gang left the drugstore.
The frosty air passed through the coat.
The girl at the booth sold fifty bonds.
The grass curled around the fence post.
The hat brim was wide and too droopy.
The heart beat strongly and with firm strokes.
The hogs were fed chopped corn and garbage.
The ink stain dried on the finished page.
The lawyer tried to lose his case.
The lazy cow lay in the cool grass.
The lease ran out in sixteen weeks.
The meal was cooked before the bell rang.
The navy attacked the big task force.
The pearl was worn in a thin silver ring.
The play seems dull and quite stupid.
There are more than two factors here.
The rope will bind the seven books at once.
The salt breeze came across from the sea.
These days a chicken leg is a rare dish.
The set of china hit the floor with a crash.
The ship was torn apart by the sharp reef.
The snow was a flop from the very start.
The slush lay deep along the street.
The small pup gnawed a hole in the sock.
The soft cushion broke the man’s fall.
The source of the huge river is the clear spring.
The stray cat gave birth to kittens.
The swan dive was far short of perfect.
The term ended in late June that year.
The two met while playing on the sand.
The urge to write short stories is rare.
The wagon moved on well-oiled wheels.
The walled town was seized without a fight.
The wide road shimmered in the hot sun.
The wrist was badly strained and hung limp.
The young girl gave no clear response.