6-2023

Examining the Effect of Talker Familiarity Using Familiar and Unfamiliar Talkers on Noise-Vocoded Speech Perception in Normal-Hearing Listeners: A Training Study

Daria N. Collins
The Graduate Center, City University of New York

EXAMINING THE EFFECT OF TALKER FAMILIARITY USING FAMILIAR AND UNFAMILIAR TALKERS ON NOISE-VOCODED SPEECH PERCEPTION IN NORMAL-HEARING LISTENERS: A TRAINING STUDY

by

DARIA COLLINS

A dissertation submitted to the Graduate Faculty in Audiology in partial
fulfillment of the requirements for the degree of Doctor of Audiology,
The City University of New York

2023

APPROVAL

Examining the Effect of Talker Familiarity Using Familiar and Unfamiliar Talkers on Noise-Vocoded Speech Perception in Normal-Hearing Listeners: A Training Study

A capstone research project submitted by

Daria Collins

This manuscript has been read and accepted for the Graduate Faculty in Audiology in satisfaction of the capstone project requirement for the degree of Doctor of Audiology.

Approved: April 2023

Meital Avivi-Reich, Advisor

Dorothy DiToro, Executive Officer

THE CITY UNIVERSITY OF NEW YORK

ABSTRACT

Examining the Effect of Talker Familiarity using Familiar and Unfamiliar Talkers on Noise-Vocoded Speech Perception in Normal-Hearing Listeners: A Training Study

by

Daria Collins

Advisor: Meital Avivi-Reich

Auditory training studies utilizing stimuli that are applicable to real-world processing of speech have been shown to improve speech perception abilities in normal hearing populations, those with hearing loss, and cochlear implant wearers. In particular, exposing normal hearing adults to noise-vocoded speech via auditory training studies has been shown to not only simulate the perceptual experience of a cochlear implant wearer, but have demonstrated promising improvements on speech perceptual abilities via the training paradigm. Additionally, studies have highlighted various variables that impact speech perception, including, talker familiarity. Talker familiarity has been shown to enhance speech perception both in listeners with normal hearing and those with hearing loss. This study aims to create a multi-session training program using a population of normal hearing adults during which listeners would be exposed to either natural or noise-vocoded sentences. After completing three training sessions over a week, participants will be asked to recognize noise-vocoded sentences spoken by the same talker they have been exposed to during training as well as an unfamiliar talker they have not been exposed to previously. In particular, this portion of the larger study will focus on talker familiarity and the impacts of training listeners on familiar versus unfamiliar speakers in relation to noise vocoded speech perception performance. This study is unique in that it is the first study to investigate both talker familiarity and noise vocoded speech familiarity in the scope of a multi-session auditory training program. The results of the suggested study may also illuminate which training conditions could potentially optimize the improvement in noise-vocoded speech perception. The findings from this study may be useful for further research pertaining to those with hearing loss and/or those with cochlear implants and potentially assist in designing a preimplantation training program to assist patients’ transition to CI.

ACKNOWLEDGMENTS

Thank you Mom, Dad, and Dina for your endless support throughout my graduate school journey, and for pushing me to not only reach for the stars, but to look beyond them and aim even higher. Your love and guidance have helped me achieve my dreams, and for that, I am eternally thankful. Thank you to Dr. Barbara Weinstein for being my mentor, for always encouraging me to pursue life-changing opportunities, and most of all, for believing in me. You have inspired me to not only continually learn and grow, but to also be an advocate for those that I wish to serve. To Dr. DiToro and Dr. Wortsman, thank you for being monumental in my journey in becoming an audiologist. Thank you to my co-author Alexis Leiderman for being my friend all these years, for sticking with me through thick and thin (virtually and in-person), and for completing this study together; we made it! To Madeline Campbell, Robin Axelrod, and Gila Orenshein, thank you for dedicating your time and efforts in helping Alexis and I to develop this training study. Finally, an immense thank you to my capstone mentor, Dr. Avivi-Reich, for all your support and guidance in helping to materialize this study.

CONTENTS

TABLES

This Capstone Project is dedicated to Mom, Dad and Dina.

1. Background

Auditory training studies have been shown to improve speech perception abilities. Many of these studies train participants utilizing a combination of methodologies and stimuli that are applicable to real-world situations, such as training participants on the recognition of sentences (Barker, 2006; Davis et al., 2005; Sheldon, Pichora-Fuller & Schneider, 2008; Stacey & Summerfield, 2007). Other studies have investigated the effects of incorporating talker-specific information within the training and testing protocols and analyzing the resulting performance on tasks of speech perception (Barker, 2006; Jesse, McQueen & Page, 2007; Domingo et al., 2020; Kreitewolf et al., 2017; Mullennix, Pisoni, and Martin, 1989; Nygaard & Pisoni, 1998; Nygaard, Sommers & Pisoni, 1994; Yonan & Sommers, 2000). Particularly, exploring the training of talker-specific information, such as if familiar versus unfamiliar talkers boosts speech perception, is of interest with auditory training due to benefits in the accuracy of processing linguistic information (Levi, Winters & Pisoni, 2011; Nygaard & Pisoni, 1998; Nygaard, Sommers & Pisoni, 1994; Yonan & Sommers, 2000), generalizing to novel stimuli (Kreitewolf et al., 2017), and potential improvements in the speed of speech perception (Jesse, McQueen & Page, 2007; Stacey and Summerfield, 2007). The purpose of this capstone is to suggest a novel auditory training program using normal-hearing listeners on tasks of speech perception of noise-vocoded speech utilizing familiar and unfamiliar talkers. This portion of the study will investigate the effects of familiar versus unfamiliar talker on noise-vocoded speech perception performance. The other portion of the study by Alexis Leiderman (2023) will focus on the effects of noise-vocoded signal exposure on speech perception.

Speech perception is an important, everyday skill which allows the listener to discriminate between distinct elements of speech, identify those elements, and interpret them. The overall speech perception mechanism is thought to be a combination of perceptual auditory processing and active cognitive processing, which influences the flexibility of using different cognitive elements to understand spoken language (Nusbaum & Schwab, 1986; Nusbaum & Magnuson, 1997), and passive cognitive processing, which oversees the sequencing of cognitive operations for speech perception (Nusbaum & Schwab, 1986). Neurophysiological mapping of the auditory cortex suggests parallel auditory representations pertaining to talker-specific indexical information, the recovery of phonetic categories, as well as information regarding speech events and context (Dick et al., 2010, Hackett, 2007, Woods et al., 2009). These processes work together to enable overall communication. Speech perception is crucial for communication, and a disruption in this process can have implications on one's communication ability. This disruption is overarching, as it may also have implications for one's ability to succeed in work and school environments, forming interpersonal relationships, and with one's overall self-image. For those who have difficulties with understanding speech, various interventions can be considered, such as the use of amplification devices and/or the use of communication and compensatory strategies. For those with considerable difficulties with speech perception, auditory training may be considered.

The goal of auditory training is to help listeners make meaningful and purposeful distinctions and interpretations between certain speech sounds through use of a systematic training paradigm (Olson, 2015). A study conducted by Boothroyd (2010) found that auditory training may improve confidence with communication, improve social participation, and improve one's overall quality of life. Auditory training has been offered in-person or via computer-based programs, both of which have been shown to be effective in enhancing this skillset (Cosby, 2011; Gil & Iorio, 2010; Henshaw & Ferguson, 2013; Stropahl et al., 2020; Stacey et al., 2010; Stacey & Summerfield, 2007). Traditionally, auditory training aims to improve speech perception by training listeners to make perceptual distinctions between sounds that are presented in a systematic manner (Henshaw & Ferguson, 2013). Auditory training can be broken down into two types of activities: synthetic and analytic (Olson, 2015). Paradigms that use drill-like activities are thought to be analytic, whereas those that use sentence identification or passage comprehension are synthetic training. Recently, auditory training programs have evolved from focusing on phonemic discrimination, to tasks that build on skills that are more transferable to real-word communication (Tye-Murray et al., 2017). Studies, such as Humes et al. (2009) selected the 600 most frequently used words that are encountered in over 90% of all conversations and trained listeners on those words, finding significant improvements in open-set recognition of speech. Another study conducted by Barcroft et al., (2011) utilized meaning-oriented training by tasking listeners to not only distinguish between word pairs, but having listeners map those linguistic forms to their meaning by using a picture identification task and mapping the words to their respective picture representations, demonstrating improvements in word recognition. Training studies with sentences as stimuli, which are more applicable to real-world speech perception and add in contextual factors, have been found to be an effective training strategy (Barker, 2006; Davis et al., 2005; Sheldon, Pichora-Fuller & Schneider, 2008; Stacey & Summerfield, 2007). Overall, there is no one way that auditory training can be conducted, and the use of stimuli and the tasks that are learned all have individual distinctions.

Auditory training has been investigated in a variety of populations and has been shown to enhance speech perception performance. A particular topic of interest is investigating the underlying mechanisms that correlate with improvements in speech perception. Overall, speech processing is impacted by multiple cognitive factors. For example, studies have demonstrated the relationship between motivation, context, and cognition on speech and communication (Kiessling et al., 2003; Pichoa-Fuller & Singh, 2006; Rudner et al. 2011). Stacey and Summerfield (2007) theorize that speech perception performance may improve following auditory training due to a remapping of prior existing linguistic knowledge following exposure to stimuli, heightened phonemic discrimination, or even an improvement in the mechanisms that affect one's ability to attend to auditory information. Overall, it has been established that speech perception is influenced by active and passive cognitive processes, therefore it is reasonable to assume that auditory training may affect and shape these processes in some way. These speculations call to light the need for more research to understand the higher-level processes that contribute to speech perception, particularly with noise-vocoded speech.

One such high-level mechanism that has been found to affect speech perception is related to talker characteristics. Talker characteristics are acoustic features that are utilized to recognize talkers. This is also known as indexical characteristics. These characteristics include pitch, timbre, the acoustic effect of how the message is articulated, anatomy (Kuwabara & Takagi, 1991), dialect (Clopper & Pisoni, 2004), speaking rate (Sommers, Nygaard & Pisoni 1994), emotional state (Murray & Arnott, 1993), and age (Hartman & Danauer, 1976; Linville & Fisher, 1985). All these factors introduce large amounts of variability into the speech signal, but despite this, listeners are able to understand speech from the same, as well as different talkers fairly easily. Thus, to have adequate speech perception, Levi (2015) suggests that perceptual abilities should be flexible to properly adapt to variabilities present in the signal. These talker specific characteristics are perceived and memorized with the speech message, and this in turn shapes speech perception (Barker, 2017). There are several theories that aim to explain the mapping of talker characteristics, as explained in Magnuson et al. (2021). Once such theory suggests that talker-specific characteristics and phonemic categories are integrated within the talker's own voice characteristics, resulting in this information being normalized (Gerstman, 1968; Ladefoged & Broadbent, 1957; Miller, 1989; Nearey, 1989; Nusbaum & Morin, 1992; Potter &Steinberg, 1950; Rakerd &Vebrugge, 1987; Strange, 1989; Syrdal & Gopal, 1986; Traunmuller, 1981). With this theory, mapping for talker-specific information and characteristics in relation to the acoustic signal and phonemic categories must all be done in real-time. One drawback is that it may lead to increased demands of attention processing as talkers are changed, which may impact performance of speech perception (Wong et al., 2004). Another theory is that talker-specific auditory exemplars are encoded in memory around various linguistic categories, meaning that phonemic structure and talker specific acoustic information are encoded together (Johnson, 1994,1997,2005; Pierrehumbert, 2002). Speech signals stimulate stored similar exemplars, which boosts recognition, and the efficiency of recognition directly relates to the number of exemplars that are similar to the incoming signal.

Studies have shown that indexical information, or information that encodes talker specific characteristics, is important for speech processing, such as word recall and recognition (Barker, 2006; Levi, Winters & Pisoni, 2011; Nygaard & Pisoni, 1998; Nygaard, Sommers & Pisoni, 1994; Olson, 2015; Sommers & Pisoni, 1994; Yonan & Sommers, 2000). This talker information is rarely learned explicitly, but rather incidentally while the listener is being engaged in speech comprehension. Therefore, implicit training with stimuli spoken by multiple talkers is applicable to real-world situations. While training with multiple talkers may be advantageous and worth exploring, the normalization hypothesis of talker specific processing information supports a decrease speech perception with increased talker variability in training, which may result in increased errors and slower performance on tasks of speech perception (Mullennix, Pisoni, & Martin 1989), therefore training with one talker may be more beneficial.

Literature suggests that talker familiarity can enhance speech perception, which is known as the familiarity benefit. This benefit has been found in implicit and explicit training (Barker, 2006; Kreitewolf et al., 2017; Levi, 2015 Mullennix, Pisoni, and Martin, 1989). The concept is that, if the speaker is familiar to the listener, this enhances lexical access. Lexical access allows one to access semantic meaning, which in turn allows enhanced recognition and perception of speech. In studies where listeners were explicitly trained long-term to identify voices of native (Nygaard & Pisoni, 1998; Nygaard, Sommers & Pisoni, 1994; Yonan & Sommers, 2000) or non-native talkers (Levi, Winters & Pisoni, 2011), they found that not only were subjects able to identify what a talker had said, but that subjects processed linguistic information more accurately with speakers that they were familiarized to compared to when they were presented the information by novel talkers. The familiarity benefit also persists when subjects are trained and tested on multiple talkers (Nygaard & Pisoni, 1998). Even more interestingly, talker familiarity that is implicitly learned is generalized to novel stimuli, particularly when subjects were given novel sentences, which suggests that talker information is not an instance specific exemplar (Goldinger, 1998), but is transferable to novel stimuli (Kreitewolf et al., 2017). Additionally, listeners are faster at speech perception, particularly in a lexical decision task, when exposed to the same talker producing targeted sounds (Jesse, McQueen & Page, 2007), which may prove to be beneficial for auditory training efficiency. While there have been various studies that look at the benefits of talker familiarity and speech perception, further research is needed to determine the amount of time necessary to yield a familiar talker advantage, which may be helpful in maximizing familiar talker effects in training studies (Domingo et al., 2020). Additional research is also warranted to identify what exact indexical information is integrated and how that integration occurs in order to create the talker familiarity effect (McLaughlin et al., 2019).

Interestingly, clinical populations may experience talker familiarity differently, which may have implications on speech processing abilities. For example, Best et al., (2018) found that older and younger populations with HL perform worse on tasks of talker identification in noise and in quiet compared to populations without HL. Ching and Dillon (2013) highlight that reduced audibility, reduced frequency selectivity through broadened auditory filters within the cochlea, cochlear dead regions, as well as reduced cognition all play a part in an overall reduction in overall speech intelligibility. For those with CIs, the CI itself may limit the amount of indexical information perceived due to device limitations. These limitations include, but may not be limited to, the number of processing channels (Wilson, 2004), and the CI wearer's ability to process speech due to limitations within their individual auditory systems (Barker, 2006; Wilson, 2004). Investigating which particular factors of indexical, talker-specific information are or are not transmitted in CI wearers is still an area of interest. Barker (2006) suggests that CI wearers use and perceive talker-specific information differently than NH listeners. Overall, it would be of benefit to explore how much of the talker characteristics are mapped during training are applied in real-time tasks of speech processing.

Auditory training has been utilized in populations with hearing loss (HL) (Cosby, 2011; Henshaw & Ferguson, 2013; Olson, 2015; Rubinstein & Boothroyd, 1987; Stropahl, & Launer, 2020), as well as cochlear implant (CI) wearers (Barker, 2006; Castiglione et al., 2016; Reis et al., 2021; Tye-Murray, Witt, & Schum, 1995). Auditory training in normal hearing (NH) listeners has shown to be effective in improving speech perception scores (Davis et al., 2005;

McGettingan et al., 2014). Such studies train NH populations by listening to processed speech which emulates what CI wearers hear to investigate effects of the processed stimuli, training parameters, and the effects of the training on speech perception. For example, Chisolm (2012) conducted a meta-analysis of auditory training that used participants with HL and highlighted numerous studies that found that auditory training had a reliable but small post-training improvement in speech recognition performance in NH listeners. For those with HL, literature has also shown similar benefit as seen with studies using a NH population. One notable example is from Rubenstein and Boothroyd (1987), who conducted an auditory training study composed of twenty older adults with mild to moderate sensorineural hearing loss (SNHL). After four weeks of auditory training, there were statistically significant improvements in speech recognition scores.

Auditory training for those with CIs is unique in that they are relearning how to understand speech through a novel auditory signal than what is heard by NH listeners and even those with HL. A CI is a device that creates an electrical signal that is representative of the acoustic, auditory stimulus. This signal is created using an external processor, and through the use of an electrode array that is placed into the cochlea, which mirrors the function of the basilar membrane. This electrode array then sends signals directly to the auditory nerve, which travels up through the central auditory nervous system so they can be interpreted meaningfully. This processed signal is vastly different than what is normally produced in an NH ear, or even what is perceived by those with HL, thus those who receive a CI must undergo intensive (re)habilitative therapy in order to learn how to process and interpret this new auditory input. Auditory (re)habilitation is important for many with CIs due to the broad outcome variability that persists regarding speech recognition due to residual hearing (Blamey et al., 2013) as well as the spectrotemporal resolution of the devices themselves (Friesen et al., 2001). Therefore, the types of auditory training utilized for CI wearers can be categorized as being passive or active in nature. Passive training is where listeners adapt to hearing using a CI through unstructured and unsupervised activities, whereas active auditory training engages those in listening tasks that target phonemic learning, specific word understanding, or overall sentence comprehension (Drouin & Theodore, 2020). Overall, auditory training can be beneficial for CI users who demonstrate a range of speech perception capabilities, by providing opportunities that contain meaningful linguistic information, and by providing opportunities within situations that are difficult (Drouin & Theodore, 2020).

Of particular interest are training studies that have utilized processed speech that simulates the acoustic information that is heard by CI wearers, called noise-vocoded speech. Originally developed by Bob Shannon and explored in Shannon et al. (1995), noise-vocoded speech is processed by dividing the original speech signal into logarithmically spaced frequency bands, extracting the amplitude envelope for each band, then using them to modulate noise. Finally, those noise bands are recombined to create the processed speech. One of the most important parameters to highlight when creating this stimulated, processed speech is the number of noise bands used. Noise-vocoded stimuli used in auditory training has been studied using a range of noise bands, from as little as four (Shannon et al., 1995) to as many as ten bands, where the use of more bands translates to easier speech perception compared to the use of fewer bands (Davis, 2005, Davis & Johnsrude, 2003; Loizou, Dorman, & Tu, 1999; Shannon et al., 1995; Sheldon et al., 2008). Overall, having as little as four bands is intelligible with training, whereas up to ten bands is highly intelligible with less training (Davis, 2005; Shannon et al., 1995). Among these auditory training studies that have used processed speech in NH populations, findings have shown that this processed speech not only reflects what a CI wearer perceives, but the training resulted in enhanced speech perception (Faulkner et al., 2012; Stacey & Summerfield, 2007; Shannon et al., 1995). Overall, such findings utilizing noise-vocoded speech in NH populations have been beneficial in investigating if using such training programs in a CI population would be effective.

The training benefit found in NH populations using noise-vocoded speech has also been demonstrated in training studies with a CI population. An intensive word-based task created by Fu et al. (2005) led to significant improvements on three measures of speech perception: vowel discrimination, consonant discrimination, and sentence perception in CI wearers. Additionally, a follow-up study of Stacey and Summerfield (2007) conducted by Stacey et al. (2010) found significant improvements in consonant discrimination in CI wearers after undergoing auditory training. This study found no significant improvements on sentence or vowel discrimination tests, highlighting a need for future studies to further investigate the effectiveness of auditory training for CI wearers.

In summary, auditory training studies have been shown to improve speech perception abilities in NH listeners, those with HL, and CI wearers. An improvement has been demonstrated in studies utilizing stimuli that are applicable to real-world situations, such as sentences. Studies that have utilized scenarios that reflect and affect real-world processing of speech perception, such as the use of a familiar versus different talkers, have also shown benefit in terms of improving speech perception. The purpose of this capstone is to investigate talker familiarity and the impacts of training listeners on familiar versus unfamiliar speakers in relation to noise vocoded speech recognition performance. Participants will be trained on a familiar target talker across three training conditions. Participants will be separated into one of four groups to investigate the interaction of talker familiarity and noise vocoded speech familiarity to assess the best training protocol that influences noise-vocoded Speech learning, as well as the effects of Talker Familiarity on Speech Perception measures. This study is unique in that it is the first study to investigate talker familiarity and speech perception utilizing noise-vocoded speech sentence stimuli in the scope of a multi-session auditory training program. Investigating these interactions can perhaps provide additional information regarding how they impact speech perception, and may further shed light on these processes in a simulated CI population. Additionally, Using a NH population and investigating the effectiveness of this training protocol on speech perception abilities can further our understanding in designing and applying novel training protocols for future clinical use, such as in designing a preimplantation training program to assist patients' transition to a CI.

2. Proposal

This training study aims to investigate the effect of Talker Familiarity on auditory training when testing noise-vocoded Speech Perception in normal hearing young adults. In particular, normal-hearing individuals will be trained using either natural or noise-vocoded Speech utilizing familiar and unfamiliar talkers. Talker familiarity has been found to improve speech perception, as it provides indexical lexical cues such as pitch, timbre, anatomy, dialect, speaking rate, emotional state, and even age that can introduce additional assisting information when listening to a signal that is unfamiliar and novel. It is hypothesized that training using noise-vocoded speech with a familiar talker will result in a greater improvement of noise-vocoded speech perception than seen when training using vocoded speech of an unfamiliar talker. However, when natural speech is used during training the benefit of talker familiarity will be lesser than when exposing the listeners to noise-vocoded speech, some of the indexical lexical cues will be unidentified once the speech is noise-vocoded. By focusing on analyzing the two variables of speech (Talker Familiarity and familiarity of Type of Speech), this study could provide insight as to factors that contribute to the learning of distorted speech. The suggested study investigated the effects of 2 main variables and the possible interaction between them. This portion of the study will focus on understanding the main effect of Talker Familiarity when a listener is trained on hearing a familiar talker versus hearing a non-familiar talker. The other portion of the study, conducted by Alexis Leiderman (2023), will focus on understanding the underlying main effect involved with learning a novel distorted speech stimulus.

To investigate the effects of these two variables, four conditions will be designed that will feature combinations of the two independent variables. Manipulation of these variables will help to elucidate the improvement of speech perception within these various conditions.

The main aims of the proposed study are:

1. To examine how training with familiar versus unfamiliar talkers improves speech recognition of noise-vocoded speech in normal-hearing individuals.
2. To examine how training with processed speech improves noise-vocoded speech perception in normal-hearing individuals. This aim will be addressed in a different portion of the study conducted by Alexis Leiderman.
3. To explore how Talker Familiarity and Type of Speech (natural vs. noise-vocoded) may interact when training normal hearing adults to perceive noise-vocoded speech.

The long-term goal of this joint study is to determine which combination of variables, such as familiar versus unfamiliar target speaker, as well as normal voice vs. noise-vocoded voice, could lead to the optimal training outcome to improve noise-vocoded speech perception.

The proposed design of the experiment will require participants to consent to a week-long training protocol consisting of three training sessions followed by a test session upon completion of the training. Each training session and test session will be held on a separate day to allow enough time for consolidation (Stacey et al., 2007).

3. Hypotheses

Based on information available in the current literature, the following hypotheses have been made:

1. Groups A and B will demonstrate an improved ability to correctly recognize noise-vocoded speech across training sessions and during the test session, recognizing more target words during testing than Group C and D.
2. During testing, Groups A and B will perform better when presented with noise-vocoded sentences spoken by a familiar talker (used during training) than when presented with sentences spoken by an unfamiliar talker.
    - This study is novel in that it has not been established how much of the indexical information learned from exposure to natural speech during training could be applied and used as an assisting cue when listening to a noise-vocoded speech of the same talker.
3. Groups A and B will recognize more target words within unfamiliar sentences spoken by the familiar talker than with sentences spoken by the unfamiliar talker within the test session.

   - Kreitewolf et al. (2017) highlighted that talker familiarity can be generalized to novel stimuli. This effect has not yet been explored utilizing noise-vocoded stimuli.

4. During testing, Groups C and D will perform slightly better when presented with noise-vocoded sentences spoken by a familiar talker (used during training) than when presented with sentences spoken by an unfamiliar talker. We expect Talker Familiarity to have a smaller effect for Groups C and D, since previous exposure during training was only to the natural voice of the familiar talker. We also expect Groups C and D to perform poorer due to indexical cues that are lost in the speech signal once it is noise-vocoded.

4. Methods

4.1. Subjects

Twenty-four normal-hearing participants between the ages of 18 to 25 will be recruited. This age range was selected to reduce the likelihood of a participant having HL, as statistically, this age range is less likely to present with hearing loss (Haile et al., 2021). Additionally, screening for normal hearing sensitivity is important, as all participants must have normal hearing to ensure audibility of noise-vocoded speech. The participants will be randomly assigned to either Groups A through D, resulting in six participants in each group. The reasoning behind having this sample size was based on the power analysis that concluded having twenty-four participants will have a significant effect to determine overall improvement in scores. To screen individuals and ensure participants are as homogeneous as possible, participants will complete a questionnaire prior to participating (see Appendix A). Inclusion criteria will include: ensuring the participants are within the targeted age range, no self-reported hearing loss, no delay in acquisition of speech, no history of excessive noise exposure, similar educational status among participants in all four groups, similar linguistic status/ experience (native English-speaking individuals), and will ensure participants are in good health. The questionnaire was created using Gorilla Software and will include participant consent, HIPAA guidelines, and experiment-specific information as well.

Participants will be recruited from the New York City tri-state area via flyers and recruitment materials posted around CUNY-affiliated institutions. Digital versions of these materials will also be sent to the institutions to be distributed electronically. Finally, participants will be monetarily compensated for participation in the study.

4.2 Procedures

This study will be conducted via the Gorilla Experiment Builder, which is a computer-based experimental software. An in-person version of this study has also been outlined, of which the procedure can be seen in Appendix B. The study will occur over the span of a week. Participants will be sent a link to access the experiment which will consist of three training sessions, followed by a test session that will be conducted one to two days after the completion of the third training session. Pre-recorded sentences will be administered during the three training sessions and the test session via headphones on the individual's personal computer. Participants will be instructed to record the specified target words in an answer box, which will be displayed on the computer screen, of the sentences that they heard for each trial. All responses will be recorded automatically and sent back to the primary investigators through the Gorilla Experiment Builder. Percent correct will be scored based on the number of target words that were repeated correctly by each participant.

Regarding the effectiveness of conducting an auditory training study in person or online, studies, such as Stacey and Summerfield (2007) and Stacey et al. (2010), have shown success with utilizing similar computer-based, noise-vocoded speech auditory training programs. In particular, Stacey and Summerfield (2007) conducted an online training study on the training of the perception of noise-vocoded speech and found improvement in speech perception amongst participants. What makes this study novel is that the effects of talker familiarity in the training of noise-vocoded speech in an online training study has yet to be investigated. Additionally, Faulkner et al., 2012 and Stacey and Summerfield (2007) found that when training NH populations using computer and live-voice auditory training, computer training is as effective as live-voice training, which may be easier to conduct and translate for a CI population. If the study is conducted remotely, the participants will be reminded before each session to ensure that they are in a quiet room to partake in the sessions with no distractions to reduce the likelihood of extraneous variables.

Overall, literature supports using one to three training sessions (Davis et al., 2005; Sheffert et al., 2012; Stacey & Summerfield, 2007) and having each training session and final test session be between 20 to 30 minutes long (Davis et al., 2005; Rosen et al., 1999; Sheffert et al., 2012; Stacey & Summerfield, 2007). Participants will not be allowed to complete more than one training session per day. This decision was made to maximize the perceptual learning from each training session by allowing time for consolidation (Stacey et al., 2007). In terms of the final test session, participants will then be exposed to four vocoded conditions: sentences from training sessions that are using the same talker as training sessions with their voice vocoded (condition 1), novel sentences not from training sessions that are using the same talker as training sessions with their voice vocoded (condition 2), sentences from training sessions that use a different talker with their voice vocoded (condition 3), and novel sentences not from training sessions that use a different talker with their voice vocoded (condition 4); see Table 1.

Testing Conditions			
Condition	**Sentence Type**	**Talker**	**Voice**
Condition 1	Sentences from Training Session 1, 2 or 3	Talker 1	Noise-vocoded
Condition 2	Novel sentences	Talker 1	Noise-vocoded
Condition 3	Sentences from Training Session 1, 2 or 3	Talker 2	Noise-vocoded
Condition 4	Novel sentences	Talker 2	Noise-vocoded

Table 1: Various testing conditions present in the test session used to assess the effects of manipulation of the variables of Talker Familiarity and noise-vocoded Speech familiarity.

When assessing the effects of the two independent variables (Talker Familiarity and noise-vocoded speech familiarity), the focus of this portion of the overall study will be on evaluating the differences in performance of noise-vocoded speech recognition between individuals who were trained on familiar versus unfamiliar talkers. To assess the effectiveness of how the two independent variables interact, four groups will be created: Group A, Group B, Group C, and Group D (Table 2).

Training Groups

Group	Talker	Voice
A	Talker 1	Vocoded
B	Talker 2	Vocoded
C (Control)	Talker 1	Natural
D (Control)	Talker 2	Natural

Table 2: Various groups that will be created to assess the effects of manipulation of the variables of talker familiarity and noise-vocoded speech familiarity. Individuals will be randomly assigned to each of the above groups.

Group A will be trained with Talker 1 utilizing voice-vocoded stimuli across the training sessions. Group B will be trained with Talker 2 utilizing voice-vocoded stimuli across the training sessions. Groups C and D will serve two purposes. First, they will be utilized to investigate the familiarity benefit in the training of non-vocoded speech across the three Trials for Talker 1 and Talker 2. Second, to address concerns regarding distinguishing between improvements in speech perception caused by the training itself (training-related learning) from improvements caused through repeated exposure to the testing materials and procedure (procedural learning), two control groups will be included as well (Stacey et al., 2010). Group C will be trained with Talker 1 with natural voice across training sessions, and will serve as the first control, and Group D will be trained with Talker 2 with natural voice across training sessions, serving as the second control.

4.3 Stimuli

For each of the three training sessions, 50 BKB sentences will be administered in a randomized order. This number was selected to ensure there is enough data to analyze within-training effects. Additionally, these sentences in particular were chosen due to being an effective and appropriate stimulus for adults (Raman et al., 2001). The BKB sentence battery includes 21 lists of non-recurring sentences that contain 16 sentences per list, and 50 keywords (Bench et al., 1979). The final test session will consist of sentences participants had been trained on, as well as novel sentences that they had not heard before. Each sentence has a differing amount of target words, ranging from three to five. Participants will be prompted to input the target words in the Gorilla Experiment Builder software, and scoring of these words is done automatically within the software. Participants will type the target words with no spacing and will be separated using a comma (for example: dog,ate,food).

The BKB sentences were produced by two talkers of the same gender, within the same age-range to reduce potential confounds. Talkers A and B recorded the stimuli in a sound-booth at the CUNY Graduate Center using a Neumann KMS 106 condenser microphone and pop filter on a tripod boom microphone stand with a Zoom F8 field recorder connected by a Mogami XLR cable. Each of the 16 sentences from the 21 sentence lists for BKB sentences were recorded three times. Two scorers listened to all three recordings of the 336 sentences independently and rated the best utterance out of the three. The criteria included choosing an utterance which sounded natural but with the least variability in tonality and intonation. In total, 300 sentences were chosen to be used by each of the two talkers. After this, a 50Hz high-pass filter was applied to each of the sentences to eliminate unwanted noise. RMS was used to equalize the intensity of each sound file to have an average intensity of 60 dB SPL via Praat. These sound files were used as the natural talker stimuli for the experiment and were also used to create the noise-vocoded sentences.

Shannon et al. (1995) developed the protocol used to create noise-vocoded stimuli. Spectral information from each sentence is removed and replaced by frequency-specific information within a broad frequency region with a band-limited noise. Then, the signal is divided into several frequency bands, and the amplitude envelope from the signal is extracted from each band using high-wave rectification and low-pass filtering. Literature supports the use of 6 to 10 frequency bands to vocode (Davis et al., 2006; Roberts et al., 2010; Sheldon et al., 2008; Stacey & Summerfield, 2007); 8 frequency bands were chosen to vocode. Low-pass filters with cutoff frequencies of 16Hz, 50Hz, 160Hz, and 600Hz can be used for extracting the amplitude envelope (Sheldon et al., 1995); this study chose a cutoff frequency of 50Hz. The envelope signal was used to modulate spectrally limited white noise by the same bandpass filter that was used in the original analysis band. Finally, all bands were summed and were used to create the final noise-vocoded sentence.

Based on the above parameters, noise-vocoded sentences were created following similar guidelines within AngelSim: Cochlear Implant and Hearing Loss Simulator, V1.08.01. Appendix C outlines the parameters that were applied to each sentence for vocoding.

4.4 Independent and Dependent Variables

The independent variables in this study are as follows: Talker Familiarity (two levels: familiar or unfamiliar speaker), Type of Speech (two levels: noise-vocoded or natural), Training Session (three levels: one through three), and Trial (1-50).

The dependent variable is the Speech Recognition performance.

To control confounding variables, participants will be assigned randomly to each group. Additionally, the questionnaire will be used as a screener to ensure overall homogeneity of the participants. Environmental factors will be controlled remotely by instructing the participants to complete each session in the same room each day and ensure the environment is quiet. They will be reminded to complete the entire experiment using the same equipment and set-up each time. The Gorilla Experimental Software is useful in that it records the type of computer that each participant accesses the study from. Additionally, the Gorilla Experimental Software was chosen in part to being compatible with a variety of computer operating systems, therefore it is anticipated that most participants recruited will be able to complete the training with their devices.

4.5 Additional Conditions

Upon completion of the study, each participant will receive monetary compensation.

5. Data Analysis

5.1 Within-Subjects Data

Each participant will be scored on how many of the 250 target words they got correct in each training session (Training Sessions 1, 2, or 3), as well as how many of the target words they get correct in the test session. To properly assess between-training effects, it will be necessary to obtain each participant's percentage correct of target words for each session and the means of those scores. It is also of note to calculate their percentage correct on each condition of the test session, and their overall percentage correct of all target words during the test session. To assess within-training effects, the scores of target words identified correctly will be assessed at the beginning vs. the end of the same training session. Finally, age, gender, and training session/ test session time will also be considered in the analysis as well.

5.2 Between-Subjects Data

Between-subject data analysis will be done to analyze trends seen in each of the four groups across the training sessions (Group A: trained on the Talker 1 with voice vocoded; Group B: trained on Talker 2 voice vocoded; Group C: trained on Talker 1 with normal voice; Group D: trained on Talker 2 with normal voice). Additionally, it is of particular interest to see which group performed the best overall during the final test session. Analyzing these trends is beneficial in understanding the most ideal conditions that led to maximal perceptual learning.

5.3 Statistical Analysis

The scores for each participant will be recorded for each session. Additionally, the mean of each of the three training sessions will be calculated for each of the groups (A-D). The mean scores for Sessions 1 through 3 will be compared to the performance on the test session for each participant, and between groups. ANOVA will be used to analyze the data to see if there are any main effects as well as possible interactions between the main effects of Type of Speech (natural versus vocoded), Talker (familiar versus unfamiliar), Trial, and Session. ANCOVA will be conducted to investigate potential contributors to individual variability. Finally, paired sample t-tests will be used to ensure similarities between each group as per inclusion criteria, and a post-hoc test for correction will be used as well.

6. Methodological Summary

6.1 Anticipated Results

It is expected that participants within Group A (listening to Talker 1 with noise-vocoded speech during training) will show improvements incorrectly identifying noise-vocoded target words across Training Sessions 1 to 3, as well as during test session compared to their performance at the beginning of training. It is also hypothesized that these participants will demonstrate an increased identification of target words within noise-vocoded sentences when hearing Talker 1, the familiar talker, than when presented with sentences spoken by an unfamiliar talker due to the familiarity benefit. The effects of being trained with the familiar talker will also result in more correct target words identified within new vocoded sentence stimuli spoken by Talker 1 within the test session than when present those participants are presented with new vocoded sentence stimuli spoken by Talker 2.

Similar findings are expected for participants within Group B (Talker 2 with noise-vocoded speech). These participants will show improvements in identifying noise-vocoded target words within sentences in performance across Training Sessions 1 to 3, as well as within the test session. Also, these participants will demonstrate an increased identification of target words within noise-vocoded sentences when hearing Talker 2, the familiar talker, than when presented with sentences spoken by an unfamiliar talker, due to the familiarity benefit. The effects of being trained with the familiar talker will result in more correct target words identified within new vocoded sentence stimuli spoken by Talker 2 within the test session than when present those participants are presented with new vocoded sentence stimuli spoken by Talker 1.

Regarding participants in Group C (Talker 1 with natural voice), we expect performance of correctly identifying target words of noise-vocoded speech to be poorer than the performance of Groups A and B, who were trained on noise-vocoded speech during the test session. We do expect, however, that performance will be slightly better when presented with sentences spoken by the familiar talker (Talker 1) than with the unfamiliar talker (Talker 2). This effect of Talker Familiarity will be a smaller effect due to the nature of this group's training. Previous exposure was only to the natural voice of the familiar talker, so performance with a degraded signal, despite being the same talker, will impact the identification of target words.

For participants in Group D (Talker 2 with natural voice), we expect performance of identifying correct target words of noise-vocoded speech to be similar to those within Group C. Group D will perform poorer on the speech recognition task compared to Groups A and B, who were trained on noise-vocoded speech during the test session. We do expect, however, that performance will be slightly better when presented with sentences spoken by the familiar talker (Talker 2) than with the unfamiliar talker (Talker 1). Similar to Group C, the effect of Talker Familiarity will be a smaller effect due to the nature of this group's training, and lack of exposure to noise-vocoded sentences prior to the final test session.

For our control groups, Group C and D, having participants complete the training procedure will result in procedural learning of the task. Thus, their performance can be compared to the performance of the other participants within training Groups A and C, reflecting any stimuli-specific perceptual learning.

One final prediction that will be seen in all conditions that look at the interaction between familiar versus unfamiliar talker is that literature shows that those trained on a familiar talker and are assessed on sentences spoken by said familiar talker, the participants will be faster at identifying correct target words. Jesse, McQueen & Page (2007) found that listeners are faster at tasks of speech perception when exposed to the same talker producing targeted sounds, which may be particularly beneficial in auditory training efficiency.

6.2 Implications and Significance

This study may assist in understanding the learning processes involved in noise-vocoded speech perception training, as well as the effect talker familiarity may have as an assisting cue. As outlined earlier, those with CIs are tasked with interpreting and making meaning of novel sounds, thus requiring re-training of the modified auditory system (Barker, 2006; Castiglione et al., 2016; Reis et al., 2021; Tye-Murray, Witt, & Schum, 1995). While there are various programs that are utilized in this population to train speech perception skills, there has yet to be a study investigating the effects of the training of noise-vocoded, CI-like speech and talker familiarity. Literature also lacks studies investigating if the familiarity benefit withstands the processing found within noise-vocoded stimuli, which emulates what CI wearers hear. Literature supports that those with hearing loss across the age span perform worse on tasks of talker identification in noise and quiet (Best et al., 2018). Those with more severe hearing losses, like CI wearers, who experience even more widespread damage and distortion of the auditory system have overall reduced intelligibility (Ching &Dillon, 2013), suggesting that indexical talker information within the distorted system is impacted (Barker, 2006). Thus, investigating the effect of CI-like noise vocoded speech in relation to talker familiarity first in a population of NH listeners is of interest. The findings could imply as to the effectiveness of training patients pre-CI implantation to perceive certain voices (Barker, 2006).

Additionally, studies have been conducted utilizing auditory training in-person or via computer-based programs, and both methods have been shown to be effective in enhancing this skillset (Cosby, 2011; Gil & Iorio, 2010; Henshaw & Ferguson, 2013; Stropahl et al., 2020; Stacey et al., 2010; Stacey & Summerfield, 2007). However, there is still a need for more literature of high-quality evidence that examines the efficacy of a computer-based auditory training for those with hearing loss (Henshaw & Furgoson, 2013). If this study is conducted remotely, it would provide additional evidence regarding the effectiveness of online training.

Of note, it would also be of use to study the interaction of noise-vocoded speech and talker familiarity by training participants with multiple familiar talkers instead of one. Literature has found that 'high-variability' training, or training with several, unfamiliar talkers, is more effective than training with just one talker (Barker, 2006; Humes et al., 2006; Kreitewolf et al., 2017; Olson, 2015; Stacey and Summerfield 2007). A particularly large advantage when training with multiple unfamiliar talkers is that it may allow listeners to separate talker-specific and lexically specific information. This is important because these are confounded in studies that only use one talker for training (Stacey & Summerfield, 2007). High-variability training with unfamiliar talkers has even been shown to be effective in training non-native speakers to perceive speech of different languages (Bradlow & Bent, 2003; Bradlow et al., 1997; Clopper & Pisoni, 2004; Lively et al., 1993, 1994; Logan et al., 1991; Wang et al., 1999). Additionally, high-variability training is more generalizable to new talkers (Lively et al., 1993). As opposed to the increased speed with speech perception tasks when trained and tested with one talker, Mullennix, Pisoni, and Martin (1989) found that when listeners are presented with multiple unfamiliar talkers rather than a single familiar talker in a word recognition task, listeners are slower to identify words, which may be a disadvantage. Overall, Stacey and Summerfield (2007) highlight that training programs should contain materials spoken by multiple talkers to be the most effective in influencing speech perception, which is important to consider when developing novel training programs. It would be of interest to see if the speed in which participants identifying a signal with less indexical information is impacted when trained with multiple talkers, and if the familiarity benefit is impacted.

If a familiarity benefit enhances lexical access as predicted, it would be advantageous to see the amount of time it takes for participants to acquire said benefit. Domingo et al. (2020) investigated the presence of talker familiarity in relation to the target speaker, the target speaker being a spouse or a friend. The study showed while the benefit was not influenced by the length of relationship, which spanned from as little as 1.5 years and up to 52 years in length, the familiarity benefit reached a plateau at around 1.5 years. In relation to the training of speech perception in the scope of an auditory training protocol, the speed in which a participant improves speech perception measures is an important factor in the efficiency of the program. Thus, if talker familiarity is noted when training on noise-vocoded speech and demonstrates improvement in speech perception tasks, as well when utilizing novel stimuli, it would be interesting to investigate how soon that familiarity effect is encoded within memory.

Overall, this training study will help in understanding the underlying processes that are involved in learning a new acoustical input through auditory training. As outlined in the literature, it is a useful tool in assisting those learning how to interpret novel acoustical input. Finally, this study may help elucidate, through the manipulation of talker familiarity and manipulation of noise-vocoded familiarity, understanding of the underlying higher-level neuronal and learning mechanisms that are involved in learning speech.

7. Appendixes

7.1 Appendix A: Questionnaire

All participants, regardless if the procedure is conducted in person or remotely, will receive the following sample questionnaire. The purpose of including a questionnaire is to ensure all participants are as similar as possible and to decrease individual variability as much as possible.

(Provide general description of experiment here)
(Contact information i.e. preferred email address/phone number)

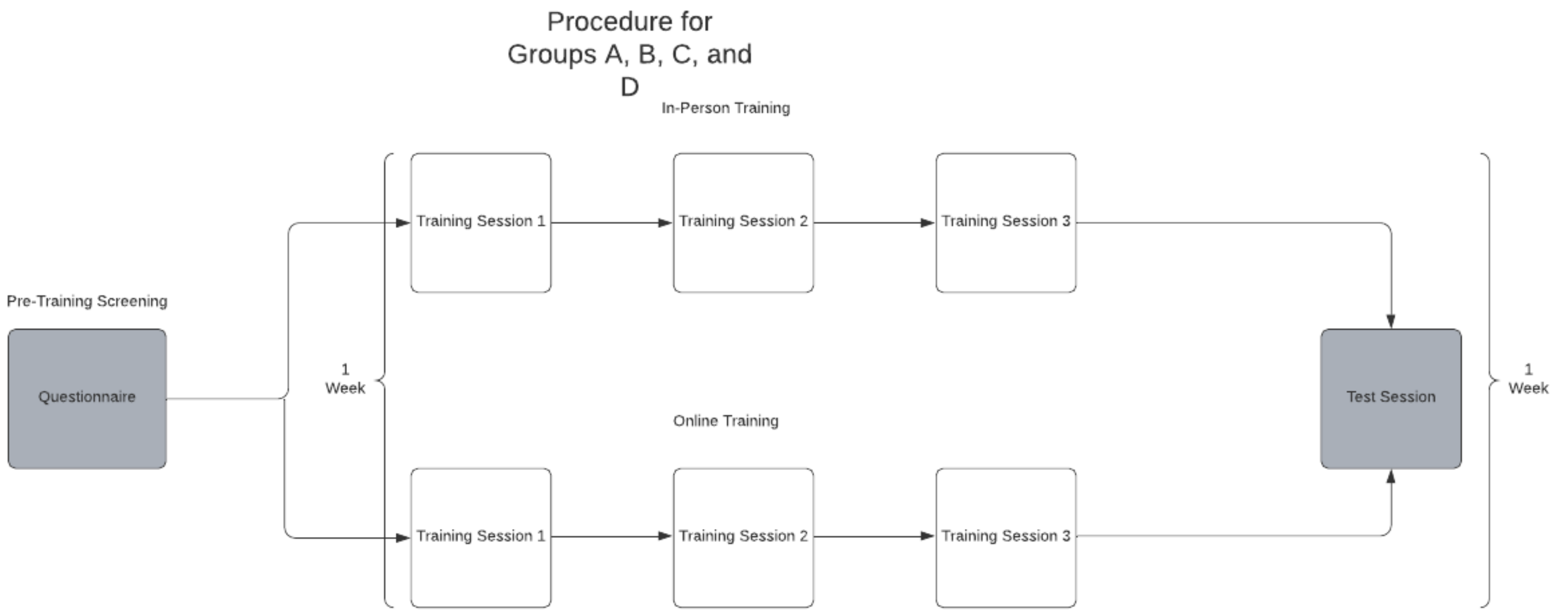
Name:

1. Please provide your age: ___
2. Race: ___
3. Gender: ___
4. Preferred language: ___
5. Highest level of education completed:
    a. High school (GED)
    b. College (B.A./B.S.)
    c. Post-baccalaureate (Master's/Doctoral)
6. Approximate yearly income:
    a. $0-$20,000
    b. $21,000-$40,000
    c. $41,000-$60,000
    d. $61,000 +
7. Have you been diagnosed with hearing loss/suspect you have hearing loss?
    a. No
    b. Yes, please elaborate: _____
8. Do you have a history of ear infections?
    a. Yes
    b. No
9. Have you ever seen a speech-language pathologist or received language services?
    a. No
    b. Yes, please elaborate: _____
10. Do you or have you ever received services for any conditions related to cognition, attention, or memory?
    a. No
    b. Yes, please elaborate: _____
11. Do you have a family history of hearing loss?

a. No
b. Yes, please elaborate: _____

12. Do you have any history of trauma to the head?
    a. No
    b. Yes, please elaborate: _____
13. Do you experience dizziness, vertigo, or a loss of balance?
    a. No
    b. Yes, please elaborate: _____
14. Do you have tinnitus (ringing, buzzing, hissing) in your ear(s)?
    a. No
    b. Yes, please elaborate and specify (how frequent, which ear, is it provoked by certain situations?): _____
15. Do you have a history of noise exposure? Such as (but not limited to) military, occupational, recreational (such as loud music exposure)?
    a. No
    b. Yes, please elaborate: _____
16. Do you listen to music via headphones or earbuds?
    a. No
    b. Yes, please elaborate:
        i. Average hours a day? ____
        ii. At what level approximately?
            1. 25% full volume
            2. 50% full volume
            3. 75% full volume
            4. 100% full volume

7.2 Appendix B: *Procedure for Training*



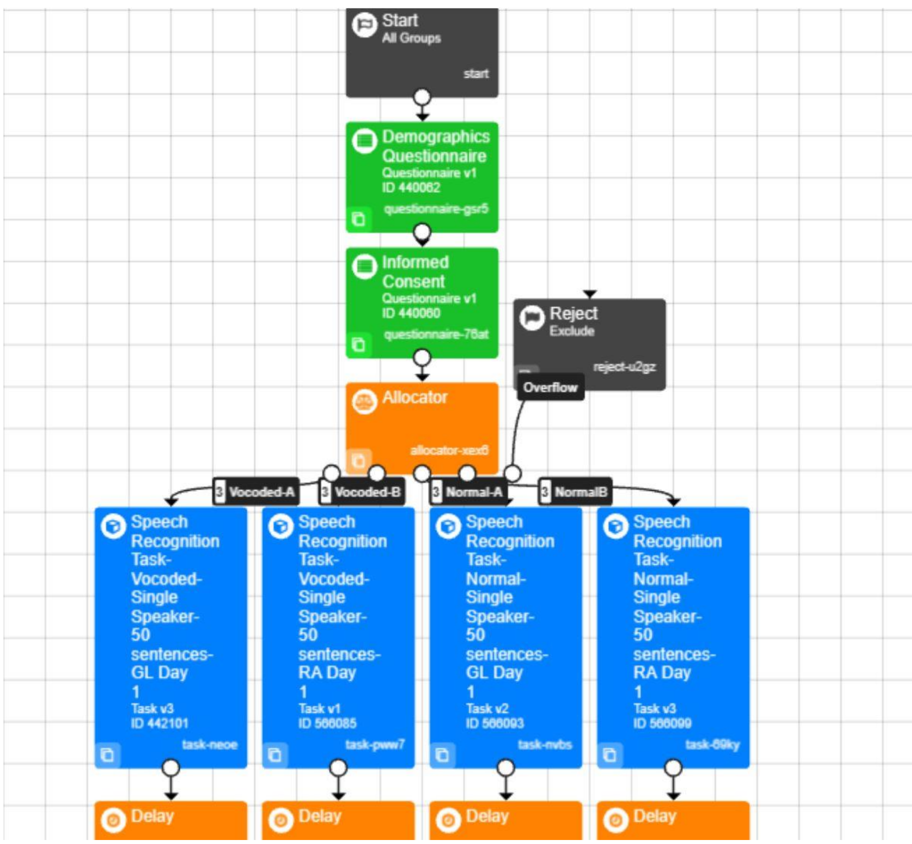
7.3 Appendix C: *Procedure for Vocoding Sentences*

The following parameters in the following chronological order were applied to those sentences that were vocoded. Vocoding was applied based on the parameters described by Fitzgerald et al. (2013) using AngelSim: Cochlear Implant and Hearing Loss Simulator,V1.08.01. Each of the processed sentences after high pass filter was applied via Audacity was downloaded and vocoded using the following steps:

1. The processor selected was a Noise/Sinewave Vocoder
2. The carrier type for the vocoder was selected as a White noise carrier (noise vocoder)
3. The number of channels stimulated and the number of spectral channels was both input to be 8 channels (Davis et al., 2006; Roberts et al., 2010; Sheldon et al., 2008; Stacey & Summerfield, 2007)
4. The analysis filter type set to Greenwood function
5. The Analysis Filter Type was selected to be a frequency range of 200-7000Hz and 24dB per octave filter slope for analysis filter
6. The envelope detection cutoff frequency was selected to be 160Hz and 24dB per octave filter slope
7. Carrier filter type was set to Greenwood function
8. The Carrier Filter Design was selected to be a frequency range of 200-7000Hz and 24dB per octave filter slope for carrier filter

7.4 Appendix D: *Gorilla Protocol*

The training program below was created using Gorilla Experiment Builder Program. Participants will be sent a link through which they can access the study after completing a demographics questionnaire and informed consent. Participants will be randomly assigned one of four groups and will then complete three training sessions. If more than twenty-four participants are recruited, they will be sent to the "reject" branch. The blue rectangles represent training sessions, and the orange boxes below represent twenty-four hour "delays". Not shown are the remaining two training sessions and delay boxes for each of the four groups, as well as the final test session.



8. References

Barker, B. A. (2006). *An examination of the effect of talker familiarity on the sentence recognition skills of cochlear implant users* [Doctor of Philosophy, University of Iowa]. https://doi.org/10.17077/etd.9ogph1cp

Barcroft, Joe et al. "Tailoring Auditory Training to Patient Needs with Single and Multiple Talkers: Transfer-Appropriate Gains on a Four-Choice Discrimination Test." *International journal of audiology* 50.11 (2011): 802–808. Web.

Best, Virginia et al. "Talker Identification: Effects of Masking, Hearing Loss, and Age." *The Journal of the Acoustical Society of America* 143.2 (2018): 1085–1092. Web.

Bench, J., Kowal, Å., & Bamford, J. (1979). The Bkb (Bamford-Kowal-Bench) Sentence Lists for Partially-Hearing Children. *British Journal of Audiology*, *13*(3), 108–112. https://doi.org/10.3109/03005367909078884

Bent, T., & Bradlow, A. R. (2003). The interlanguage speech intelligibility benefit. *journal of the Acoustical Society Wof America*, *114*(3), 1600-1610. https://doi.org/10.1121/1.1603234

Blamey P, Artieres F, Baskent D, et al. Factors affecting auditory performance of postlingually deaf adults using cochlear implants: an update with 2251 patients. *Audiol Neurotol.* 2013;18(1):36-47

Boothroyd, Arthur. "Adapting to Changed Hearing: The Potential Role of Formal Training." *Journal of the American Academy of Audiology* 21.9 (2010): 601–611. Web.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /O/ and /I/: Some effects of perceptual learning on speech production. Journal of the Acoustical Society of America, 101, 2299-2310.

Castiglione, A., Benatti, A., Velardita, C., Favaro, D., Padoan, E., Severi, D., Pagliaro, M., Bovo, R., Vallesi, A., Gabelli, C., & Martini, A. (2016). Aging, Cognitive Decline and Hearing Loss: Effects of Auditory Rehabilitation and Training with Hearing Aids and Cochlear Implants on Cognitive Function and Depression among Older Adults. *Audiology and Neurotology*, *21*(Suppl. 1), 21–28. https://doi.org/10.1159/000448350

Ching, Teresa Y C, and Harvey Dillon. “A Brief Overview of Factors Affecting Speech Intelligibility of People with Hearing Loss: Implications for Amplification.” *American journal of audiology* 22.2 (2013): 306–309. Web.

Clopper, C. G., & Pisoni, D. B. (2004). Homebodies and army brats: Some effects of early linguistic experience and residential history on dialect categorization. Language Variation and Change, 16(1), 31–48.

Cosby, J. L. (2011). *Impact of Auditory Training on Speech Perception and Cognitive Abilities in Older Adults with Hearing Loss*. https://thescholarship.ecu.edu/handle/10342/3563

Davis, M. (2003). An Introduction to Noise-Vocoded Speech Matt Davis. Retrieved December 03, 2020, from http://www.mrc-cbu.cam.ac.uk/personal/matt.davis/vocode/

Davis, M. H., & Johnsrude, I. S. (2003). Hierarchical processing in spoken language comprehension. Journal of Neuroscience, 23, 3423–3431.

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., Taylor, K., & McGettigan, C. (2005). Lexical Information Drives Perceptual Learning of Distorted Speech: Evidence From the Comprehension of Noise-Vocoded Sentences. *Journal of Experimental Psychology: General*, *134*(2), 222–241. https://doi.org/10.1037/0096-3445.134.2.222

Dick, F., Tierney, A.T., Lutti, A., Josephs, O. Sereno, M.I., & Weiskopf, N. (2012). In vivo functional and myeloarchitectonic mapping of human primary auditory areas. *Journal of Neuroscience, 32,* 16095–16105.

Domingo, Y., Holmes, E., & Johnsrude, I. S. (2020). The benefit to speech intelligibility of hearing a familiar voice. Journal of Experimental Psychology: Applied, 26(2), 236–247. https://doi.org/ 10.1037/xap0000247

Drouin, Julia R., and Rachel M. Theodore. "Leveraging Interdisciplinary Perspectives to Optimize Auditory Training for Cochlear Implant Users." *Language and linguistics compass* 14.9 (2020): n. pag. Web.

Faulkner, A., Rosen, S., & Green, T. (2012). Comparing live to recorded speech in training the perception of spectrally shifted noise-vocoded speech. *The Journal of the Acoustical Society of America*, *132*(4), EL336–EL342. https://doi.org/10.1121/1.4754432

Fitzgerald, M. B., Sagi, E., Morbiwala, T. A., Tan, C. T., & Svirsky, M. A. (2013). Feasibility of real-time selection of frequency tables in an acoustic simulation of a cochlear implant. *Ear and hearing*, *34*(6), 763–772. https://doi.org/10.1097/AUD.0b013e3182967534

Friesen L. M., Shannon R. V., Baskent D., Wang X. (2001). Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants. *The Journal of the Acoustical Society of America*, 110(2), 1150–1163. https://doi-org.ezproxy.gc.cuny.edu/10.1121/1.1381538

Fu, Q.-J., Galvin, J., Wang, X., & Nogaki, G. (2005). Moderate auditory training can improve speech performance of adult cochlear implant patients. *Acoustics Research Letters Online*, *6*(3), 106–111. https://doi.org/10.1121/1.1898345

Gerstman, L. J. (1968). Classification of self-normalized vowels. *IEEE Transactions on Audio Electroacoustics, AU-16*, 78–80.

Gil, D., & Iorio, M. C. M. (2010). Formal auditory training in adult hearing aid users. *Clinics*, *65*(2), 165–174. https://doi.org/10.1590/S1807-59322010000200008

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105*(2), 251–279. https://doi.org/10.1037/0033-295X.105.2.251

Greenwood, D. D. (1990). A cochlear frequency-position function for several species—29 years later. *The Journal of the Acoustical Society of America*, *87*(6), 2592–2605. https://doi.org/10.1121/1.399052

Hackett, T.A. (2007). Organization and correspondence of the auditory cortex of humans and nonhuman primates. In J.H. Kass (Ed.), *Evolution of the nervous system*, (pp 109 –119). Oxford, UK: Elsevier.

Haile, Lydia M et al. "Hearing Loss Prevalence and Years Lived with Disability, 1990–2019: Findings from the Global Burden of Disease Study 2019." *The Lancet (British edition)* 397.10278 (2021): 996–1009. Web.

Henshaw, H., & Ferguson, M. A. (2013). Efficacy of Individual Computer-Based Auditory Training for People with Hearing Loss: A Systematic Review of the Evidence. *PLOS ONE*, *8*(5), e62836. https://doi.org/10.1371/journal.pone.0062836

Humes, L. E., Burk, M. H., Strauser, L. E., & Kinney, D. L. (2009). Development and efficacy of a frequent-word auditory training protocol for older adults with impaired hearing. *Ear and hearing*, *30*(5), 613–627. https://doi.org/10.1097/AUD.0b013e3181b00d90

Jesse, A., McQueen, J. M. & Page, M. (2007). The locus of talker-specific effects in spoken-word recognition. In Trouvain, J. & Barry, W. J. (eds), International Congress of Phonetic Sciences (ICPhS 2007), 1921–4. Dudweiler: Pirrot.

Johnson, K. (1994). Memory for vowel exemplars. *Journal of the Acoustical Society of America, 95*, 2977.

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In K. Johnson & J. W. Mullennix (Eds.), *Talker Variability in Speech Processing* (pp. 145–166). San Diego: Academic Press.

Johnson, K. (2005). Speaker normalization in speech perception. In D.B. Pisoni & R. Remez (Eds.), *The Handbook of Speech Perception*. Oxford: Blackwell Publishers. pp. 363–389.

Kiessling J, Pichora-Fuller MK, Gatehouse S, et al. (2003). Candidature for and delivery of audiological services: Special needs of older people. Int J Audiol 42:2S92-2S101.

Kreitewolf, J., Mathias, S. R., & von Kriegstein, K. (2017). Implicit Talker Training Improves Comprehension of Auditory Speech in Noise. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.01584

Kuwabara, H., & Takagi, T. (1991). Acoustic parameters of voice individuality and voice-quality control by analysis-synthesis method. *Speech Commun.*, *10*, 491-495.

Ladefoged, P., and Broadbent, D. E. (1957). Information conveyed by vowels. *Journal of the Acoustical Society of America, 29*, 98–104.

Levi, S. V., Winters, S. J. & Pisoni, D. B. (2011). Effects of cross-language voice training on speech perception: Whose familiar voices are more intelligible? Journal of the Acoustical Society of America 130(6), 4053–62.

Linville, Sue Ellen, and Hilda B Fisher. “Acoustic Characteristics of Perceived Versus Actual Vocal Age in Controlled Phonation by Adult Females.” The Journal of the Acoustical Society of America 78.1 (1985): 40–48. Web.

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/: II. The role of phonetic environment and talker variability in learning new perceptual categories. Journal of the Acoustical Society of America, 94, 1242-1255.

Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y., & Yamada, T. (1994). Training Japanese listeners to identify English /r/ and /l/: III. Long-term retention of new phonetic categories. Journal of the Acoustical Society of America, 96, 2076-2087.

Loizou, P. C., Dorman, M., & Tu, Z. (1999). On the number of channels needed to understand speech. Journal of the Acoustical Society of America, 106(Pt. 1), 2097–2103.

Magnuson, James S. et al. “Talker Familiarity and the Accommodation of Talker Variability.” *Attention, perception & psychophysics* 83.4 (2021): 1842–1860. Web.

McGettigan, C., Rosen, S., & Scott, S. K. (2014). Lexico-semantic and acoustic-phonetic processes in the perception of noise-vocoded speech: Implications for cochlear implantation. *Frontiers in Systems Neuroscience, 8*. https://doi.org/10.3389/fnsys.2014.00018

McLaughlin, D.E., Carter, Y.D., Cheng, C.C. *et al.* Hierarchical contributions of linguistic knowledge to talker identification: Phonological versus lexical familiarity. *Atten Percept Psychophys* 81, 1088–1107 (2019). https://doi-org.ezproxy.gc.cuny.edu/10.3758/s13414-019-01778-5

Miller, J. D. (1989). Auditory-perceptual interpretation of the vowel. *Journal of the Acoustical Society of America, 85,* 2114–2134.

Mullennix, J. W., Pisoni, D. B., & Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. Journal of the Acoustical Society of America, 85, 365-378.

Murray, Iain R, and John L Arnott. “Toward the Simulation of Emotion in Synthetic Speech: A Review of the Literature on Human Vocal Emotion.” *The Journal of the Acoustical Society of America* 93.2 (1993): 1097–1108. Web.

Nearey, T. M. (1989). Static, dynamic, and relational properties in vowel perception. *Journal of the Acoustical Society of America, 85,* 2088–2113.

Nusbaum, H. C., & Magnuson, J. S. (1997). Talker normalization: Phonetic constancy as a cognitive process. In K. Johnson & J. W. Mullennix (Eds.), Talker variability in speech processing (pp. 109 –132). San Diego, CA: Academic Press.

Nusbaum, H. C., & Morin, T. M. (1992). Paying attention to differences among talkers. In Y. Tohkura, Y. Sagisaka, & E. Vatikiotis-Bateson (Eds.), *Speech Perception, Speech Production, and Linguistic Structure*, pp. 113–134. Tokyo: OHM.

Nusbaum, H. C., & Schwab, E. C. (1986). The role of attention and active processing in speech perception. In E. C. Schwab & H. C. Nusbaum (Eds.), Pattern recognition by humans and machines: Vol. 1. Speech perception (pp. 113–157). San Diego, CA: Academic Press.

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science, 5*(1), 42–46. https://doi.org/10.1111/j.1467-9280.1994.tb00612.x

Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics, 60*(3), 355–376. https://doi.org/10.3758/BF03206860

Olson, A. D. (2015). Options for Auditory Training for Adults with Hearing Loss. *Seminars in Hearing, 36*(4), 284–295. https://doi.org/10.1055/s-0035-1564461

Pichora-Fuller, M. Kathleen, and Gurjit Singh. "Effects of Age on Auditory and Cognitive Processing: Implications for Hearing Aid Fitting and Audiologic Rehabilitation." *Trends in amplification* 10.1 (2006): 29–59. Web.

Potter, R., & Steinberg, J. (1950). Toward the specification of speech. *Journal of the Acoustical Society of America, 22*, 807–820.

Pierrehumbert, J. (2002) Word-specific phonetics. In C. Gussenhoven and N. Warner (Eds.), *Laboratory Phonology 7*, pp. 101–139. Berlin: Mouton de Gruyter.

Rakerd, B. & Verbrugge, R. R. (1987). Evidence that the dynamics information for vowels is talker independent in form. *Journal of Memory and Language, 26*,558–563.

Raman G, Lee J, Chung M, et al. Effectiveness of Cochlear Implants in Adults with Sensorineural Hearing Loss. Agency for Healthcare Research and Quality (US), Rockville (MD); 2011.

Reis M, McMahon CM, Távora-Vieira D, Humburg P, Boisvert I. Effectiveness of Computer-Based Auditory Training for Adult Cochlear Implant Users: A Randomized Crossover Study. *Trends in Hearing*. 2021;25. doi:10.1177/23312165211025938

Rubinstein, A., & Boothroyd, A. (1987). Effect of two approaches to auditory training on speech recognition by hearing-impaired adults. *Journal of speech and hearing research*, *30*(2), 153–160. https://doi.org/10.1044/jshr.3002.153

Rudner M. , R ö nnberg J. & Lunner T . 2011 . Working memory supports listening in noise for persons with hearing impairment . J Am Acad Audiol , 22 , 156 – 167 .

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech Recognition with Primarily Temporal Cues. *Science*, *270*(5234), 303–304. https://doi.org/10.1126/science.270.5234.303

Sheldon, S., Pichora-Fuller, M. K., & Schneider, B. A. (2008). Priming and sentence context support listening to noise-vocoded speech by younger and older adults.

The Journal of the Acoustical Society of America, *123*(1), 489–499. https://doi.org/10.1121/1.2783762

Sommers, M. S., Nygaard, L. C., & Pisoni, D. B. (1994). Stimulus variability and spoken word recognition. I. Effects of variability in speaking rate and overall amplitude. Journal of the Acoustical Society of America, 96, 1314-1324.

Stacey P. C., Raine CH, O'Donoghue GM, Tapper L, Twomey T, & Summerfield AQ. (2010). Effectiveness of computer-based auditory training for adult users of cochlear implants. *International Journal of Audiology*, *49*(5), 347–356. https://doi.org/10.3109/14992020903397838

Stacey, P. C., & Summerfield, A. Q. (2007). Effectiveness of computer-based auditory training in improving the perception of noise-vocoded speech. *The Journal of the Acoustical Society of America*, *121*(5 Pt1), 2923–2935. https://doi.org/10.1121/1.2713668

Stropahl, M., Besser, J., & Launer, S. (2020). Auditory Training Supports Auditory Rehabilitation: A State-of-the-Art Review. *Ear and Hearing*, *41*(4), 697–704. https://doi.org/10.1097/AUD.0000000000000806

Strange, W. (1989). Dynamic specification of coarticulated vowels spoken in sentence context. *Journal of the Acoustical Society of America*, *85*, 2135–2153.

Syrdal, A. K. and Gopal, H. S. (1986). A perceptual model of vowel recognition based on the auditory representation of American English vowels. *Journal of the Acoustical Society of America*, *79*, 1086–1100.

Traunmuller, H. (1981). Perceptual dimension of openness in vowels. *Journal of the Acoustical Society of America*, *69*, 1465–1475.

Tye-Murray, Nancy et al. “Auditory Training for Adults Who Have Hearing Loss: A Comparison of Spaced Versus Massed Practice Schedules.” *Journal of speech, language, and hearing research* 60.8 (2017): 2337–2345. Web.

Tye-Murray, N., Witt, S., & Schum, L. (1995). Effects of Talker Familiarity on Communication Breakdown in Conversations with Adult Cochlear-Implant Users. *Ear & Hearing*, *16*(5), 459–469.

Wilson, B.S. (2004). Engineering Design of Cochlear Implants. In: Zeng, FG., Popper, A.N., Fay, R.R. (eds) Cochlear Implants: Auditory Prostheses and Electric Hearing. Springer Handbook of Auditory Research, vol 20. Springer, New York, NY. https://doi.org/10.1007/978-0-387-22585-2_2

Woods, D. L, Stecker, G.C., Rinne T,. Herron T.J., Cate, A.D., Yund, E.W., Liao, I., & Kang, X. (2009). Functional maps of human auditory cortex: Effects of acoustic features and attention. PLoS One 4:e5183.

Wong, P.C.M., Nusbaum, H.C., & Small, S.L. (2004). Neural bases of talker normalization. *Journal of Cognitive Neuroscience, 16*, 1173–1184.

Yonan, C. A., & Sommers, M. S. (2000). The effects of talker familiarity on spoken word identification in younger and older listeners. Psychology and Aging, 15(1), 88–99. https://doi.org/10.1037/0882-7974.15.1.88