August 2015

The Effect of Sleep on Perceptual Learning and Memory Consolidation

Vanessa Claire Irsik
University of Nevada, Las Vegas, irsikv@unlv.nevada.edu

THE EFFECT OF SLEEP ON PERCEPTUAL LEARNING AND MEMORY CONSOLIDATION

By

Vanessa C. Irsik

Bachelor of Arts in Psychology
University of Nevada, Las Vegas
2012

A thesis submitted in partial fulfillment
of the requirements for the

Master of Arts - Psychology

Department of Psychology
College of Liberal Arts
The Graduate College

University of Nevada, Las Vegas
August 2015

UNLV | GRADUATE COLLEGE

We recommend the thesis prepared under our supervision by

Vanessa C. Irsik

entitled

The Effect of Sleep on Perceptual Learning and Memory Consolidation

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

Master of Arts - Psychology
Department of Psychology

Joel Snyder, Ph.D., Committee Chair

Erin Hannon, Ph.D., Committee Member

Jefferson Kinney, Ph.D., Committee Member

Gabriele Wulf, Ph.D., Graduate College Representative

Kathryn Hausbeck Korgan, Ph.D., Interim Dean of the Graduate College

August 2015

ABSTRACT

The Effect of Sleep on Perceptual Learning and Memory Consolidation

by

Vanessa C Irsik

Dr. Joel S. Snyder, Examination Committee Chair

Associate Professor of Psychology

University of Nevada, Las Vegas

An ability to segregate speech accurately is essential given that most auditory environments contain other overlapping conversations or environmental noise. While perceiving speech among background noise can be difficult in and of itself, those with hearing impairments can experience considerable difficulty. While training has been shown to benefit perceptual segregation of trained sounds, it is unclear how such training transfers to sounds not included in a training regimen. The current study aimed to address this question by training listeners on a portion of sounds during a vowel segregation task, and subsequently testing on both the trained sounds and untrained sounds. Additionally, the dependency on sleep in consolidating generalization was investigated by testing listeners at two additional time points: before sleep (12 hours later) and after sleep (24 hours later). Finally, neural correlates specific to generalization was investigated by recording brain activity (EEG) during all test and training sessions. Trained listeners significantly improved on trained and untrained vowel pairs, demonstrating training-induced learning and generalization. The control group also significantly improved across test sessions, demonstrating testing-induced learning. Spatio-temporal analyses of EEG

data revealed that generalization learning was paralleled by a source configuration change, while rote learning was paralleled by a change in the power of the neural response. These results confirm that learning gained through speech segregation is generalizable to new sounds, as well as revealed a neural pattern of activity that may index the processes responsible for transferring learning to untrained sounds. Finally, time and additional practice appear to greater contribute to learning overall, as compared to sleep.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Processing speech is an important part of everyday life given that it provides a means of communication with others in our environment. In fact, obtaining an ability to do this is so fundamental that it is considered a major milestone during development. Despite the complicated steps our sensory system must carry out in order to process simple and complex sounds, most people can communicate with relative ease. This process involves first detecting the physical signal, then segmenting and identifying individual words or auditory objects from a continuous acoustic stream, and often segregating the message from other overlapping noises. Difficulty can arise when executing any one of these steps as well as from attempting to adjust to speaker variability. Measuring speech properties (e.g., fundamental frequency and formant patterns) reveal that a word spoken by two different speakers or even a word spoken by the same person twice can be markedly different, otherwise referred to as inter- and intra-speaker variability (Nygaard & Pisoni, 1998). Those suffering from hearing impairments as a result of injury or aging can experience exceptional difficulty in processing both simple and complex sounds. For example, cochlear implant (CI) users, stroke victims, and older adults all have difficulty with segregating concurrent speech or perceiving speech in noise, as well as can have issues with localizing sounds in their environment (Alain et al., 2006; Bamiou et al., 2012; Grieco-Calub, Saffran, & Litovsky, 2009; Ingvalson & Wong, 2013; Tessier, Weill-Chounlamountry, Michelot, & Pradat-Diehl, 2007). Understanding how healthy listeners learn to perceive speech signals in an effortless manner can perhaps shed some light on how to ameliorate the difficulties of

those who have impaired hearing, as well as help determine under what circumstances perceptual abilities can be improved in normal hearing individuals.

The literature on perceptual learning has grown immensely as a result of efforts to determine what conditions encourage learning and promote accurate speech perception. Training has been reported as effective in improving performance on a wide range of tasks such as word recognition/identification (Sebastian-Galles, Dupoux, Costa, & Mehler, 2000), frequency discrimination (Demany, 1985; Wright & Sabin, 2007), and perceiving speech in noise (Van Engen, 2012). Some factors that appear to facilitate learning include a training regimen containing highly variable stimuli (Samuel & Kraljic, 2009), use of visual cues paired with training stimuli (Bernstein, Auer, Eberhardt, & Jiang, 2013), and short training sessions lasting no longer than one hour (Molloy, Moore, Sohoglu, & Amitay, 2012). Research on the effect of natural experience on perception appears to suggest that exposure to a particular stimulus over time enhances perception and becomes a type of informal training. For example, those who have experience hearing the speech of the deaf score significantly higher than inexperienced listeners when tested on their ability to transcribe deaf speech (McGarr, 1983; Klimacka, Patterson, & Patterson, 2001). Musicians are another great example of the positive effects of training on perception. Becoming adept at playing an instrument requires consistent analysis of complex auditory signals. This necessitates close attention to be paid to acoustic features such as pitch and harmony, and often involves segregation of overlapping melodies (Zendel & Alain, 2009). Extended periods of practice seem to result in musicians possessing enhanced abilities in segregating concurrent complex sounds (Zendel & Alain, 2009), perceiving speech in noise (Parbery-Clark, Skoe, Lam, &

Kraus, 2009), and pitch discrimination (Tervaniemi, Just, Koelsch, Widmann, & Schroger, 2005), all of which demonstrate the powerful effect of training on perception.

It seems safe to conclude that training can be effective in improving perceptual acuity; however, it is not always clear what sort of learning underlies observed behavioral improvements. Although there are several subtypes of perceptual learning, only two will be focused on here, namely rote learning and generalization. Rote learning involves improvement only on trained material, while generalization involves using what was learned during training with untrained or unencountered stimuli. An example of rote learning would be learning to perceive the word "dog" only when spoken by a specific person. An example of generalization would be learning to perceive the word "dog" in any situation regardless of the speaker. Although either type of learning can result in enhanced perception, generalization is the more advantageous of the two. If we were to only advance by rote learning, we would need to be exposed to every existing sound in order to process and recognize it at a later date which could be a nearly impossible task. Possessing the ability to generalize can make a training session even more useful since what was learned can be used in new situations. This can allow a trained listener the potential to communicate with previously unencountered speakers or better perceive degraded speech in noisy environments. Therefore, making a distinction between what type of learning is behind performance improvement can be helpful in not only understanding how listeners adapt to improve their perception, but also in assisting in the creation of better training programs in the future. If we better understand what conditions induce rote learning or generalization, then training regimens can be created and implemented accordingly.

Generalization has been researched using several different tasks and stimulus types such as degraded or accented speech. This can allow insight about certain populations who struggle with a particular sound type. Degraded or vocoded (i.e., synthesized) speech is often compared to what CI users hear, and accented speech can be useful in assessing how listeners learn to perceive speech sounds from an unfamiliar language. For example, Schwab, Nusbaum, and Pisoni (1985) investigated the effect of training using synthesized speech. Participants significantly improved on the task without ever hearing the same word twice during training or test, demonstrating generalization of training to untrained words. Improvements were long lasting and still present at a six month follow up test. Using a similar paradigm, Greenspan, Nusbaum, and Pisoni (1988) trained participants on identifying speech using individual words as well as full sentences. Again, participants significantly improved on the identification task without ever hearing the same word twice during training or test. Performance of those who trained with repeated stimuli only improved similarly to those trained with novel stimuli when the repeated stimuli set was large and highly variable. This is a particularly useful finding for those interested in creating training programs since it suggests that participants can spend less time training and still get comparable results. Francis, Baldwin, and Nusbaum (2000) investigated whether participants could be trained to attend to specific acoustic cues when identifying phonemes, and whether this training could be generalized and applied to novel stimuli. Participants were indeed able to attend to the cue they were trained on, and use their training in identifying new untrained phonemes. Bradlow, Pisoni, Akahane-Yamada, and Tohkura (1997) were able to demonstrate generalization with participants trained with accented speech. Feedback was

used to train Japanese listeners to discriminate between the English /r/ and /l/ during a word identification task. Participants significantly improved performance on trained words as well as were able to generalize training to new words spoken by new speakers. These improvements were long lasting and still present at a follow up test three months later (Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999). Together, these studies demonstrate the utility of generalizing given that learning could be extended beyond training. This is of the utmost importance if training is to be useful in situations outside of the laboratory.

Although it seems clear that participants can generalize in multiple situations involving speech, there are certainly instances when this effect was not found. Bradlow and Bent (2003) reported that participants could only generalize training on perception of accented speech when the training session was highly variable and contained multiple talkers. Training with a single talker did not result in generalization. Likewise, Eisner and McQueen (2005) reported talker specific learning in their experiment which tested the perception of accented speech. Given that there is some variability in the reports in this area it is worth considering why generalization is possible in one setting and not another, or for that matter, why it is possible at all. As has been discussed above, long term exposure or practice generally results in enhanced perception of the encountered sound. It is widely accepted that behavioral improvements are possible due to structural or neurochemical changes that the brain undergoes in response to incoming stimuli, often referred to as experience-dependent changes (Karni, 1996). While this provides a straight forward mechanism for rote learning, it does not provide a satisfying explanation for generalization learning. An underlying principle of generalization is that there is some

shared component between trained and untrained material. In some cases, participants are trained on a task with one set of sounds and are then tested on the same task with new sounds (Gaspar et al., 2013). This suggests a specific skill or rule was learned, which can be applied in many situations even if the stimuli change. In other cases, participants are trained on identifying sounds or words which might share spectral or temporal characteristics, but which differ semantically (Schwab, Nusbaum, & Pisoni, 1985; Greenspan, Nusbaum, & Pisoni, 1988). Whether or not any shared components are processed by the same higher or lower level neural networks may be what enables a transfer of learning. For instance, it has been shown that primary and secondary auditory cortices process and represent spectral and temporal features of speech (Obleser, Elbert, Lahiri, & Eulitz, 2003; Pasley et al., 2012). It may be possible that training on a set of words or sounds sufficiently activates and strengthens a wide range of frequency sensitive neural networks, which then facilitates the perception of new unencountered words (i.e., generalization). This may explain how listeners were able to generalize learning in the paradigms which involved identifying synthetic speech (Schwab, Nusbaum, & Pisoni, 1985; Greenspan, Nusbaum, & Pisoni, 1988). This however remains an unanswered question. Further investigation is necessary to investigate what processes underlie a learning transfer, and whether this is a general mechanism or perhaps specific to a stimulus or task.

In order to make any claims about the effect of training on neural processing, it is important to measure the physiological changes that parallel or mediate improvement. These changes are commonly measured by recording brain activity during a perceptual task using functional magnetic resonance imaging (fMRI), electroencephalography

(EEG), magnetoencephalography (MEG), or transcranial magnetic stimulation (TMS). Although all of these techniques can be useful, each one offers different advantages. For instance, using EEG or MEG to record event-related potentials (ERP) or fields (ERFs), respectively, can be useful due to the high temporal resolution which enables measurement of changes related to precise moments during stimulus presentation. Several studies have used ERPs during a speech perception task in order to investigate what areas of the brain undergo changes after training. For example, the N1, P2 and N1c ERP amplitude peaks have all been reported to reflect changes that result in or parallel behavioral improvements in a variety of experiments measuring auditory perception including oddball detection tasks (Tong, Melara, & Rao, 2009), pitch discrimination tasks (Bosnyak, Eaton, & Roberts, 2004), as well as tests which train listeners to discriminate between different voice onset cues (Tremblay, Shahin, Picton, & Ross, 2009). Results typically include increases in amplitude or shortened latency which are importantly coupled with behavioral improvements and reduced response times. Musicians have been found to have enhanced amplitude in P2 and N1c components as compared to non musicians when presented with tonal stimuli (Shahin, Bosnyak, Trainor, & Roberts, 2003). Enhancements are suggested to be a result of the long term training that accompanies musicianship.

Experiments investigating ERP training effects during other speech perception tasks have also been reported. For example, Reinke, He, Wang, and Alain (2003) recorded ERPs while participants trained on a vowel segregation task. During the task, listeners were concurrently presented two vowels and asked to respond by button press which vowels they believe they heard. In order to address preliminary skill at the task and

the effect of training, listeners completed a pretest and a post-test. The trained group's performance improved during the pretest as well as significantly improved after the training sessions at post-test, while the control group (untrained) experienced the same minimal improvement during the pretest but no additional learning at post-test. Importantly Reinke et al. (2003) also found a decrease in latency in the N1c and increase in amplitude in the P2 at both right and left temporal electrodes at post-test for only those that received training. Alain, Snyder, He, and Reinke (2007) also investigated perceptual learning by using the same vowel segregation task while measuring the brain responses that paralleled improvement in performance. To accomplish this, ERPs were recorded during five blocks of trials during a pretest and five blocks of trials during a post-test a week later. Training occurred during the intervening week. No ERPs were recorded during training. Pretest analyses revealed that an early component, Ta from 100-140 ms at temporal electrodes, had enhanced amplitude as block number increased for both groups. Likewise performance improved as block number increased for both groups at pretest; however, only those who received training during the week significantly improved from pre-test to post-test. No significant changes were observed in the N1 or P2 during the pretest for either group as a function of block. The post-test ERPs of the trained group showed a dramatic increase in amplitude of Ta and P2 when compared to pretest, but did not show any enhancements during the post-test as a function of block. In contrast, the post-test ERPs of the untrained group showed significant within-session changes in amplitude of the Ta and P2, but did not show amplitude differences when compared to pretest. The differences in the ERP results between the two groups are likely due to one group receiving extended training. The trained group’s ERPs underwent

changes during the training session in the intervening week while the untrained group appeared to start undergoing changes during the post-test since they received no prior training. The above reports seem to conclusively show certain components that reflect perceptual learning, although the use of repeated stimuli prevents any inferences to be made about neural changes that represent or are the result of generalization. Further research is therefore needed in order to see if ERP components such as the N1 and P2 can be informative on the differences in neural processing underlying different learning types.

In addition to distinguishing between different types of learning, it is also important to consider how the improvement unfolds over time. This can be an important variable to measure since the speed at which learning unfolds can be an indicator of the different neural mechanisms underlying advanced performance. Faster within-session improvement could reflect quick neurochemical changes that strengthen synaptic connectivity between neurons in a network important for the trained task. Slow learning, represented by slow between-session improvements may reflect structural synaptic changes that take longer to complete (Gilbert, 1994; Karni & Bertini, 1997). In addition, there are other modulating factors that affect the time course of learning. For example, the learning reported in the studies reviewed thus far may be due to a memory consolidation process that is sleep dependent instead of being simply time dependent. Although there is some debate as to the most appropriate definition of memory consolidation, it is essentially a process that transforms a new memory into a stronger more stable form that is resistant to decay or interference (Stickgold & Walker, 2007). Sleep dependent memory consolidation would then be a similar process that is dependent on and evident after a period of sleep (Diekelmann, Wilhelm, & Born, 2009). An

important component of this is that similar effects are not seen during an equal time spent awake, and that consolidation is halted if sleep deprivation occurs (Stickgold, 2005). In addition, the duration or intensity of wakefulness has a direct impact on the time spent in certain sleep cycles (e.g., REM or SWS), and the amplitude of oscillatory activity during sleep (Stickgold, 2005; Tononi & Cirelli, 2012). Enhanced amplitude of oscillatory activity during sleep has even been shown to occur in the same brain regions that were involved in training on a task just before sleep (Huber, Ghilardi, Massimini, & Tononi, 2004). Similar results have been found in studies which induced neural potentiation using TMS (Huber et al., 2007). What is perhaps most important is that the intensity and length of time spent in sleep cycles such as REM and SWS is positively correlated with behavioral improvement (Stickgold & Walker, 2007).

There are numerous accounts reporting the positive effects of sleep on learning in a wide range of tasks in both the visual and auditory domain. For example, Mednick et al. published two reports in which they tested whether a short sleep period, such as a nap, could have a positive effect on learning on a visual discrimination task. In the first study, the author's found that a short sleep period reversed performance deficits that were the result of exhaustion from numerous training sessions. Those who spent an equivalent period of time awake and resting were not able to recover any performance loss (Mednick et al., 2002). In a later study, Mednick et al. (2003) reported that participants improved as much on a visual discrimination task after a short nap as those who slept for a full night. Again, those who did not sleep experienced no such performance enhancement. Tucker, Tang, Uzoh, Morgan, and Stickgold (2011) investigated the effect of sleep and reward on learning, and were able to demonstrate that a night of sleep was more effective in

enhancing performance on a visual identification task than the offer of a monetary reward. This importantly demonstrates that this effect cannot be attributed to a simple desire or intent to improve.

Reports within the auditory domain also suggest a benefit of sleep on learning. Gaab, Paetzold, Becker, Walker, and Schlaug (2004) measured the degree of auditory learning on a pitch discrimination task using sine wave tones, and found a significant increase in performance on the task after a night of sleep versus those who spent the same period of time awake. Similarly, Gottselig et al. (2004) trained participants on an auditory tone sequence task and investigated whether sleep, a restful break without sensory input, an unrestful break with sensory input (rested while watching a movie), or no break had differential effects on learning. Both sleep and a restful break had a significant impact and enhanced performance on the pitch discrimination task. Fenn, Nusbaum, and Margoliash (2003) looked at the effect of sleep on generalization and used a paradigm similar to that of Schwab et al. (1985) and Greenspan et al. (1988), which tested how well participants identified synthesized speech. All participants significantly improved as a result of training; however, the greatest improvement was observed in those who had a night of sleep before testing.

While it is clear that there is often a positive relationship between sleep and learning, explanations as to why or how this is accomplished are not entirely agreed upon. For example, the dual-process hypothesis suggests that it is certain sleep cycles that enhance performance on certain tasks, namely rapid eye movement (REM) sleep and slow wave sleep (SWS) consolidate procedural and declarative learning, respectively

(Diekelmann et al., 2009). The sequential model suggests it is a certain combination and order of REM and SWS that will strengthen learning (Conte & Ficca, 2013). The synaptic downscaling hypothesis suggests that it is the slow wave oscillations during SWS that reduce the metabolic resources necessary to maintain strengthened neural connections, thereby resulting in greater ease in completing trained tasks (Tononi & Cirelli, 2003, 2006). A reason as to why one theory or hypothesis has not gained more following could be the fact that the reports of the effects of sleep on learning are not entirely consistent. For example, the size and sometimes the presence of a facilitative sleep effect appears dependent on the type of stimuli (e.g., visual vs. auditory stimuli, or simple vs. complex sounds), or the type of learning measured in an experiment (e.g., procedural, declarative) (Conte & Ficca, 2013). It has been suggested that memory consolidation is not a unitary phenomenon, and contains a series of steps (Stickgold & Walker, 2005). It may be the case that certain neural systems may not be able to complete these steps while still awake. While there are a multitude of studies examining sleep dependent rote learning, sleep dependent generalization remains a relatively untouched topic. Given that generalization itself is not a unified construct and may be accomplished in different ways, further investigation is necessary to determine whether generalization learning is invariably enhanced by sleep or if the different strategies that it can be accomplished change its dependency on sleep to consolidate.

CHAPTER 2

CURRENT STUDY

There are several research questions that remain unanswered. While prior research confirms the possibility of generalization when training to better perceive individual words/phonemes (e.g., Bradlow & Bent, 2003; Eisner & McQueen, 2005; Greenspan, Nusbaum, & Pisoni, 1988), it is not clear whether past trends suggest a general rule for the conditions needed to induce a transfer of learning, or whether they just represent a rule specific to word/phoneme identification. Given that most environments are rarely silent and often contain several overlapping conversations or sounds, an ability to segregate co-occurring sounds is an important skill to master. Despite this, listeners with hearing impairments (e.g., Grieco-Calub, Saffran, & Litovsky, 2009), older adults (e.g., Alain et al., 2006), or individuals who have suffered neurological damage to auditory cortices all can experience difficulty in accurately perceiving simultaneously occurring sounds. While training has been shown to positively benefit speech segregation ability (Alain et al., 2007), it has only been addressed in studies measuring rote learning, i.e., improvement just on trained material. Since the goal of a training study should be to positively impact perception inside and beyond the laboratory, an important question remains whether training-induced learning can be applied to untrained sounds when learning to segregate speech. The current study aimed to address this question by measuring generalization (i.e., transfer of learning to untrained sounds) after listeners complete training on a vowel segregation task (adapted from Alain et al., 2007). During this task listeners are concurrently presented a pair of vowels and asked to identify both vowels. Success on this task is dependent on an ability to

perceptually segregate and appropriately group the acoustic information for each vowel, with larger differences in pitch resulting in greater ease at the task. Since pitch processing facilitates segregation, this was the variable that was focused on during training. Vowel pairs were divided into two sets for training, each which contained pairs with different frequency (i.e., pitch) separations. By training listeners on one of the two sets and testing on the full set, it is possible to address whether listeners can generalize segregation ability by measuring the performance on untrained sounds after training. Importantly, changing the pitch of a vowel changes its harmonic structure (Ruben & Vatikiotis-Bateson, 1998). Therefore, improvement on vowels with untrained frequency separations cannot just be attributed to memorization of the vowel pairs themselves during training.

Data from prior generalization studies suggest that training with an acoustically variable stimulus set can result in enhanced processing of new untrained words/phonemes (Bradlow & Bent, 2003; Greenspan, Nusbaum, & Pisoni, 1988). Speech sounds are already complex in nature given that they are composed of a series of harmonics at different frequencies. Consequently, training with a variable stimulus set may have allowed for auditory cortices to become better tuned to a wide range of frequencies, which would importantly facilitate encoding of the harmonic structure of new words/phonemes. If the current study follows past trends, then training with a set of vowels should induce generalization of any learned segregation ability to vowel pairs separated by untrained frequency separations, due to retuning of low level sensory areas. On the other hand, improvement may be dependent on learning the sounds themselves at each frequency separation. If this is the case, then no transfer should be observed.

A second unanswered question involves uncovering which modulating factors are behind generalization. Specifically, it is important to determine whether learning is time dependent or perhaps dependent on another factor such as sleep. Given that this has not been addressed in the context of speech segregation, it remains unclear what role sleep plays in consolidating segregation ability. In order to address this, multiple strategically scheduled post-tests were carried out after training. By scheduling post-tests after training as well as before and after a night of sleep, it can be observed whether sleep is integral in stabilizing learning, or if a process simply dependent on time is involved.

Lastly, while much has been found in regards to neural changes that parallel rote learning, such correlates of generalization are not known. Measuring brain activity during both learning types can make it possible to differentiate between the neural changes that result in generalization as compared to rote learning. In order to address this, EEG was recorded during all tests and training sessions. ERP analyses were done in order to identify the presence of certain components thought to be markers of perceptual learning, e.g. N1, P2 (Alain et al., 2007; Reinke et al., 2003). Additionally, spatiotemporal analyses were done to better understand the processes underlying any found ERP modulations.

CHAPTER 3

METHODOLOGY

Participants

Thirty-two normal hearing (thresholds <35 dB from 250 to 8000 Hz) fluent English speakers were recruited from the community with ages within the range of 18-35 years (M=24.12, SD= 4.14). Any person with a neurological or psychiatric disorder was excluded from participation. Each listener was compensated $10 for each hour of participation, with payment being prorated to the half hour.

Stimuli

Eight 200 ms American English vowels were synthesized using Klatt at the following fundamental frequencies: 100, 101, 103, 106, 112, and 126 Hz. The vowels include: 'AE' as in apple (/æ/), 'AH' as in hot (/a/), 'EE' as in feet (/i/), 'EH' as in met (/e/), 'ER' as in hurt (/ɜ/), 'OH' as in boat (/əʊ/), 'OO' as in hoot (/u/), 'UH as in hut (/ʌ/). This resulted in forty-eight unique single vowel stimuli. To create double-vowel pairs Praat software was used. The waveforms of two vowels were summed together with one $f0$ always being 100 Hz and the other being either 100 Hz, 101 Hz, 103 Hz, 106 Hz, 112 Hz, or 126 Hz, which are 0, 0.25, 0.5, 1, 2, or 4 semitones higher than 100 Hz, respectively. This resulted in 308 unique double vowel pairs. Stimuli were split into two groups for training and will be referred to as Training A pairs and Training B pairs. Grouping was based on the frequency separation between the vowels in a pair. Training A pairs contained all vowels with Δf_0 separations of 0.5 semitones (i.e., 100 Hz and 103 Hz), as well as Δf_0 differences of 2 semitones (i.e., 100 Hz and 112 Hz) (n= 112 pairs). Training

B pairs contains all vowel pairs with Δf_0 separations of 1 semitone (i.e., 100Hz and 106Hz), as well as Δf_0 differences of 4 semitones (i.e., 100 Hz and 126 Hz) (n= 112 pairs). Preliminary pilot testing was done to ensure equal difficulty between groups, Training A pairs (M=41.76, SE =7.65) Training B pairs (M=42.93, SE=7.30). The remaining two groupings of vowel pairs (i.e., vowel pairs with f_0s of 100 Hz and 100 Hz, and with 100 Hz and 101 Hz) could not be equally distributed to the two training groups due to there being an uneven number of vowel pairs for one of the groupings (i.e., vowel pairs with 100 Hz and 100 Hz). This was a result of there being duplicates after the aforementioned vowel summing process (e.g., combining ER100 Hz and EH100 Hz yields the same sound as EH100 Hz and ER100 Hz). Therefore, all remaining vowel pairs with Δf_0 separation of 0 semitones (i.e., 100 Hz and 100 Hz) (M=36.31, SE =7.33) or 0.25 semitone (i.e.,100 Hz and 101 Hz) (M=36.80, SE =6.22) were not used for training and will be referred to as Not Trained pairs (n= 84 pairs).

Procedure

Single-Vowel Practice. Session one began at 8:30 am. Informed consent and demographic information were obtained before the experiment. Participants were also asked to fill out a log that kept record of their sleep habits, caffeine consumption, and general level of fatigue for the seven days prior to and for the duration of the study. Afterwards participants were seated in a comfortable chair in a sound attenuated booth for testing. Stimuli were presented using Presentation software on a PC computer and played through Sennheiser headphones. An example of each vowel sound was presented after which a brief two-part Practice test was completed. During part one, a single vowel was presented and participants were asked to identify what they heard by selecting one of

eight buttons, each labeled for one of the vowels. Visual and auditory feedback was given after each response. Visual feedback was displayed on the screen and indicated whether or not the response was correct or incorrect, as well as what the correct response was for that trial. Auditory feedback consisted of a replay of the vowel sound for that trial. Both types of feedback always played regardless of whether the response was correct or incorrect. Visual feedback always displayed first and was immediately followed by auditory feedback. During part two of the test, feedback was no longer given and a score of 80% or above was required in order to continue participation in the rest of the experiment. This was done in an effort to ensure that any future difficulty with the vowel segregation task would not be the result of an inability to distinguish between or identify the individual vowels. Any participant with a score lower than 80% was excluded from participating, paid for their time, and debriefed about the study.

Pretest. Next, during the Pretest participants were diotically presented with double-vowel pairs through ER-3A headphones and asked to respond by pressing the two buttons that represented the two vowel sounds that they heard. No feedback was given about the correctness of their response. Participants were given 4000 ms to respond before the trial terminated and the next trial began. The Pretest consisted of all 308 total trials, which were presented in two blocks of 154 trials. A break was offered between each block. Electroencephalograph (EEG) signals were recorded during this test as well as all subsequent tests.

Training. Next, participants were assigned to one of three groups: Training Group A, Training Group B, or the Control group. Group assignment was quasi-random

such that an equal number of participants were in each group. Those assigned to the Control group watched a documentary (Planet Earth) for the same time period a training session would last for those in groups Training Group A or Training Group B (approx 27 min). Those assigned to either Training Group A or Training Group B completed the same task as during the Pretest, but were now given feedback on their performance. Those in the Training Group A only trained on stimuli in Training A pairs, and those in Training Group B only trained on stimuli in Training B pairs (See **Stimuli** for a breakdown). Visual feedback displayed whether their responses were correct or incorrect as well as what the correct responses were for that trial. Additional auditory feedback consisted of a replay of the double vowel pair for that trial. Both types of feedback always played regardless of whether their response was correct or incorrect. Visual feedback always displayed first and was immediately followed by auditory feedback. Participants had 4000 ms to respond after which the feedback portion of the trial would automatically play. The next trial automatically began 2000 ms after the feedback had finished. Each training set had 112 unique trials (see **Stimuli** above). All of the trials were presented once during block one, and repeated in a different order during block two resulting in a total of 224 trials per training session.

Post-test one - Immediate. Next, Post-test one took place immediately after training during session one. All methods including stimuli and procedure were the same as the Pretest.

Post-test two - 12 hours after Pretest. Post-test two occurred during session two and began at 8:30 pm on the same day. All methods including stimuli and procedure were the same as the Pretest.

Post-test three - 24 hours after Pretest. Post-test three occurred during session three and began at 8:30 am on the following day. All methods including stimuli and procedure were the same as the Pretest.

Electrophysiological Recording. Actiview software was used to record EEG signals using a BioSemi ActiveTwo system. Signals were recorded from the scalp using 64 pin-type electrodes spaced according to the 10/20 system, and eight flat-type electrodes that were placed on the face below the hairline. These points include both mastoids, both pre-auricular points, outer canthi, and inferior orbits of each eye. Common Mode Sense (CMS) and Driven Right Leg (DRL) were used as reference and ground. Signagel was applied to the scalp before electrode placement as a conductor. Voltage offsets were ensured to be below 40 μV before beginning the experiment. All electrophysiological procedures have been obtained from manuals provided by BioSemi (See http://www.biosemi.com/index.htm). EEG signals were then recorded for off line analysis.

Dependent Measures

Behavioral Data. Participants' performance was calculated as percent correct (i.e., both vowels answered correctly) separately for each Trial Type during each testing session. All trials that contain a vowel pair in which a Training Group trains with will be referred to as Trained trials. All trials that contain a vowel pair in which a Training Group

does not train with will be referred to as Untrained trials. All remaining vowel pairs, which neither group trains on (See **Stimuli** for explanation), will be referred to as Not Trained trials (See Table 1 for breakdown). In order to examine the effect of training on performance, averages were submitted to a 3 x 4 (Group [Training Group A, Training Group B, Control] x Test session [Pretest, Post-test one, Post-test two, Post-test three]) mixed design Analysis of Variance (ANOVA) with one between-subjects factor (Group) and one within-subjects factor (Test session). Additional ANOVAs and pairwise comparisons were done in order to understand any interactions. No performance differences are expected to occur between Training A and Training B Groups, however, differences are anticipated between the Trained groups and the Control Group due to the expected benefit of training. To examine generalization, averages were also submitted to a 2 x 3 x 4 (Group [Training Group A, Training Group B] x Trial Type [Trained, Untrained, Not Trained] x Test session [Pretest, Post-test one, Post-test two, Post-test three]) mixed design ANOVA, with one between-subjects factor (Group) and two within-subjects factors (Trial type and Test session). Generalization would be represented by significant improvement on Untrained trials across Test Sessions. Given that a transfer of learning would be a result of applying skill gained during training, significant improvement on Trained trials is expected to occur in parallel with generalization. Therefore, no performance differences are expected between Trained and Untrained trials across Test Sessions. Since Not Trained trials are not of equal difficulty with Trained and Untrained trials (See **Stimuli**), a transfer of learning may not be seen without extensive training.

	Trained Trials	Untrained Trials	Not Trained Trials
Training Group A	100 Hz/103 Hz pairs 100 Hz/112 Hz pairs	100Hz/106 Hz pairs 100 Hz/126 Hz pairs	100 Hz/100 Hz pairs 100 Hz/101 Hz pairs
Training Group B	100Hz/106 Hz pairs 100 Hz/126 Hz pairs	100 Hz/103 Hz pairs 100 Hz/112 Hz pairs	100 Hz/100 Hz pairs 100 Hz/101 Hz pairs

Table 1. Breakdown of stimuli for each Trial Type (i.e., Trained, Untrained, and Not Trained trials) for each Training Group.

Event-Related Potential Data. For EEG data, all analyses were done using Brain Electrical Source Analysis (BESA). Noisy electrodes were interpolated and epochs contaminated by artifacts (amplitude > 150 μV, gradient >75 μV, low signal < 0.10 μV) were automatically rejected before averaging. 1000 ms ERP epochs were then averaged separately for each condition type (i.e., Test Session and Trial Type). ERP averages were baseline corrected by subtracting the average of the 100 ms pre-stimulus period from each time point in the waveform. The data were then filtered to attenuate energy at frequencies below .5 Hz and above 30 Hz. Mean amplitude values were then extracted for the N1 at frontal, frontocentral, and central electrodes (i.e., F1, Fz, F2, FC1, FCz, FC2,C1, Cz, C2) from 90 to 130 ms, and the P2 at frontocentral, central, and centroparietal electrodes (i.e., FC1, FCz, FC2,C1, Cz, C2, CP1, CPz, CP2) from 175 to 230 ms. Amplitude values were submitted to a 3 x 4 (Group [Training Group A, Training Group B, Control] x Test session [Pretest, Post-test one, Post-test two, Post-test three]) mixed design Analysis of Variance (ANOVA) with one between-subjects factor (Training Group) and one within-subjects factors (Test session). A significant increase in

amplitude is anticipated to occur for Training A and Training B across Test Sessions, and is expected to occur in parallel with training induced learning. No differences are anticipated between Trained Groups. Significant differences are anticipated between Trained Groups and the Control Group, such that an increase in amplitude is expected for Trained listeners but not the Control Group. Mean amplitude values were also submitted to a 2 x 3 x 4 (Group [Training Group A, Training Group B] x Trial type [Trained, Untrained, Not Trained] x Test session [Pretest, Post-test one, Post-test two, Post-test three]) mixed design Analysis of Variance (ANOVA) with one between-subjects factor (Training Group) and two within-subjects factors (Trial type and Test session). This was done in an effort to address amplitude changes that are specific to generalization versus rote learning. Any pattern differences between learning types would be represented by a Trial Type x Test Session interaction. Pairwise comparisons and planned t-tests were done in order to understand the group differences that lead to any interactions.

Global Analyses. ERP components (N1, P2) were further analyzed using two additional measures that use all of the electrodes in a complementary fashion, Global Field Power (GFP) and Global Map Dissimilarity (GMD) (Murray, Brunet, & Michel, 2008). Both measures allow additional information to be gained over and above standard ERP analyses of mean amplitude at subsets of electrodes. In particular, GFP provides a reference free measure of the strength of a neural response, while GMD is a measure that determines whether conditions differ in brain source configuration by measuring topographical differences in voltage distribution, while controlling for the strength of the response. Used together, GFP and GMD can allow insight to be gained as to whether conditions differ in only the strength of the response, or if different neural populations

were active between conditions. The GFPs for each participant were baseline corrected before analyses and grand averaging in order to correct for differences in background noise between conditions. GMD was computed using Cartool. First, dissimilarity values between each time point for Test Session were computed (i.e., Pretest vs. Post 1, Post 1 vs. Post 2, and Post 2 vs. Post 3), with possible values ranging from 0 (No difference) to 2 (topographic inversion). Statistically significant differences were then computed by using a non-parametric randomization test. Significant differences at each time point between conditions are those with resulting value less than .05.

CHAPTER 4

RESULTS

Behavioral Data

Combined Trial Type. Analysis of percent correct (i.e., both vowels answered correctly) revealed a significant difference in test scores across Test Sessions on the double vowel task. This was evidenced by a significant main effect of Test Session $F(3, 87) = 28.32$, $p < .001$, $\eta^2_p = .49$. Follow up polynomial contrasts revealed a significant linear trend $F(1,29) = 65.83$, $p < .001$, $\eta^2_p = .69$, which showed a significant increase in performance across Test Sessions. A marginally significant cubic trend was also found $F(1,29) = 4.13$, $p = .051$, $\eta^2_p = .12$. Pairwise comparisons using an LSD correction were done to further explore these trends, and revealed that there was a significant increase in scores from Pretest to Post-test one (i.e., after training) ($p<.05$), and from Post-test two to Post-test three (i.e., after sleep) ($p<.05$), but not between Post-test one and Post-test two (i.e., before sleep) ($p>.05$). The interaction of Test Session x Group was not significant despite a trend for different effects during Test Session between Groups, $F(6, 87) = 2.06$, $p = .06$, $\eta^2_p = .12$, such that Training Group B performed better across Test Session than those in Training Group A or the Control group (Figure 1B). Three additional ANOVAs were done to more closely examine Test Session differences. Test Session is the within-subjects factor for each Group. The ANOVA for Training group A revealed a significant main effect of Test Session $F(3,21) = 3.12$, $p<.05$, $\eta^2_p = .30$ (See Figure 1B). Follow up polynomial contrasts revealed a marginally significant linear trend $F(1,7) = 5.51$, $p = .051$, $\eta^2_p = .44$, which suggests a linear increase in scores across Test Session. Pairwise comparisons using an LSD correction confirmed this and showed a significant difference

between Pretest and Post-test three scores. These effects would suggest that participants in Training Group A improved gradually across Test Sessions, which is likely due to a benefit of training and perhaps additional practice received through testing. The fact that there was not a significant increase in scores from Post-test one to Post-test two but there was an increase after a night of sleep at Post-test three would suggest a facilitatory effect of sleep on learning; however, the presence of a linear trend as compared to a cubic trend would suggest a more gradual improvement between tests.

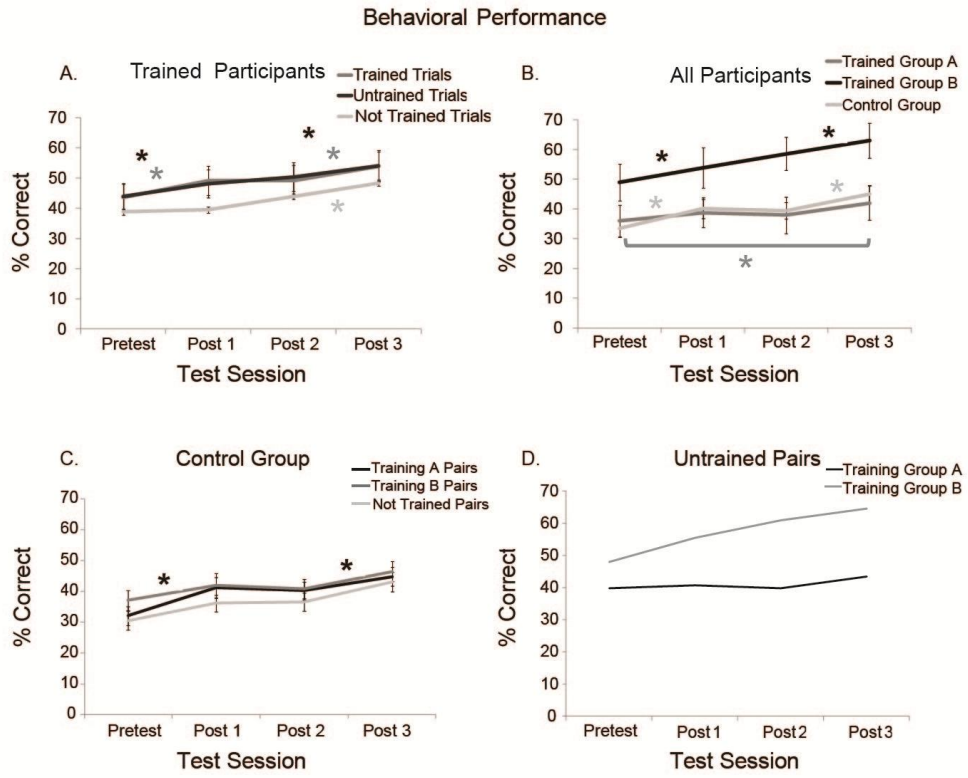

Figure 1. (A) Percent correct for Trained Participants across Test Sessions as a function of Trial Type. (B) Percent correct across Test Sessions as a function of Group. (C) Percent correct for the Control Group across Test Sessions as a function of Stimulus Set. (D) Percent correct for the Untrained trials for Training Group A and Training Group B. Significant differences between Test Sessions marked with an asterisk. Color of asterisk represents Group or Trial Type with significant differences.

Next, the analysis for Training Group B revealed a significant main effect of Test Session $F(3,21) = 13.23$, $p<.001$, $\eta^2_p= .65$. Polynomial contrasts showed that this was a significant linear trend $F(1,7) = 38.15$, $p<.001$, $\eta^2_p= .84$, which revealed a linear increase in scores across Test Session. In addition, pairwise comparisons showed a significant difference between Pretest and Post-test one (i.e., after training) ($p<.05$), between Post-test two and Post-test three (i.e., after sleep) ($p<.05$), but no difference between Post-test one and Post-test two (i.e., before sleep) ($p>.05$). These effects may suggest that training and a night of sleep had an impact on performance for Training Group B, although the overall trend was linear as was seen for Training Group A. This suggests both groups improved gradually over time.

Lastly, the analysis for the Control group revealed a significant main effect of Test Session $F(3,45) = 18.90$, $p<.001$, $\eta^2_p= .55$ (See Figure 1B). Polynomial contrasts revealed that this is a significant linear trend $F(1,15) = 37.65$, $p<.001$, $\eta^2_p= .71$, in addition to a significant cubic trend $F(1,15) = 14.02$, $p<.05$, $\eta^2_p= .48$. Pairwise comparisons showed a significant difference between Pretest and Post-test one (i.e., after training) ($p<.001$), between Post-test two and Post-test three (i.e., after sleep) ($p< .01$), but no significant difference between Post-test one and Post-test two (i.e., before sleep) ($p>.05$). These results suggest a large degree of learning due to practice through testing (testing effects), since this group did not complete any training. Additionally these participants may have experienced a benefit of sleep on learning given that no significant improvement was observed from Post-test one to Post-test two, but a benefit was seen after a night of sleep at Post-test three (i.e., cubic trend). Although this trend was largely

cubic, the data better fit a linear trend which suggests that the Control Group gradually improved as each Test Session progressed.

Split by Trial Type. In order to examine the differential effects of rote learning and generalization, analyses were done on the trained participants alone (i.e., Training Group A and Training Group B). Performance was measured across Test Session and split by Trial Type (i.e., Trained, Untrained, Not Trained). There was a significant main effect of Test Session $F(3, 42) = 14.72$, $p < .001$, $\eta^2_p = .51$, and a significant linear trend $F(1,14) = 35.94$, $p < .001$, $\eta^2_p = .72$, which revealed significant improvement of trained participants on the vowel task across Test Sessions, as was found in analyses with Trial Type combined. Pairwise comparisons revealed significant differences in the scores between Pretest and Post-test one (i.e., after training) ($p<.05$), between Post-test two and Post-test three (i.e., after sleep) ($p<.01$), but no significant difference between Post-test one and Post-test two (i.e., before sleep) ($p>.05$). There was also a significant main effect of Trial Type $F(2, 28) = 30.33$, $p < .001$, $\eta^2_p = .68$. Pairwise comparisons corrected for multiple comparisons showed that this was a result of significantly lower scores on Not trained trials over all ($p<.001$) (See Figure 1A). Next, there was also a significant Test Session x Group interaction $F(3, 42) = 3.15$, $p < .05$, $\eta^2_p = .18$, which revealed that those that were assigned to Training B had the greatest amount of improvement across Test Sessions (See Figure 1B).

Three additional ANOVAs were done to closer examine Test Session differences for each Trial Type (i.e., Trained, Untrained, Not Trained) for Trained participants. This was done to better understand the pattern of improvement that was a result of rote

learning (i.e., performance on Trained trials) versus generalization (i.e., performance on Untrained or Not Trained trials). For each Trial Type, Test Session is the within-subjects factor (i.e., Pretest, Post-test one, Post-test two, Post-test three) and Group (Training Group A, Training Group B) is the between-subjects factor. The analysis for performance on Trained trials showed a significant main effect of Test Session $F(3,42) = 9.72$, $p < .001$, $\eta^2_p = .41$, and that this was a linear trend $F(1,14) = 21.63$, $p < .001$, $\eta^2_p = .60$. Pairwise comparisons showed a significant increase in scores on Trained trials from Pretest to Post-test one (i.e., after training) ($p<.01$), from Post-test two to Post-test three (i.e., after sleep) ($p<.05$), but no difference between Post-test one to Post-test two (i.e., before sleep) ($p>.05$). No significant Group differences were found for performance on Trained trials across Test Sessions. All together these results suggest that there was significant rote learning after training, and that this occurred similarly for both Training Group A and Training Group B since no significant between-subjects effect was found. Additionally, it appears that sleep played a role in consolidating learning given that no additional learning was observed at Post-test two, yet performance improved significantly at Post-test three.

Next for Untrained trials a significant main effect of Test Session $F(3,42) = 11.72$, $p < .001$, $\eta^2_p = .44$, and a significant linear trend $F(1,14) = 39.10$, $p < .001$, $\eta^2_p = .73$, were found. Pairwise comparisons showed a significant increase in scores from Pretest to Post-test one ($p<.05$), and from Post-test two to Post-test three scores ($p<.05$), but no difference between Post-test one and Post-test two scores ($p>.05$). Additionally, a significant Test Session x Group interaction was found $F(3,42) = 6.09$, $p < .01$, $\eta^2_p = .30$, which was the result of a steeper pattern of increase in scores on Untrained pairs across

Test Session for Training Group B. Unlike the results from Trained trials, Group differences were found, such that Training Group B improved significantly on Untrained trials while Training Group A improved modestly only after a night of sleep at Post-test three (See Figure 1D).

Finally, for Not Trained trials, a significant main effect of Test Session $F(3,42) = 10.30$, $p < .001$, $\eta^2_p = .42$, and a significant linear trend $F(1,14) = 19.80$, $p < .001$, $\eta^2_p = 58$, were found. No significant Group differences were found for performance on Not Trained trials across Test Sessions. Pairwise comparisons revealed that there was no change in scores from Pretest to Post-test one, a non significant increase in scores from Post-test one to Post-test two (p=.07), but a significant increase in scores from Post-test two to Post-test three (p<.01) and from Pretest to Post-test three (p<.01). These results suggest also suggest a potential role of sleep enhancing learning, seeing that improvement was only observed after a night of sleep at Post-test three.

The improvement on Untrained and Not Trained trials may be representative of generalization learning, given that performance improved despite no special training on these sounds; however, the significant testing effects observed thus far (i.e., Control group performance) give reason to doubt that improvement on any untrained material (Untrained or Not Trained pairs) during Post-test two and Post-test three is only due to generalization. Therefore, performance on Untrained or Not Trained trials at Post-test one would be the ideal Test Session to observe a transfer of learning since only minimal testing had occurred at that point. Improvement was only seen on Untrained trials for Training Group B at Post-test one (See Figure 1D). Training Group A only improved

modestly on Untrained trials and this did not occur until Post-test three. Neither group improved on Not Trained trials until Post-test two and Post-test three. It seems then that generalization did occur, although only during Untrained trials for one of the groups.

Split by Stimulus Set. Given that the Control group did not train on any material, they could not be included in the previous analysis that compared performance across Test Session and split by Trial Type. In order to observe whether there were different patterns of improvement for the Control group on the vowels in the different stimulus sets (See **Stimuli** for a breakdown of these sets), a separate ANOVA was done, which had Test Session (Pretest, Post-test one, Post-test two, Post-test three) and Stimulus Set (Training A Pairs, Training B Pairs, Not Trained Pairs) as within-subjects factors. When used comparatively with the Trained Groups analyses (See **Behavioral Data Split by Trial Type**), this analysis can allow for patterns of training related learning (i.e., performance on Trained, Untrained, and Not Trained trials) to be differentiated from learning due to testing effects. A significant main effect of Test Session was found $F(3, 45) = 19.39$, $p < .001$, $\eta^2_p = .56$, which was the result of significant differences in performance across Test Sessions. Follow up polynomial contrasts revealed a significant linear trend $F(1,15) = 37.86$, $p < .001$, $\eta^2_p = .71$, and a significant cubic trend $F(1,15) = 14.34$, $p < .01$, $\eta^2_p = .48$, as was found for the Control Group during analyses when Trial Type was combined. Pairwise comparisons with an LSD correction were done to further break down observed trends, and showed a significant increase in percent correct from Pretest to Post-test one ($p < .001$), Post-test two to Post test three ($p < .01$), but not between Post-test one and Post-test two. In addition, a significant main effect of Stimulus Set was also found, $F(2,30) = 13.81$, $p < .001$, $\eta^2_p = .47$ (See Figure 1C). Pairwise comparisons

showed that this was due to all three Stimuli Sets being significantly different from another (all *p's* <.05). Participants scored highest on Training B pairs, and lowest on Not Trained Pairs. No interactions were found for Stimulus Set and Test Session, which suggests that improvement on all three pair types improved similarly across Test Sessions.

Training Session Data

Analysis of percent correct showed a non-significant main effect of Block, $F(1,14) = 4.072$, $p=.06$, $\eta^2_p= .22$. The Block x Group interaction was not significant, although there was a significant main effect of Group $F(1,14) = 7.63$, $p<.05$, $\eta^2_p= .34$, which was a result of those in Training Group B performing better than Training Group A overall.

Post-hoc Analyses

Additional analyses were done to examine what external factors, if any, might be contributing to the Group differences seen for performance. Prior to participating in the study, each participant completed the Practice test which tested their ability to identify each vowel in isolation. Although any listeners who struggled with the Practice test were excluded from participation (i.e., percent correct < 80%), not all included participants scored at ceiling. It is therefore possible that some participants may have started out with a greater aptitude for identifying the vowels used in this study. To explore this, correlations were done to explore the relationship between initial ability to identify vowels in isolation (i.e., single vowels) and the capacity to improve on the task (See Table 2 for a results summary).

A significant positive correlation between Practice and Pretest scores, r (28)= .40, $p < .05$, and between Practice and Post-test three scores, r (28)= .41, $p < .05$ was found, which suggests that those who could easier identify individual vowels initially were better able to segregate and identify the double vowels throughout the study. On the other hand, those who scored highly on the Practice test didn't necessarily improve the most. This was evidenced by a non-significant correlation between the Practice test and the slope of learning from Pretest to Post-test three. Further analyses were done to explore whether prior experience, such as musicianship, may have influenced scores on the double vowel task. There was a significant positive correlation between the number of years a listener spent training as a musician and Practice test scores, r (28)= .42, $p < .05$, which suggests that those who trained longer to be a musician tended to better initially identify individual vowels. Next, a non-significant correlation between years of training as a musician and both Pretest and Post-test three scores suggests that more training as a

	Practice Scores	Years of Music Training	Hours Slept Before Test	Restfulness
Pretest Score	.40*	.01	-.09	.03
Post 3 Score	.41*	.11	-.25	.19
Slope of Learning	.08	.24	.02	.31
Years of Music Training	.42*			

Table 2. Results from correlations between test scores and participant demographic information, and also between test scores and participant sleep habits. Significant correlations marked with an asterisk.

musician does not necessarily translate to greater segregation ability on the double vowel task. Hence, while those who trained longer as a musician tended to be more skilled at identifying single vowels, they were not more skilled at segregating and identifying the

double vowel pairs. In addition, a non-significant correlation was also found between the slope of learning from Pretest to Post-test three and years of training as a musician, r (28)= .24, p=.19. This would suggest that longer training as a musician does not necessarily to lead to a large benefit of training on this task.

Next, analyses were done to explore whether sleep had a role in the degree of learning for participants. Two dimensions of sleep were explored. The first was the number of hours slept before a test, and the second was how restful the sleep period was before a test. Restfulness was reported by the participant as a percentage ranging from 0% (i.e., not at all restful) and 100% (i.e., completely restful). No significant correlations were found (See Table 2).

While certain external factors may have had a small role in affecting participants' score on the vowel task, the above correlations fail in providing a satisfying explanation as to why there are such large differences in the amount of learning between Training A and Training B Groups. Although preliminary pilot testing suggested that the difficulty level of Training A pairs and Training B pairs was equal (See **Stimuli**), the frequency separation between vowels in Training B pairs is somewhat larger than the vowels in Training A pairs. Importantly, a larger frequency separation can ease the process of segregating two sounds since the pitch information for each sound would be more distinct. Therefore, it is possible that training on an easier set of vowel pairs (i.e., Training Set B) may have resulted in the larger learning effects observed for those in Training Group B. To explore this, a t-test was done to compare performance for Trained Participants on the vowels for Training A pairs and Training B pairs at Pretest. There was

a significant difference in the scores at Pretest for Training A pairs (M=41.73, SE=4.22) versus Training B pairs (M=45.73, SE=4.76), t(15)= -3.22, p < .01. This confirmed that participants did indeed have more ease in identifying the vowel pairs in Training Set B beginning at Pretest. The above results suggest that the difficulty level of training material had a large effect on how much was learned during the task, such that training with easier material resulted in more learning overall.

Event-Related Potential Data

Combined Trial Type. Analyses were first done on mean amplitude values across Test Session and Group (i.e., Training Group A, Training Group B, and Control). Analysis of mean N1 amplitude revealed a non-significant main effect of Test Session $F(3, 87) = 2.35$, $p = .07$ $\eta^2_p = .08$ (Figure 3A), although polynomial contrasts revealed a significant cubic trend $F(1,29) = 7.92$, $p < .01$, $\eta^2_p = .21$ (See Figure 3E). Pairwise comparisons showed significant differences in N1 mean amplitude between Post-test one and Post-test two ($p<.05$), but did not show differences between Pretest and Post-test three ($p=.71$).

Although some of the behavioral analyses also revealed cubic trends also, the lack of N1 mean amplitude difference found between Pretest and Post-test three suggests that we might be observing changes related to habituation (i.e., reduction in N1 amplitude from repeated stimulus presentation) and release from habituation, rather than training related changes.

Examination of P2 amplitude showed an increase in amplitude as Test Session increased, which was evidenced by a significant main effect of Test Session $F(3, 87) =$

7.47, $p < .001$, $\eta^2_p = .20$ (Figure 3A), and also by a significant linear trend, $F(1, 29) = 25.40$, $p < .001$, $\eta^2_p = 46$. In addition, a significant Test Session x Group interaction was found, $F(6, 87) = 2.74$, $p = .05$ $\eta^2_p = .16$, although there was not a significant between-subjects effect of Group ($p > .05$). The interaction was due to a different pattern of amplitude increase for each group (Figure 3F), such that Training Group B showed the largest increase in amplitude across Test Sessions, as was found with behavioral data.

Since there were Group differences in amplitude across Test Sessions, P2 amplitude was further examined for each group in separate ANOVAs. Each ANOVA had Test Session as a within-subjects factor (Pretest, Post-test one, Post-test two, Post-test three). A non-significant main effect of Test Session was found for Training Group A, $F(3, 21) = 2.27$, $p = .10$ $\eta^2_p = .24$. A significant main effect of Test Session was found for Training Group B, $F(3,21) = 8.42$, $p < .05$ $\eta^2_p = .54$ (See Figure 3C). Polynomial contrasts showed that this was a significant linear trend, $F(1,7) = 19.16$, $p < .01$ $\eta^2_p = .73$. Pairwise comparisons were done to further explore these trends and found a significant increase in P2 amplitude from Post-test one to Post-test two ($p < .01$), and from Pretest to Post-test three ($p < .05$) (See Figures 3C and 3F). Lastly, a non-significant main effect of Test Session was found for the Control Group, although there was a significant linear trend $F(1,15) = 15.76$, $p < .01$), $\eta^2_p = .51$ (See Figures 3D and 3F). Pairwise comparisons showed that this was due to a significant increase in P2 amplitude from Pretest to Post-test three ($p < .01$). With the exception of those in Training Group A, participants' P2 amplitude increased gradually which suggests that it was greater modulated by time and practice through testing rather than by sleep.

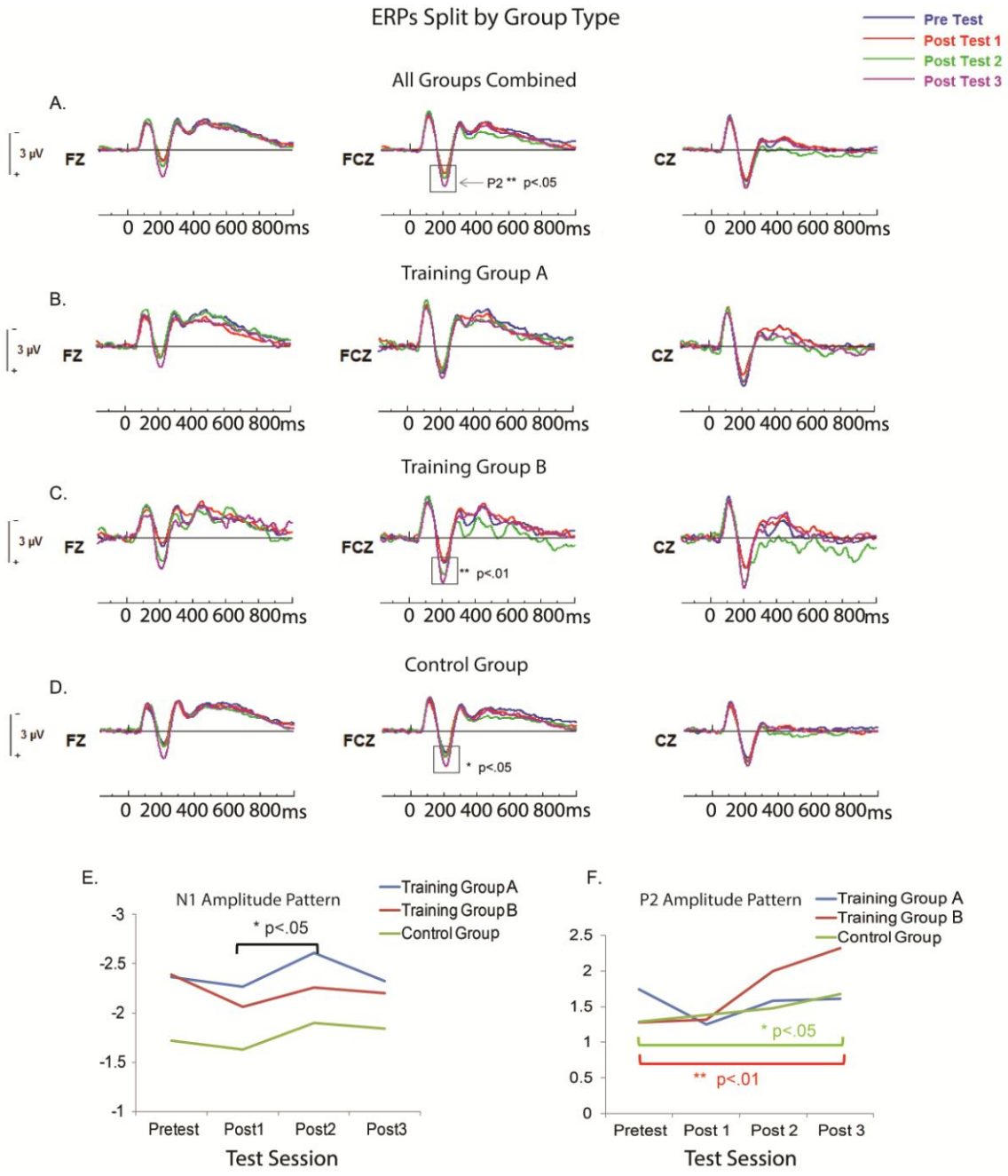

Figure 2. (A) Group mean ERPs across Test Sessions. (B) Training Group A mean ERPs across Test Sessions. (C) Training Group B mean ERPs across Test Sessions. (D) Control Group mean ERPs across Test Sessions. (E) Average N1 mean amplitude shown across Test Sessions as a function of Group. (F) Average P2 mean amplitude shown across Test Session as a function of Group. Significant differences marked with brackets and an asterisk. Bracket color will represent significant effects specific to a Group (i.e., Blue = Training Group A, Red = Training Group B, Green = Control Group) or effects only found with Groups combined (i.e., Black= All Groups).

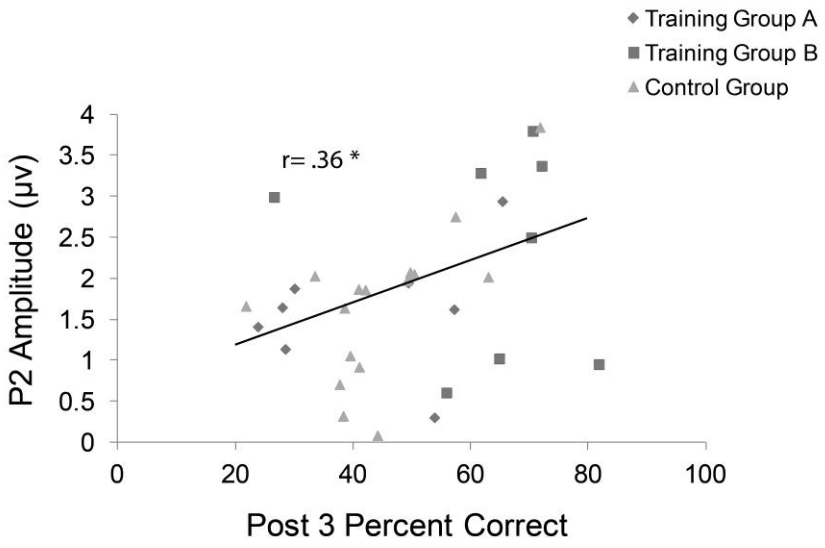

Figure 3. Scatter plot showing the relationship between P2 amplitude at Post-test three with percent correct at Post-test three.

Given that there were some Group differences in the behavioral and ERP data, additional analyses were done in order to measure the relationship between P2 amplitude and behavioral improvement. Further investigation confirmed this by showing there was a significant positive correlation between Post-test three test scores and Post-test three P2 mean amplitude, $r(30)= .36$, $p < .05$ (See Figure 4), such that those with larger P2 amplitude during Post-test three tended to score higher during that test session. This importantly suggests that the neural processing that was going on during the P2 had a large role in how a participant performed on the task.

Split by Trial Type. N1 and P2 components were then examined by Trial Type (i.e., Trained, Untrained, Not Trained trials) across Test Sessions in order to address patterns of activation that may be specific to generalization or rote learning. This analysis

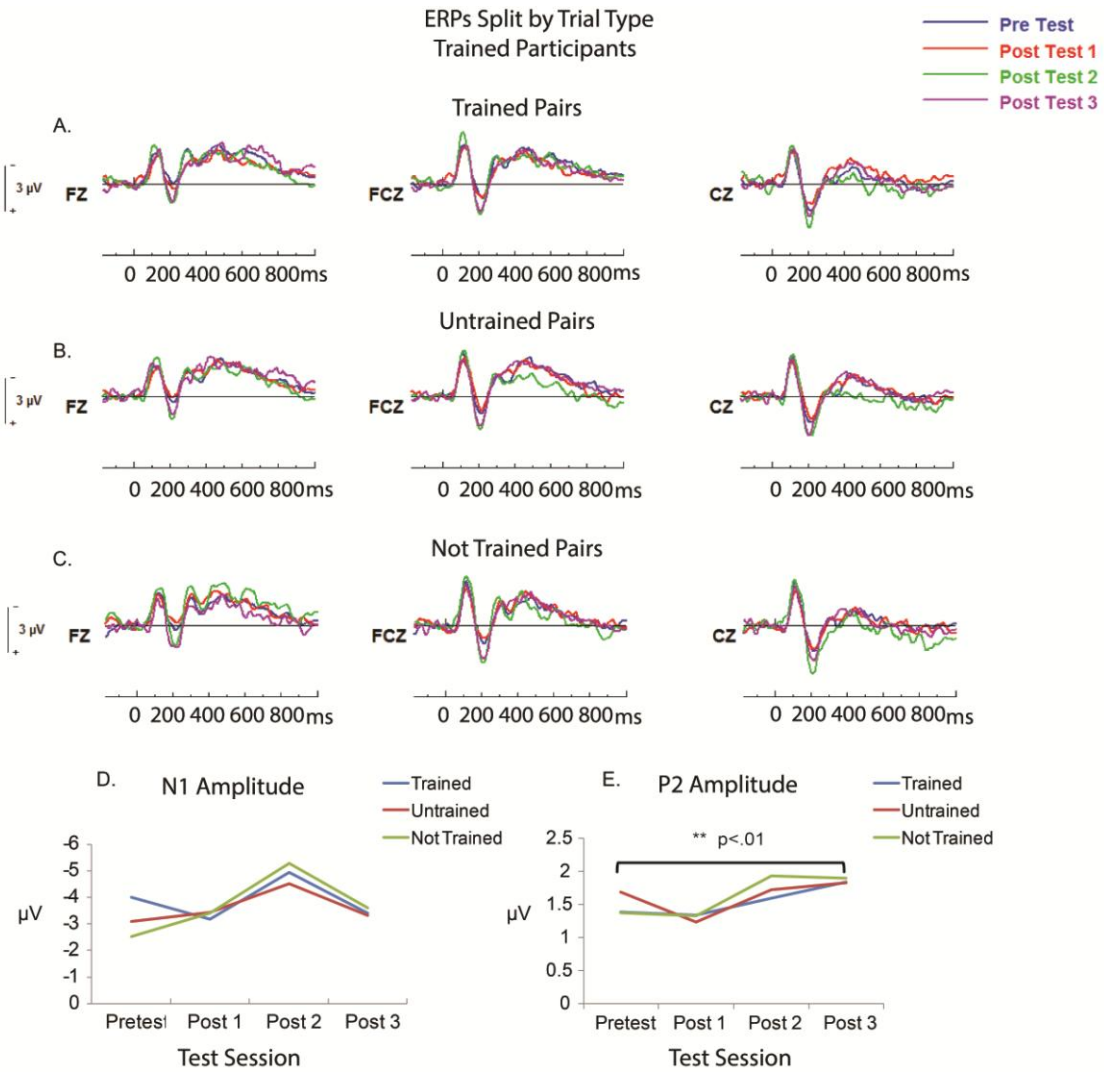

Figure 4. (A) Trained groups mean ERPs across Test Sessions for Trained trials. (B) Trained groups mean ERPs across Test Sessions for Untrained trials. (C) Trained groups mean ERPs across Test Sessions for Not Trained trials. (E) Average N1 mean amplitude for Trained Participants shown across Test Sessions as a function of Trial Type. (F) Average P2 mean amplitude for Trained Participants shown across Test Sessions as a function of Trial Type. Significant differences marked with a bracket and an asterisk. Bracket color will represent significant effects specific to a Trial Type (i.e., Blue = Trained Pairs, Red = Untrained Pairs, Green = Not Trained Pairs) or effects only found with Trial Type combined (i.e., Black= All Pairs).

was done with Trained participants only. No significant main effect was found for Test Session, although a marginally significant cubic trend was found for Test Session (p=.06) (See Figure 5D). This largely follows what was found when during analyses with Trial

Type combined, such that these amplitude modulations appear to reflect habituation (i.e., reduction in N1 amplitude from repeated stimulus presentation) rather than training related changes. In addition no main effect of Trial Type or significant interactions were found for Test Session and Trial Type (*p*'s >.05), which suggests there was the same pattern of increase and decrease in N1 amplitude across Test Session for each Trial Type (i.e., Trained, Untrained, Not Trained trials).

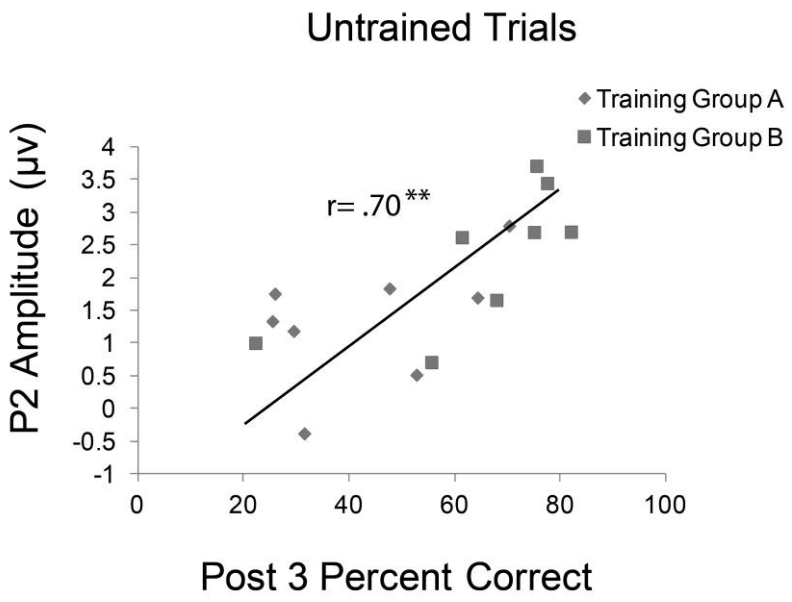

Figure 5. Scatter plot showing the relationship between P2 amplitude at Post-test three with percent correct at Post-test three for Untrained Trials for Trained Participants.

For the P2, a significant main effect of Test Session was found, $F(1,15) = 5.67$, $p < .01$, $\eta^2_p = .28$ (See Figure 5E). Polynomial contrasts showed that this was a significant linear trend $F(1,14) = 12.18$, $p < .01$, $\eta^2_p = .46$. Pairwise comparisons showed significant differences in P2 amplitude between Post-test one and Post-test two ($p<.01$), and between Pretest and Post-test three ($p<.05$). Again, no main effect of Trial Type or any interactions were found (See Figure 5E), which suggests that the pattern of amplitude

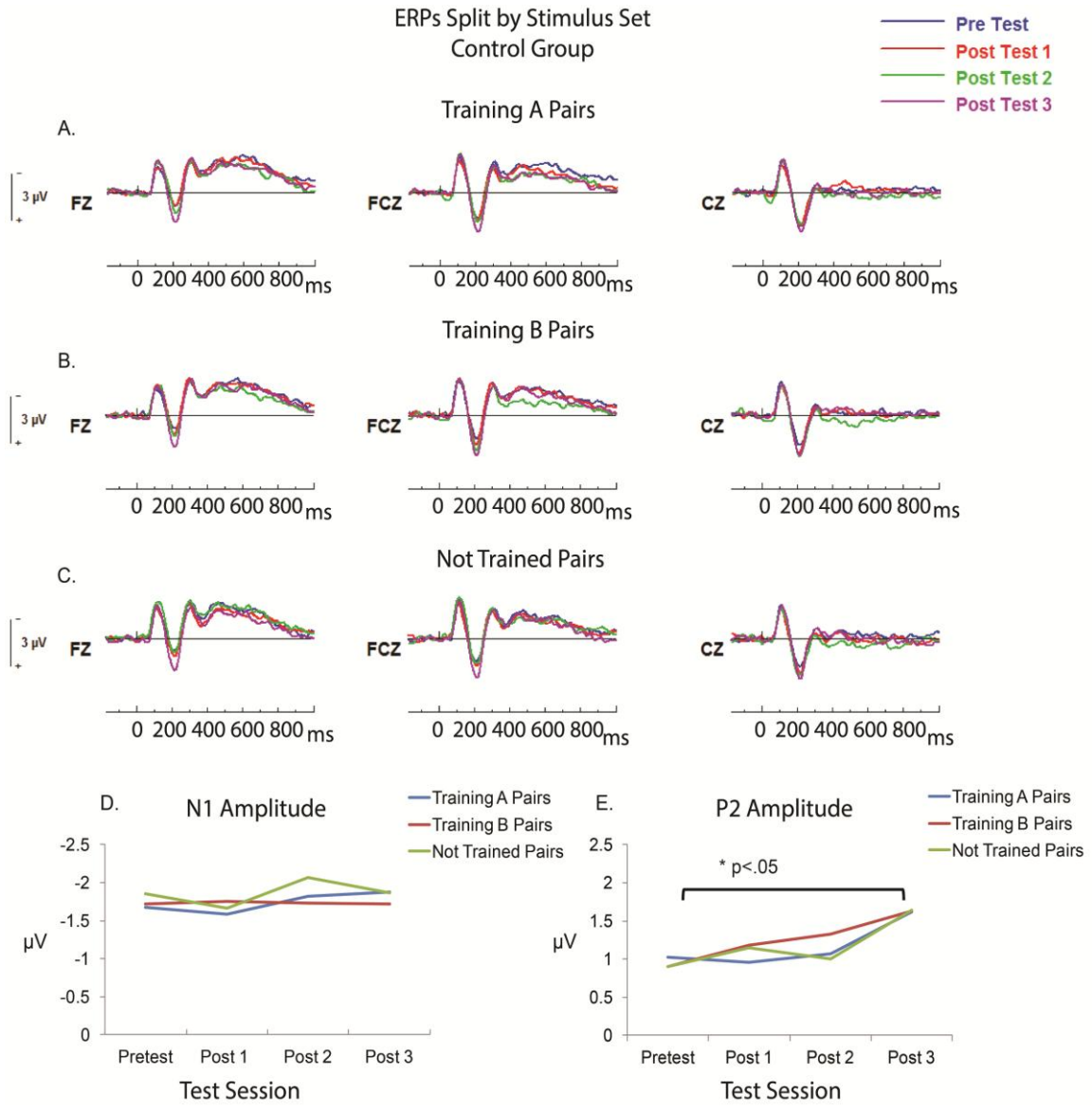

Figure 6. (A) Control Group mean ERPs across Test Sessions for Training A Pairs. (B) Control Group mean ERPs across Test Sessions for Training B Pairs. (C) Control Group mean ERPs across Test Sessions for Not Trained Pairs. (E) Average N1 mean amplitude for the Control Group, shown across Test Sessions as a function of Stimulus Set. (F) Average P2 mean amplitude for the Control Group, shown across Test Sessions as a function of Stimulus Set. Significant differences marked with a bracket and an asterisk. Bracket color will represent significant effects specific to a Stimulus Set (i.e., Blue = Training A Pairs, Red = Training B Pairs, Green = Not Trained Pairs) or effects only found with Stimulus Set combined (i.e., Black= All Pairs).

change was similar between Trial Types across Test Session. Therefore, there do not appear to be any distinct amplitude modulation patterns unfolding for rote learning versus

generalization. Since there were differences in performance on different Trial Types during behavioral analyses, and P2 amplitude was shown to be related to level of performance when Trial type was combined, further correlations were done to explore the relationship of the size of P2 amplitude and performance by Trial Type. A marginally significant positive correlation was found between Post-test three P2 amplitude during Trained trials and performance on Trained trials at Post-test three, $r(14) = .45$, $p = .07$. Next, a significant positive correlation was found between Post-test three P2 amplitude during Untrained trials and performance on Untrained trials at Post-test three, $r(14) = .70$, $p < .01$ (See Figure 6). Finally, the correlation between Post-test three P2 amplitude during Not Trained trials and performance on Not Trained trials at Post-test three was not significant, $r(14) = .38$, $p = .13$. Although the change in P2 amplitude across Test Sessions was similar for each Trial Type, the size of P2 amplitude was strongly correlated with performance, in particular for Untrained trials.

Split by Stimulus Set. Given that there were no interactions or different patterns in amplitude between Trial Types across Test Sessions for Trained participants, conclusions as to how generalization-learning-related changes (i.e., neural changes related to improvement on Untrained or Not Trained trials) differ from rote learning related changes are limited. To fully explore whether there are any different patterns in N1 or P2 amplitude change between the different Stimulus Sets, mean amplitude values for the Control Group were examined in a separate ANOVA with Test Session and Stimulus Set (i.e., Training A Pairs, Training B Pairs, Not Trained Pairs) as within-subject factors. No significant main effect of Test Session or Stimulus Set were found for N1 or P2 (p values >.05) (See Figure 7), although a significant linear trend was found for

P2 mean amplitude across Test Session, $F(1,15) = 14.23$, $p < .01$, $\eta^2_p = .48$. Pairwise comparisons showed that this was due to a significant increase in P2 mean amplitude from Pretest to Post-test three ($p<.01$). No interactions between Stimulus Set and Test Session were found, which suggests that the pattern of amplitude change is similar across the vowels in each Stimulus Set across Test Sessions.

Global Field Power Analyses

Combined Trial Type. In order to examine the strength of the neural response in an unbiased and reference free manner, GFP was calculated for the N1 and P2 time ranges used in previous analyses (i.e., 90-130 ms and 175-230 ms for the N1 and P2, respectively). Analysis of the N1 time range revealed a significant main effect of Test Session $F(3,87) = 5.35$, $p < .01$, $\eta^2_p = .15$, and a significant cubic trend, $F(1,29) = 13.63$, $p < .01$, $\eta^2_p = .32$. Pairwise comparisons revealed this was due to a significant decrease in GFP from Pretest to Post-test one ($p<.05$), a significant increase in GFP from Post-test one to Post-test two ($p<.01$), but no differences between Post-test two to Post-test three ($p>.05$), or between Pretest and Post-test three (See Figures 8A and 8E). These trends parallel what was found for the N1 during the ERP analyses, and likely reflect adaptation or habituation effects.

Next, analysis of the P2 time range showed a significant main effect of Test Session $F(3,87) = 3.789$, $p < .05$, $\eta^2_p = .11$, and a significant cubic trend, $F(1,29) = 5.85$, $p < .05$, $\eta^2_p = .16$. Pairwise comparisons revealed these effects were due to a significant reduction in GFP from Pretest to Post-test one ($p<.05$), a significant increase in GFP from Post-test-one to Post-test two ($p<.01$), but no differences between Post-test two and Post-

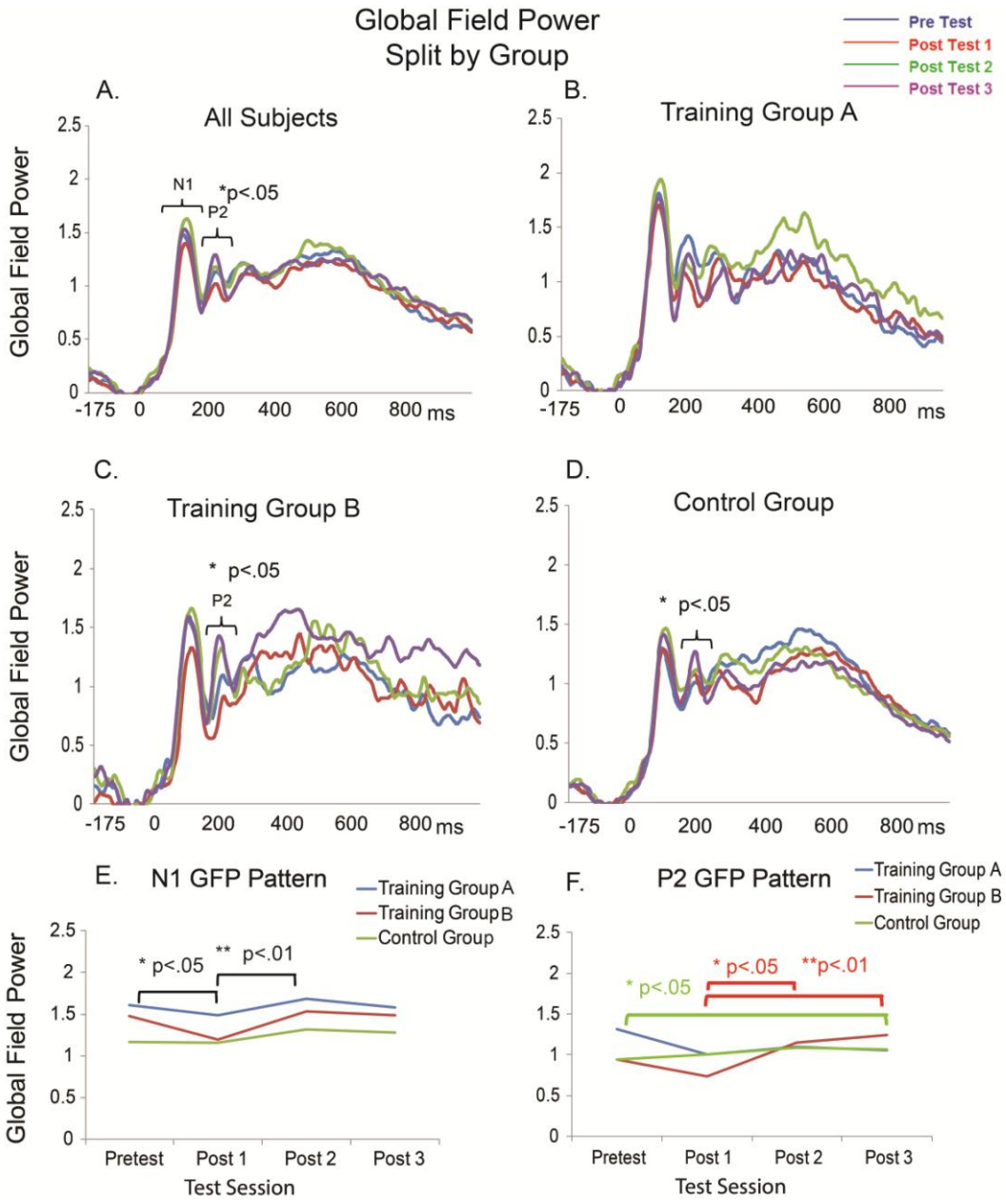

Figure 7. (A) Group GFP calculated as a function of time across Test Sessions. (B) Training Group A GFP calculated as a function of time across Test Sessions. (C) Training Group B GFP calculated as a function of time across Test Sessions. (D) Control Group GFP calculated as a function of time across Test Sessions. (E) Group average GFP during N1 time period (90-130 ms) as a function of Test Session. (F) Average GFP during P2 time period (175-230 ms) as a function of Test Session and Group. Significant differences marked with a bracket and an asterisk. Bracket color will represent significant effects specific to a Group (i.e., Blue = Training Group A, Red = Training Group B, Green = Control Group) or effects only found with Groups combined (i.e., Black= All Groups).

test three (p>.05). Also, a significant Test Session x Group interaction $F(6,87)=2.95$, $p<.05$, $\eta^2_p=.16$ was found, which was a result of differential patterns of change in strength of the neural response across Test Session for each Group (See Figure 8F).

Given that there were significant group differences in GFP across Test Sessions during the P2 time range, the groups were examined separately using three additional ANOVAs. No significant differences were found across Test Session for Training Group A ($p=.21$), although there was a marginally significant cubic trend $F(1,7)=3.81$, $p=09$, $\eta^2_p=.35$. A significant main effect of Test Session was found for Training Group B, $F(3,21)=5.17$, $p<.05$, $\eta^2_p=.42$. Polynomial contrasts showed that this was a linear trend, $F(1,7)=12.99$, $p<.01$, $\eta^2_p=.65$, such that GFP during the P2 range increased in power as each Test Session progressed. Pairwise comparisons revealed that this was due to a significant increase in GFP from Post-test one to Post-test two ($p<.05$), and from Post-test one to Post-test three ($p<.01$) (See Figure 8F). Lastly, no significant main effects were found for the Control Group's P2 GFP ($p=.27$), although a significant linear trend was found, $F(1,15)=5.93$, $p<.05$, $\eta^2_p=.28$, which was a result of a significant increase in GFP from Pretest to Post-test three ($p<.05$) (See Figure 8F). Since amplitude was the dependent measure of the earlier ERP analyses, it is not surprising that similar trends were observed here with GFP. Subsequent GMD analyses will provide information as to whether these power differences unfolded in conjunction with source configuration changes (See **Global Map Dissimilarity Results**).

Split by Trial Type. To examine potential GFP patterns specifically related to generalization or rote learning, GFP was calculated for Trained participants separately for

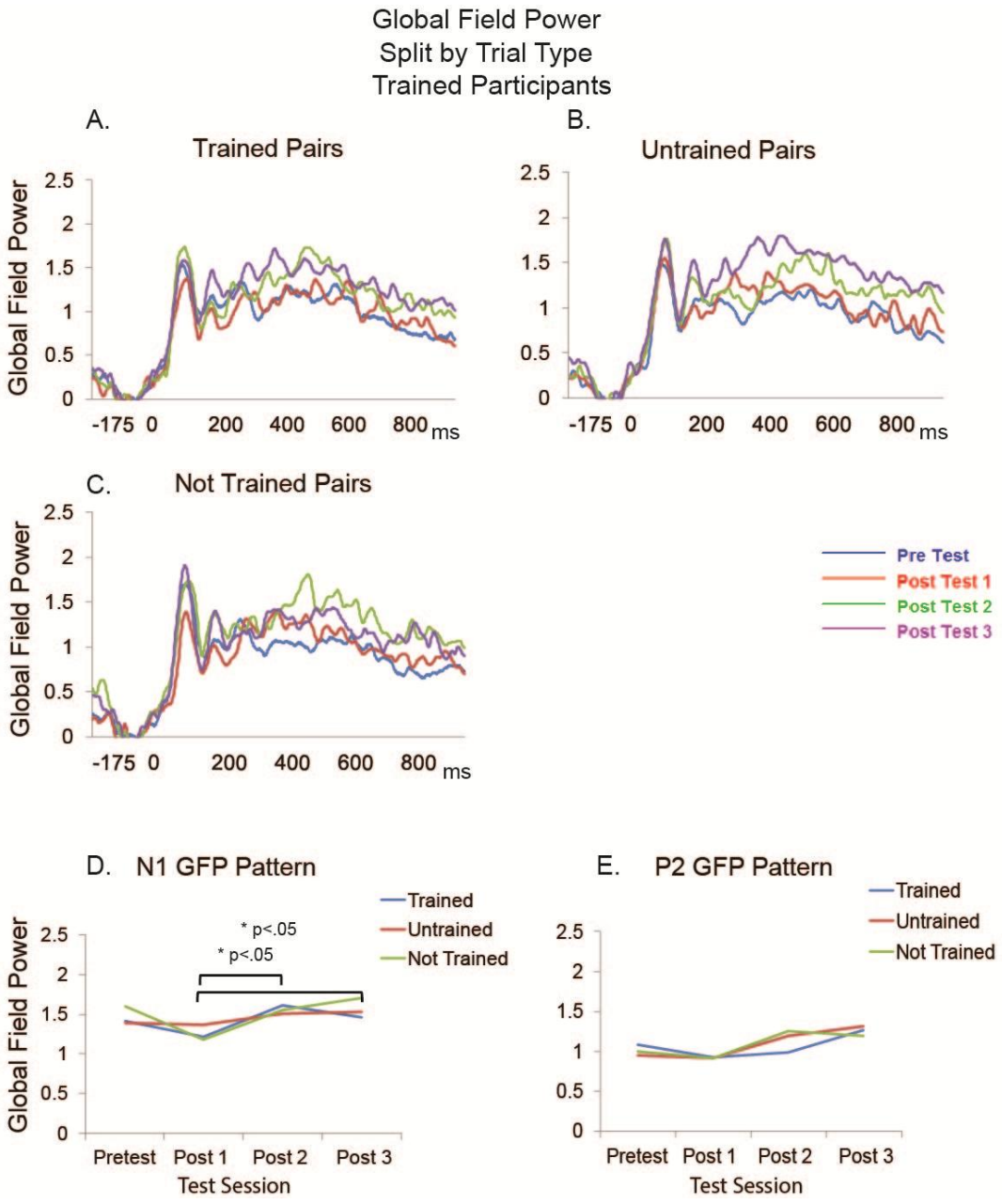

Figure 8. (A) Trained Groups GFP calculated as a function of time across Test Session for Trained trials. (B) Trained Groups GFP calculated as a function of time across Test Session for Untrained trials. (C) Trained Groups GFP calculated as a function of time across Test Session for Not Trained trials. (D) Trained Groups average GFP during N1 time period (90-130 ms) as a function of Test Session and Trial Type. (E) Trained Groups average GFP during P2 time period (175-230 ms) as a function of Test Session and Trial Type. Significant differences marked with a bracket and an asterisk. Bracket color will represent significant effects specific to a Trial Type (i.e., Blue = Trained Pairs, Red = Untrained Pairs, Green = Not Trained Pairs) or effects only found with Trial Type combined (i.e., Black= All Pairs).

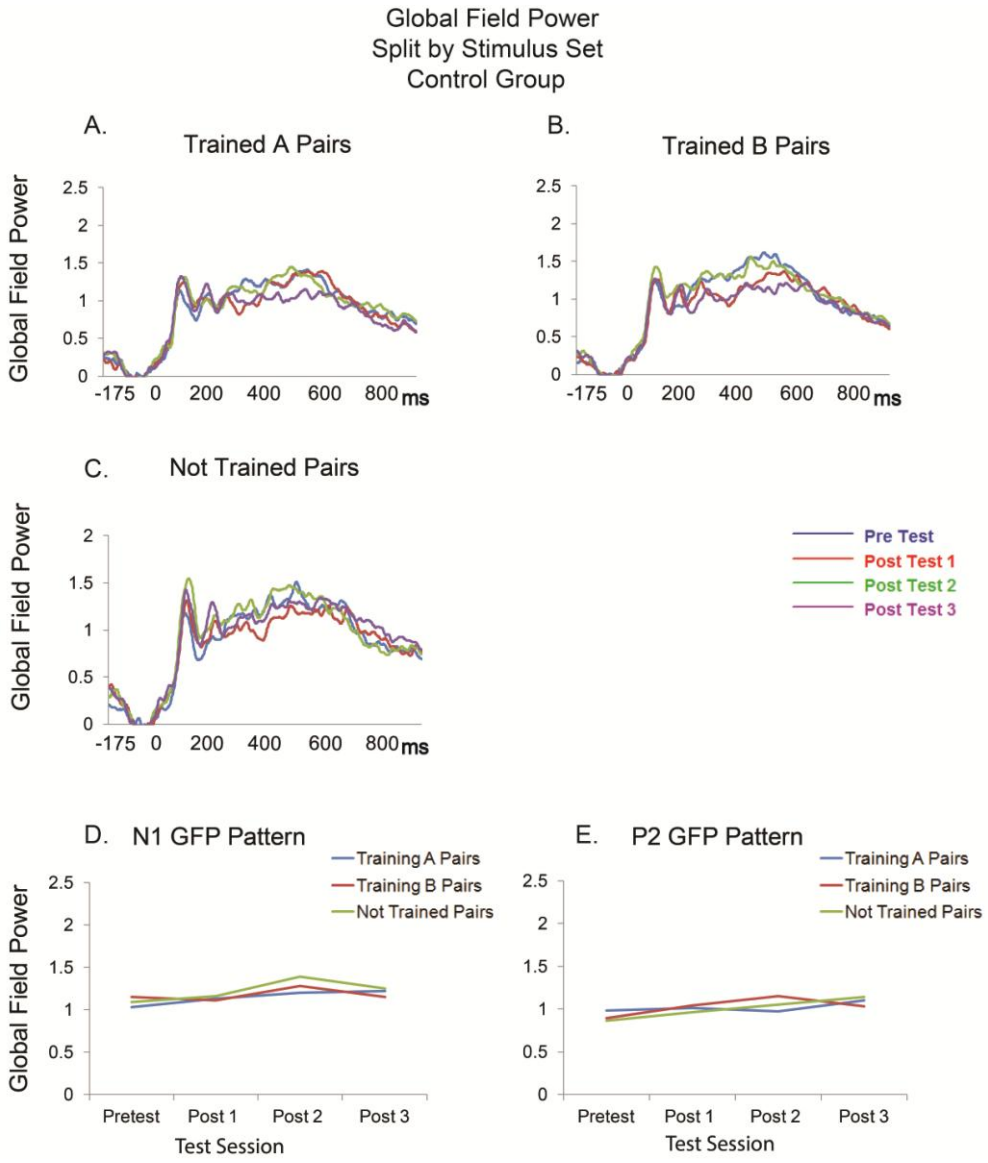

Figure 9. (A) Control Group GFP calculated as a function of time across Test Session for Training A Pairs. (B) Control Group GFP calculated as a function of time across Test Session for Training B Pairs. (C) Control Group GFP calculated as a function of time across Test Session for Not Trained Pairs. (D) Control Group average GFP during N1 time period (90-130 ms) as a function of Test Session and Stimulus Set. (E) Control Group average GFP during P2 time period (175-230 ms) as a function of Test Session and Stimulus Set. Significant differences marked with a bracket and an asterisk. Bracket color will represent significant effects specific to a Stimulus Set (i.e., Blue = Training A Pairs, Red = Training B Pairs, Green = Not Trained Pairs) or effects only found with Stimulus Set combined (i.e., Black= All Pairs).

N1 and P2 time ranges (i.e., 90-130 ms and 175-230 ms for the N1 and P2, respectively), but split by Group (Training Group A, Training Group B) and Trial Type (i.e., Trained,

Untrained, and Not Trained). This analysis was done for Trained participants only. For the N1 time range, a significant main effect of Test Session was found $F(3,42)= 4.04$, $p<.05$, $\eta^2_p=.22$ (See Figure 9D). Polynomial contrasts showed that this was a cubic trend, $F(1,15)= 6.68$, $p<.05$, $\eta^2_p=.32$. Pairwise comparisons were done to further breakdown the results, and showed a significant increase in GFP from Post-test one to Post-test two ($p<.05$), a significant increase from Post-test one to Post-test three ($p<.05$), but no difference between Pretest and Post-test three ($p>.05$), or between any other Test Sessions. No interactions were found for N1 GFP between Test Session and Trial Type (i.e., Trained, Untrained, Not Trained), limits any conclusion to be made about differential GFP patterns of generalization or rote learning (See Figure 9D); however, such trends follow what was found for N1 ERP analyses split by Trial Type. For the P2 time range, the main effect for Test Session was not significant, although it appears to be trending in that direction (p=.08). No polynomial contrasts or interactions with Trial Type or Group were significant (See Figure 9E).

Split by Stimulus Set. Next, the Control Group's GFP for N1 and P2 time periods was calculated for the Control Group alone with Test Session and Stimulus Set as within-subjects factors. For the N1, a non-significant main effect of Test Session ($p=.07$) and Stimulus Set ($p=.07$) were found. For the P2, no significant main effects of Test Session ($p=.11$) or Stimulus Set ($p=.79$) were found (See Figure 10). One possible reason for the lack of replication of ERP effects for the P2 may be a result of additional variance due to splitting GFP by Trial Type. GFP would be more sensitive to noise in the data as compared to ERP analyses, since GFP uses all of the electrodes and ERP analyses only use a small subset of electrodes.

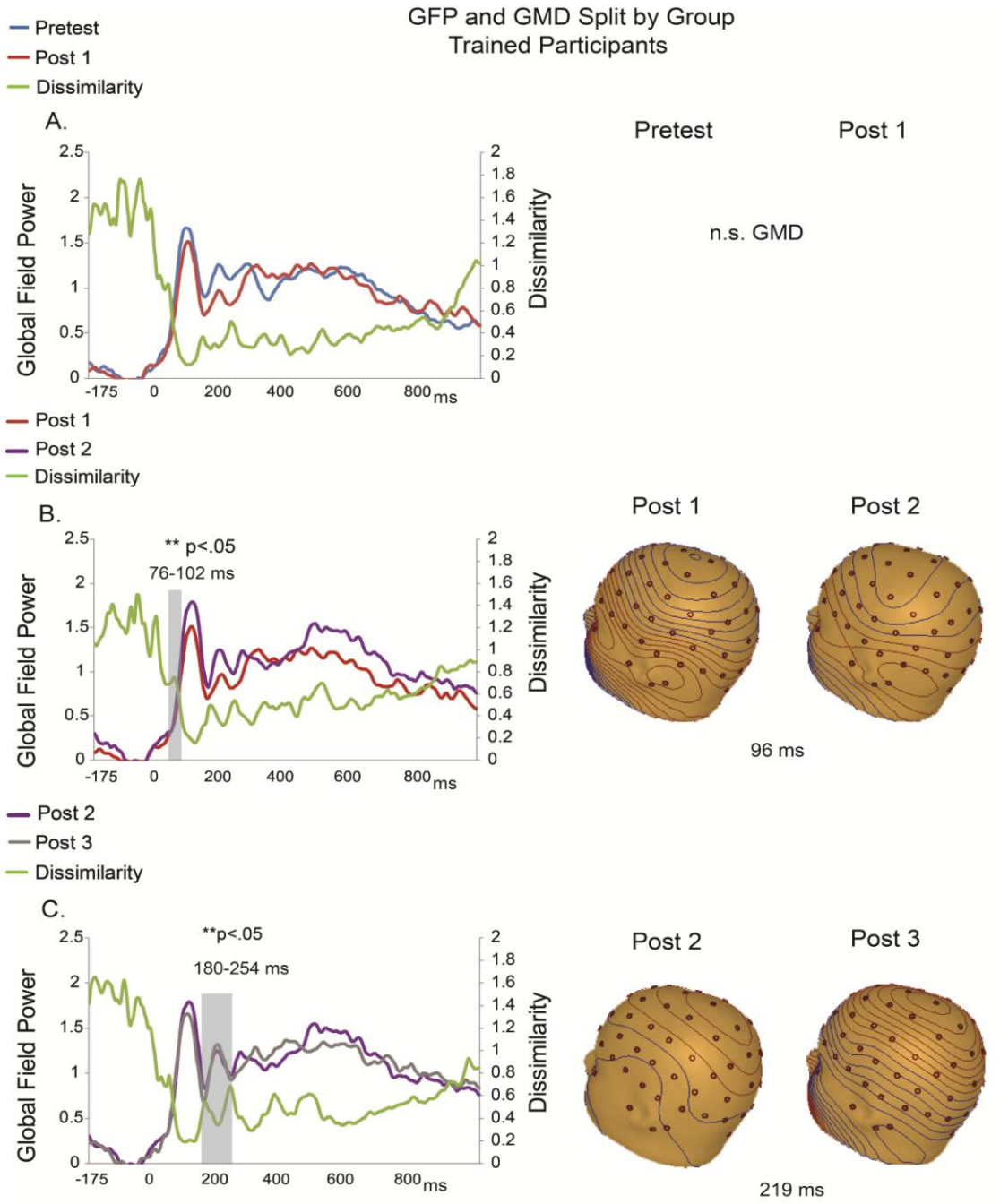

Figure 10. (A) Trained Groups GFP and Dissimilarity values for Pretest and Post 1 as a function of time. (B) Trained Groups GFP and Dissimilarity values for Post 1 and Post 2 as a function of time. (C) Trained Groups GFP and Dissimilarity values for Post 2 and Post 3 as a function of time. Time ranges with significant GMD differences, if any, marked with a grey box and asterisk. Topographies during a significant GMD time range shown to the right.

Global Map Dissimilarity Analyses

Combined Trial Type. To better understand whether the above GFP differences were additionally accompanied by differential source configuration, GMD was calculated for all Groups. Analyses were done for Trained participants (i.e., Training Group A and Training Group B pooled together) and the Control Group with Trial Type pooled (i.e., Trained, Untrained, and Not Trained). A summary of the results can be found in Table 3. Comparisons were done as follows to examine how voltage topographies changed as each Test Session progressed: Pretest vs. Post-test one, Post-test one vs. Post-test two, and Post-test two vs. Post-test three. For Trained participants, significantly different topographies were found between Post-test one and Post-test two (i.e., before sleep) during the N1 time range (76-102 ms) (See Figure 11B), such that the negative voltage distribution on the top of the head appears to have shifted to be more anterior (i.e., towards the front of the head) and left lateral, while the positive voltage distribution on the side of the head appears to have narrowed and shifted to be more posterior (i.e., towards the back of the head). Also, significantly different topographies were found between Post-test two and Post-test three (i.e., after sleep) during the P2 time range (180-254 ms) (See Figure 11C), which consisted of the positive voltage distribution shifting more anterior and becoming more narrow.

In addition, the Control Group had significantly different topographies between Post-test two and Post-test three (i.e., after sleep) just before the N1 time range (45-84 ms), and during the P2 time range (152-248) (See figure 12C). The topographical differences just before the N1 time range consisted of left lateralized shift of the negative

voltage distribution on the top of the head. The topographical differences during the P2 time range consisted of an anterior shift in the positive voltage on top of the head.

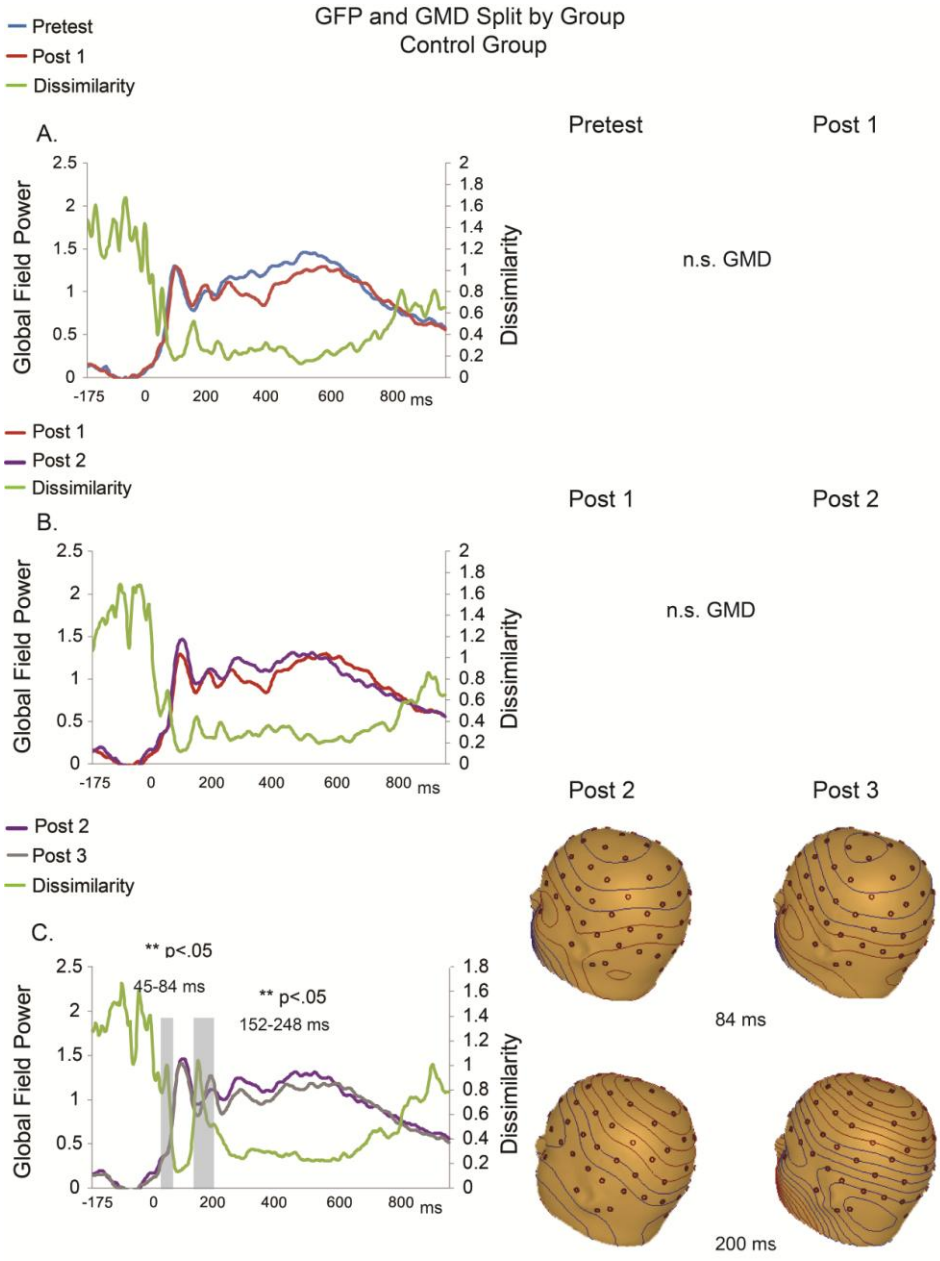

Figure 11. (A) Control Group GFP and Dissimilarity values for Pretest and Post 1 as a function of time. (B) Control Group GFP and Dissimilarity values for Post 1 and Post 2 as a function of time. (C) Control Group GFP and Dissimilarity values for Post 2 and Post 3 as a function of time. Time ranges with significant GMD differences, if any, marked with a grey box and asterisk. Topographies during a significant GMD time range shown to the right.

	Trained Participants	Control Group
Pretest vs. Post 1	n.s.	n.s.
Post 1 vs. Post 2	76-102 ms	n.s.
Post 2 vs. Post 3	180-254 ms	45-84 ms 152-248 ms

Table 3. Summary of the results from the GMD analysis for Trained Participants and the Control Group with Trial Type combined. Possible test comparisons and resulting significant time ranges listed. Comparisons which did not result in a significant topographical difference marked "n.s.".

Split by Trial Type. Next, GMD was calculated for Trained Participants with Trial Type split (i.e., Trained, Untrained, Not Trained). This was done to detect source configuration differences that occurred during rote learning (i.e., performance on Trained trials) and generalization (i.e., performance on Untrained or Not Trained trials). Test Sessions were compared as mentioned previously (i.e., Pretest vs. Post-test one, Post-test one vs. Post-test two, and Post-test two vs. Post-test three). Results are summarized in Table 4. No topographical differences were found for Trained trials during N1 or P2 time ranges when comparing Test Sessions, which suggests no source configuration changes as each session progressed.

On the other hand, significant differences were found for Untrained trials between Pretest and Post-test one (i.e., after training) during the N1 time range (84-104 ms) (See Figures 13D and 14A). Topographical map changes consisted of the negative voltage distribution on the top of the head becoming more narrow and the positive voltage distribution on the side of the head shifting more superior (i.e., upward). These changes represent a shift in the neuronal networks used to process Untrained trials, and may

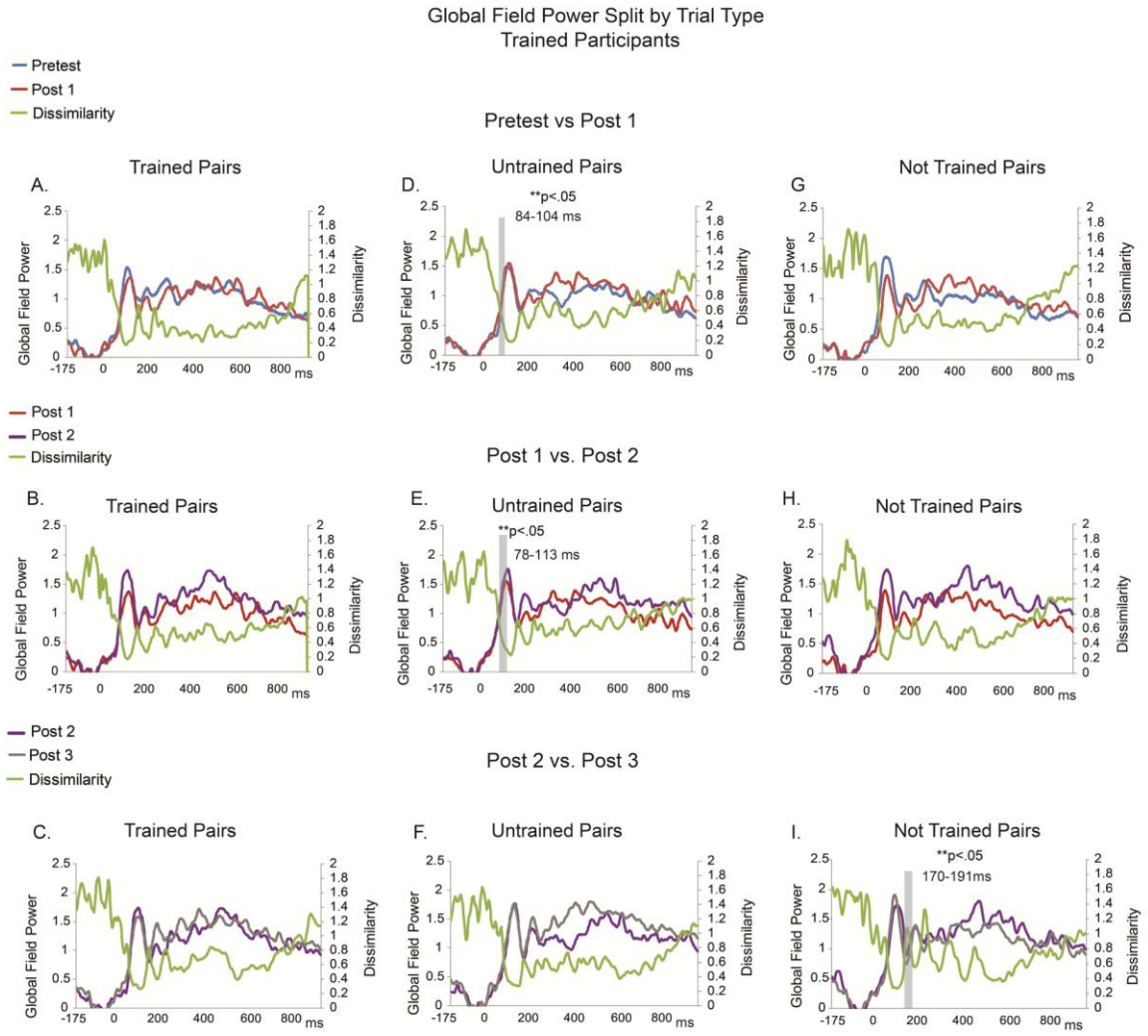

Figure 12. (A) Trained Groups GFP and Dissimilarity values for Pretest and Post 1 as a function of time for Trained trials. (B) Trained Groups GFP and Dissimilarity values for Post 1 and Post 2 as a function of time for Trained trials. (C) Trained Groups GFP and Dissimilarity values for Post 2 and Post 3 as a function of time for Trained trials. (D) Trained Groups GFP and Dissimilarity values for Pretest and Post 1 as a function of time for Untrained trials. (E) Trained Groups GFP and Dissimilarity values for Post 1 and Post 2 as a function of time for Untrained trials. (F) Trained Groups GFP and Dissimilarity values for Post 2 and Post 3 as a function of time for Untrained trials. (G) Trained Groups GFP and Dissimilarity values for Pretest and Post 1 as a function of time for Not Trained trials. (H) Trained Groups GFP and Dissimilarity values for Post 1 and Post 2 as a function of time for Not Trained trials. (I) Trained Groups GFP and Dissimilarity values for Post 2 and Post 3 as a function of time for Not Trained trials. Time ranges with significant GMD differences, if any, marked with a grey box and asterisk.

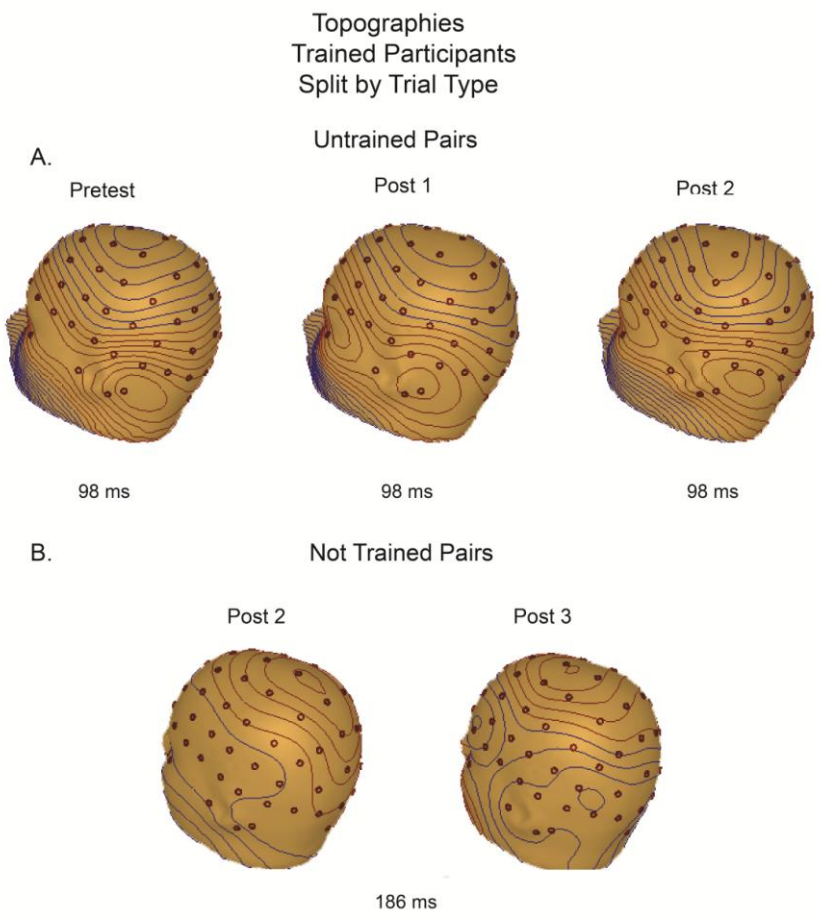

Figure 13. (A) Topographies shown that represent the time period which significant GMD was found between Pretest and Post 1, and Post 1 and Post 2 for Untrained trials. (B) Topographies shown that represent the time period which significant GMD was found between Post 1 and Post 2 for Not Trained trials. Shown for Trained participants only.

	Trained Trials	Untrained Trials	Not Trained Trials
Pretest vs. Post 1	n.s.	84-104 ms	n.s.
Post 1 vs. Post 2	n.s.	78-113 ms	n.s.
Post 2 vs. Post 3	n.s.	n.s.	170-191 ms 221-262 ms

Table 4**.** Summary of the results from the GMD analysis for Trained Participants when split by Trial Type. Possible test comparisons and resulting significant time ranges listed. Comparisons which did not result in a significant topographical difference marked "n.s.".

represent the neural changes that enabled an improvement on these trials from Pretest to Post-test one (i.e., generalization). An important point is that these changes occurred in the absence of any training on Untrained trials. Significant topographical map changes were also found for Untrained trials between Post-test one and Post-test two (i.e., before sleep) during the N1 time range (78-113 ms) (See Figure 13E and 14A). Map changes consisted of a left lateralized shift of the negative voltage on top of the head, and posterior and superior shift in the positive voltage on the side of the head.

Significant differences were also found for Not Trained trials between Post-test two and Post-test three (i.e., after sleep) during the P2 time range (170-191 ms) (See Figure 13I and 14B). Map changes consisted of an anterior shift in the positive voltage on top of the head. These two topographical map changes likely reflect source configuration differences that were made possible due to additional testing and experience with the vowel pairs.

Split by Stimulus Set. Finally, in order to examine whether there are any different trends between the Control Group and Trained Participants for the different Stimulus Sets, GMD was calculated for the Control Group alone but split by Stimulus Set. See Table 5 for a summary. For Training A Pairs, significantly different topographies were found near the N1 time range (78-100 ms) between Post-test one and Post-test two (i.e., before sleep) (See Figures 15B and 16A), and during the P2 time range (174-219 ms) between Post-test two and Post-test three (i.e., after sleep) (See Figures 15C and 16B). Topographical changes during the N1 time range consisted of an anterior shift in the negative voltage distribution on the top of the head and a superior shift in the positive

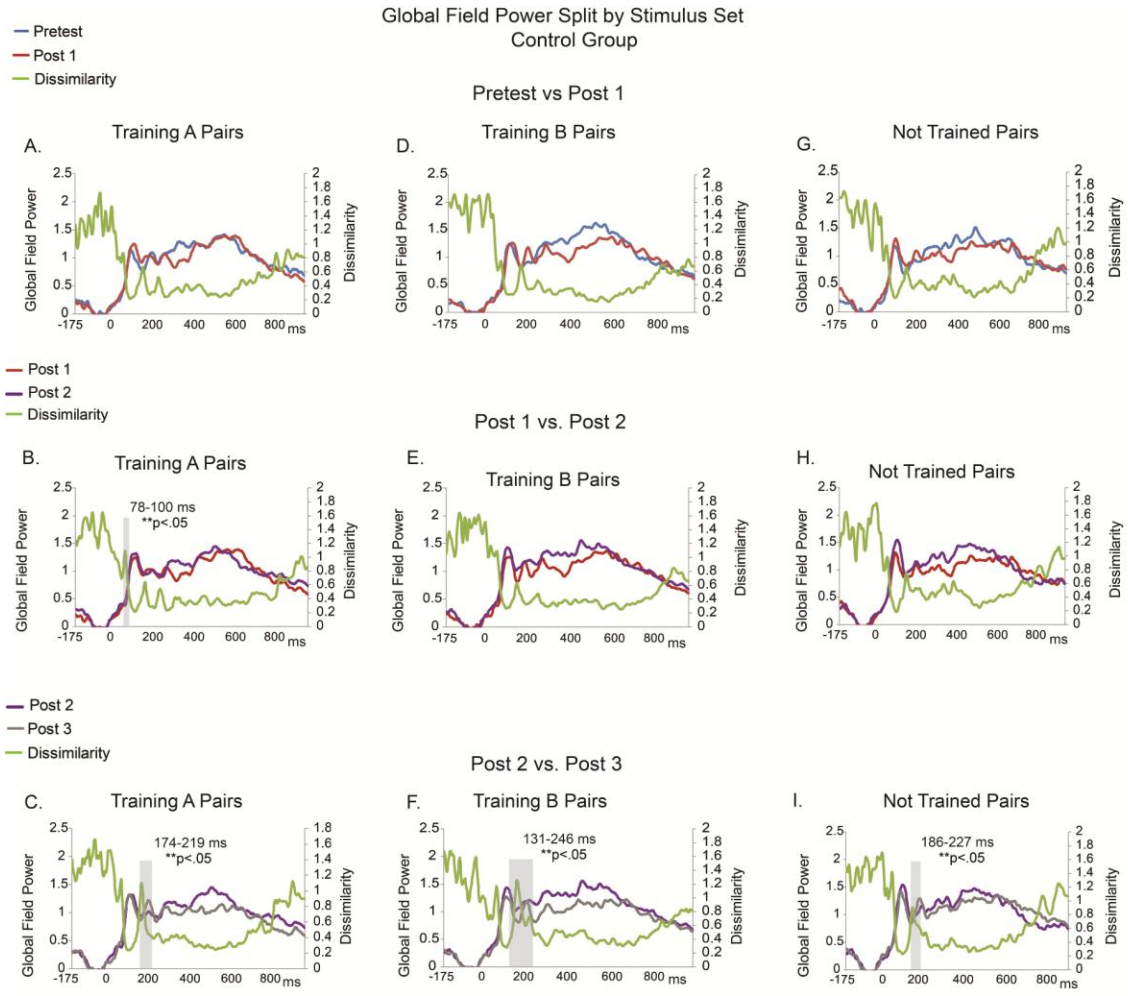

Figure 14. (A) Control Group GFP and Dissimilarity values for Pretest and Post 1 as a function of time for Training A pairs. (B) Control Group GFP and Dissimilarity values for Post 1 and Post 2 as a function of time for Training A pairs.. (C) Control Group GFP and Dissimilarity values for Post 2 and Post 3 as a function of time for Training A pairs. (D) Control Group GFP and Dissimilarity values for Pretest and Post 1 as a function of time for Training B pairs. (E) Control Group GFP and Dissimilarity values for Post 1 and Post 2 as a function of time for Training B pairs. (F) Control Group GFP and Dissimilarity values for Post 2 and Post 3 as a function of time for Training B pairs. (G) Control Group GFP and Dissimilarity values for Pretest and Post 1 as a function of time for Not Trained pairs. (H) Control Group GFP and Dissimilarity values for Post 1 and Post 2 as a function of time for Not Trained pairs. (I) Control Group GFP and Dissimilarity values for Post 2 and Post 3 as a function of time for Not Trained pairs. Time ranges with significant GMD differences, if any, marked with a grey box and asterisk.

voltage on the side of the head. Topographical changes during the P2 time range consisted of an anterior shift in the positive voltage on the head and an inferior shift in

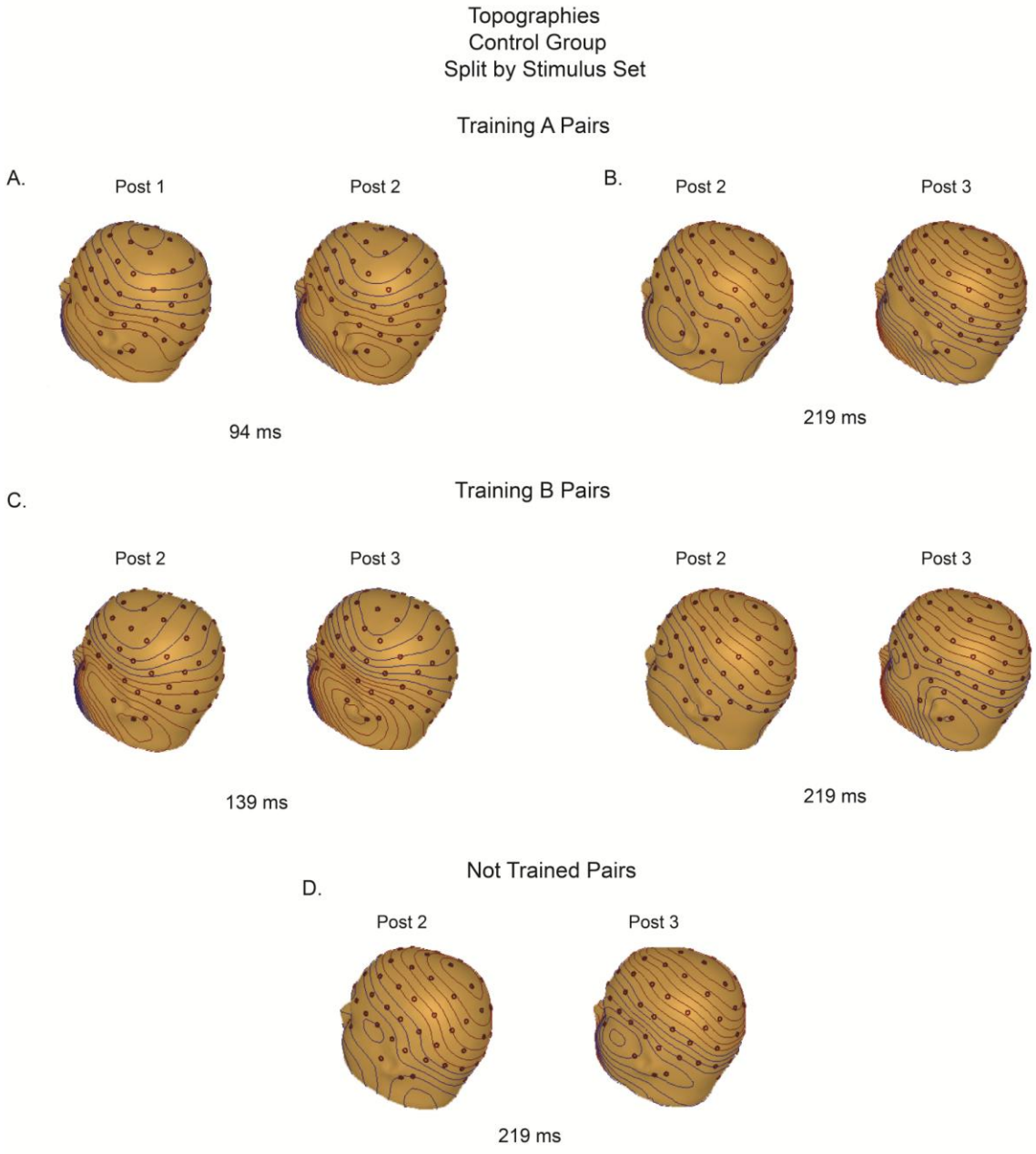

Figure 15. (A) Topographies shown that represent the time period which significant GMD was found between and Post 1 and Post 2 for Training A Pairs (B) Topographies shown that represent the time period which significant GMD was found between and Post 2 and Post 3 for Training A Pairs. (C) Topographies shown that represent the time period which significant GMD was found between Post 2 and Post 3 for Training B Pairs. Two exemplars are shown for this comparison due to the long time range that was found to be significantly different. (D) Topographies shown that represent the time period which significant GMD was found between Post 2 and Post 3 for Not Trained trials. Shown for Control group only.

	Training A Pairs	Training B Pairs	Not Trained Pairs
Pretest vs. Post 1	n.s.	n.s.	n.s.
Post 1 vs. Post 2	78-100 ms	n.s.	n.s.
Post 2 vs. Post 3	174-219 ms	131-246 ms	186-227 ms

Table 5**.** Summary of the results from the GMD analysis for the Control Group when split by Stimulus Set. Possible test comparisons and resulting significant time ranges listed. Comparisons which did not result in a significant topographical difference marked "n.s.".

the negative voltage on the side of the head. For Training B Pairs, significantly different topographies were found during a range that stretched from the end of the N1 time range to the end of the P2 time range (131-246) between Post-test two and Post-test three (i.e., after sleep). Due to the length of this time range, two time points are shown (See Figures 15F and 16C). Topographical changes near the end of the N1 time range consisted of an anterior left lateralized shift in the negative voltage on the top of the head, and a slight anterior shift in the positive voltage on the side of the head. Topographical changes during the P2 time range consisted of an anterior shift in the positive voltage on the top of the head. Finally, significantly different topographies were found for Not Trained trials during the P2 time range (186-227 ms) between Post-test two and Post-test three (i.e., after sleep) (See Figures 15I and 16D). Topographical changes during the P2 time range again consisted of an anterior shift in the positive voltage on the top of the head. Since the Control Group did not participate in any training, the differences seen as each Test Session progressed are likely the result of additional practice by testing (See Table 5 for a summary).

Follow up GMD Analyses

In order to examine whether there were any group GMD differences between Training Group A and Training Group B, the GMD analyses were done again separately for each group. Test sessions were compared in the same fashion as previous GMD analyses (i.e., Pretest vs. Post-test one, Post-test one vs. Post-test two, and Post-test two vs. Post-test three). Comparisons were first made between groups with Trial Type combined, then split by Trial Type. The analysis with Trial Type combined only showed one significant topographical difference which occurred between Post-test two and Post-test three (See Table 6), during the end of the P2 time range (229-256 ms). This difference was only found for Training Group A. No significant differences were found to occur within the N1 or P2 time range when making any of the Test Session comparisons with combined Trial Type for Training Group B. Next, the analysis with Trial Type split revealed only one significant topographical difference (211-256 ms). This difference occurred for Training Group B during Not trained trials when comparing Post-test two and Post test three when split by Trial Type (See Table 7).

A reason for the differences seen here as compared to previous analyses is likely due to a poor signal to noise ratio as a result of splitting the Trained Participants into two groups, and then further splitting by Trial Type, all which leaves very few ERP trials per analysis. Individual trial ERPs can vary quite a bit for a number of reasons. For instance, basal activity (background EEG noise) can be difficult to differentiate from event-related activity if a high enough trial number is not met. In addition metabolic constraints or varying refractory periods can also have an impact on the latency of neuronal firing. If too few trials are used, an effect may be washed out due to small latency differences

between trials. Despite these drawbacks, there were at least two of the same effects found here (Compare Table 2 to Table 5, and Table 3 to Table 6).

	Training Group A	Training Group B
Pretest vs. Post 1	n.s.	n.s.
Post 1 vs. Post 2	n.s.	n.s.
Post 2 vs. Post 3	229-256 ms	n.s.

Table 6. Summary of the results from the follow up GMD analyses for Training Group A and Training Group B. Possible test comparisons and resulting significant time ranges listed. Comparisons which did not result in a significant topographical difference marked "n.s.".

	Trained Pairs	Untrained Pairs	Not Trained Pairs
Pretest vs. Post 1	n.s.	n.s.	n.s.
Post 1 vs. Post 2	n.s.	n.s.	n.s.
Post 2 vs. Post 3	n.s.	n.s.	211-256 ms

Table 7. Summary of the results from the follow up GMD analysis for Training Group B when split by Trial Type. Possible test comparisons and resulting significant time ranges listed. Comparisons which did not result in a significant topographical difference marked "n.s.".

CHAPTER 5

DISCUSSION

Generalization and Speech Segregation

There were three main goals of this study. The first was to investigate whether listeners could learn to better segregate speech sounds by generalization, as has been seen with word/phoneme identification tasks (e.g., Bradlow & Bent, 2003; Eisner & McQueen, 2005; Greenspan, Nusbaum, & Pisoni, 1988). In order to address this, two training groups trained on one third of the total vowel pairs by receiving feedback on their performance, and then completed three post-tests where all vowel pairs were presented. Any improvement on Untrained vowel pairs at post-test would be considered a transfer of learning. A Control group also completed the pretest and post-tests but did not complete any training so that the true value of training could be seen by comparing the degree of improvement between those who trained (training effects) with those that did not (testing effects). All groups improved at the task which was evidenced by a significant increase in percent correct (See Figure 1B). Since participants in all groups did improve, no additional benefit of training was observed above testing effects. In fact, most of the group differences were seen as a result of the Training B participants learning more and scoring higher than other groups beginning at Pretest (See Figure 1A), rather than any specific effect of training overall. When this task has been used previously, no learning effects in the Control Group were found like the ones seen here (cf. Alain et al., 2007; Reinke et al., 2003), which is likely due to some methodological differences between past studies and the current one. For instance, the current study used twice the number of trials and did so within one 24 hour period, whereas a five day period

separated the two testing sessions (i.e., Pretest and Post-test) in Alain et al., 2007 and Reinke et al., 2003. The trial number difference was due to more vowel pairs being used here, which was done to improve the signal to noise ratio of the ERP data. Unfortunately, the presence of such large testing effects makes it difficult to make any clear inferences about generalization at later post-tests; however, the significant improvement in performance on Untrained vowel pairs at Post-test one for Training Group B does appear to be the best evidence of a transfer of learning here given that it occurred in the absence of additional exposure to the Untrained sounds (See Figure 1D). In order to make generalization learning effects more clear in future studies examining speech segregation, different stimuli should be used in each test session.

While such large performance differences between Groups were not expected, the presence of such effects has led to two interesting conclusions. First, the difficulty of the stimuli in a training set had a large effect on how much participants learned throughout the task and whether generalization occurred (See **Post hoc analyses**). It is possible that training with double vowels with larger frequency separations allowed for easier segregation which allowed participants to acquire a better representation of each vowel in a pair. This may have been more challenging when training with difficult trials. Future studies should consider the difficulty of a training set when creating future training regimens. A second interesting finding was the fact that there was not an added benefit of training over simply testing. This raises an important question as to whether feedback is truly necessary to improve at a task, or whether it is the case that completing a certain threshold of trials is sufficient for learning. Although previous studies (i.e., Alain et al., 2007; Reinke et al., 2003) have found that trained listeners scored higher in the vowel

segregation task compared to untrained listeners, it is not clear whether the same benefit could be seen by testing on the same amount of trials with no feedback. In a review article, Beste and Dinse (2013) report that learning without explicit training can be accomplished if exposure to a sound is extensive and occurs within a short enough time window. If the stimulus is not presented enough times then learning may not be accomplished. Therefore, it appears that learning can be accomplished by simply testing or training with feedback, although there might be a point where testing without being corrected (i.e., feedback on incorrect trials) would cease to be helpful. Without being corrected and having the opportunity to pair incoming acoustic information with the appropriate label, it is possible that participants may reach a point where advancement at a task is halted. Further studies should be done which in order to explore whether feedback on performance is indeed necessary to improve perception.

Neural Correlates of Generalization

A second goal was to explore how generalization learning is represented in the brain, and whether this differs from rote learning. To explore this, EEG was recorded during all tests and then first examined using traditional ERP analyses. N1 and P2 components were focused on since there have been numerous published studies which confirm their reliable malleability to learning. ERP analyses of N1 amplitude do not appear to reflect actual learning effects given that the cubic pattern of increase and decrease in amplitude does not parallel that of any observed learning patterns. Instead, they seem to better represent habituation effects. On the other hand, analyses of P2 amplitude do indeed seem to reflect the pattern of learning for each Group since the pattern of increase in amplitude is paralleled by a similar pattern of improvement in

scores on the vowel task (See Figures 1B and 3F). Furthermore, the size of P2 amplitude has a relation to how well a listener does on the task. This was evidenced by a significant positive correlation between P2 amplitude and scores at the final post-test, such that those with larger P2 amplitude tended to score higher on the task (See Figure 4). Unfortunately no differences were seen in the pattern of increase of N1 or P2 during ERP analyses as a function of Trial Type (i.e., Trained, Untrained, Not Trained) across Test Session (See Figures 5D and 5E); however, the size of P2 amplitude at Post-test three was highly correlated with Post-test three scores for Untrained trials (See Figure 6). This may reflect enhanced sensory processing which allowed for improvement on the task during these trials. It follows then that failure to improve on Untrained trials may be a result of failed sensory processing during this time range. While it is clear that the sensory processing which occurs during the P2 time range plays an important role in participants' ability to segregate the double vowels in this task, these traditional ERP analyses do not appear to show any processing differences during generalization that are distinct from what would be expected during rote learning.

Two additional complementary analyses were done on the EEG data to further breakdown neural processing and generalization during the vowel task: a Global Field Power (GFP) analysis and a Global Map Dissimilarity (GMD) analysis. GFP offers a reference free and unbiased measure of the strength of the neural response. On the other hand, GMD measures whether the scalp distribution of voltage differs between conditions, while ignoring amplitude differences. Used together they can be very useful since they make it possible to distinguish between true amplitude differences between conditions (i.e., GFP), and differences between conditions that are a result of different

source configurations (i.e., GMD). When examined separately the Control Group did not show any large differences in GFP for the N1 time range, however a significant increase in GFP was found during the P2 time range from Pretest to Post-test three (See Figure 8F). The GMD analysis revealed a significant difference in the topographical maps between Post-test two and Post-test three just before the N1 time range (45-84 ms) and during the P2 time range (152-248 ms) (See Figure 12C). Importantly, these analyses provide a more detailed breakdown of the processing changes that paralleled improvement. Specifically, while power differences appear to underlie behavioral improvement during the Test Sessions before a night of sleep, improvement observed at Post-test three appears to have occurred by additionally recruiting a different pattern of neural sources.

Next, the GFP analysis for Trained participants during the N1 time range revealed a significant increase in the power of the neural response from Post-test one to Post-test two. In addition, GMD revealed a significant difference in the topographical distribution of voltage during this same time range between Post-test one and Post-test two (See Figure 11B). As was discussed for the N1 during ERP analyses, it is not clear how much these GFP patterns reflect actual learning effects rather than habituation effects; however, the significant GMD differences may reflect neural changes that facilitated learning. For the P2 time range, the GFP analysis for Training Group A did not show significant power differences across Test Sessions, while those in Training Group B showed a significant increase in GFP from Post-test one to Post-test two and from Post-test one to Post-test three (See Figure 8F). Additionally, when Trained groups were pooled together GMD analyses showed a significant difference between the topographical maps of Post-test two

and Post-test three (See Figure 11C). When GMD was computed for Trained groups separately, significant differences were only seen for Training Group A between Post-test two and Post-test three (See Table 6). It is not clear whether differences were not found for Training Group B due to a loss of power from splitting the Groups during GMD analyses, or if the significant GMD when groups were pooled was truly driven by Group A. When reducing the number of trials in an analysis or average, EEG data can become incredibly noisy. Increasing noise would make it very difficult to detect topographical differences between conditions. Altogether, unlike the Control Group, the Trained participants appeared to recruit different neural resources and increased the power of the neural response both before and after a night of sleep. The difference between groups may be a result of the fact that the Trained subjects completed a training session where they had additional exposure to different vowels which may have resulted in greater potentiation of the neural circuits in auditory cortices.

It should also be noted that there were differences in the results found between certain GFP and ERP analyses (e.g., Compare figures 3F and 8F), although generally the same patterns were found in both analyses. A likely reason for this is the fact that GFP uses all the electrodes whereas only a few are included when examining ERP effects. It is possible that using too narrow a range of electrodes in analyzing ERPs may be too conservative. Using all of the electrodes importantly considers the entire scalp distribution all at once, which allows for an unbiased assessment of the brain state as a whole. On the other hand, GFP may not be suited to detect effects which are restricted to a very narrow scalp distribution. Additionally, GFP is more dramatically affected by noise since adding random variation across the scalp can drown out actual power

differences. Although there can be drawbacks to using GFP, the central benefit is that it can be used in conjunction with GMD to breakdown what can be known from traditional ERP analyses; however, these drawbacks should be considered in future studies when deciding whether GFP and GMD are appropriate for an experiment.

Although these findings shed more light as to what neural condition lead to improvement, they do not directly address generalization. Rote learning is defined here as enhanced perception that is due to repeated exposure or training. On the other hand, a transfer of learning is defined here as using what was learned during training to enhance the perception of an untrained or unencountered sound. From a neural perspective, repeated exposure to a sound or training strengthens the responding network and potentially results in larger amplitude of a neural response. It might be expected then that a transfer of learning after training would be possible as result of the recruitment of different neural sources during the processing of untrained material that were strengthened due to training. In order to address this possibility GMD was further used to examine how source configurations changed for each Trial Type (Trained, Untrained, Not Trained) across Test Session. As was mentioned above, the ideal Post-test to observe possible generalization in the current study is Post-test one which occurred directly after training. Interestingly, only during Untrained trials were significantly different topographical maps found between Pretest and Post-test one, which occurred during the N1 time period (See Figure 13D and 14A). This suggests the source configuration changed significantly when processing Untrained trials despite no additional training with Untrained sounds. Importantly, there was also an increase in percent correct during Untrained trials at Post-test one for Training Group B (See Figure 1D), suggesting

generalization. Not trained trials could also be a Trial Type where a transfer of learning could occur, however no improvement in scores were seen from Pretest to Post-test one and no differences in GMD were found from Pretest to Post-test one. If this effect is truly generalization, then it should only occur with those who trained. Indeed, the Control group did not show any significant changes in source configuration from Pretest to Post-test one for any Stimulus Set (See Figure 15), however, they do show ERP amplitude and GFP increases across Test Session which accompanied behavioral improvement. While the behavioral data and GMD analyses for Untrained trials appear to suggest that Trained participants were indeed able to generalize at Post-test one, future studies which control for testing effects should be done to confirm this.

A drawback of GMD is that while it can be used to examine whether different sources are active during different conditions, it does not provide information about exactly where the different sources originate. Therefore, it is not clear exactly what areas of the brain were active that made a transfer of learning possible. Since they occurred during the time periods in which the N1 and P2 occur it may be that changes occurred in the brain regions responsible for generating these components (i.e., auditory cortices). Segregating and identifying two concurrent vowels requires the encoding and grouping of appropriate peaks of energy, otherwise referred to as harmonics. If the appropriate harmonics for two vowels are not grouped accurately, then the representation for both of them would not be distinct enough and would likely result in difficultly identifying one or both of the vowels. It is possible certain vowels have harmonics in close enough ranges that one population of neurons may process different harmonics of two different vowels. If this is true, then training on one set of vowels may inadvertently strengthen the neurons

that process both the harmonics of Trained and Untrained vowels. Importantly, such an explanation aligns the current studies data with previous trends concerning generalization, specifically the finding where a transfer of learning was best elicited when the stimuli set was large and highly variable (cf. Bradlow & Bent, 2003; Eisner & McQueen, 2005). Consistency between studies may suggest that a general mechanism that generalization with speech sounds is accomplished; specifically that tuning low level sensory areas can facilitate processing of new unencountered speech sounds.

Sleep-dependent Generalization

The final goal of this study was to address whether sleep was necessary in order to consolidate generalization, as was found in Fenn et al., 2003. Given that in this study it is difficult to separate generalization (i.e., improvement on Untrained trials) from learning from exposure at test sessions after Post-test one, questions about sleep dependent generalization cannot be addressed with this data set. Although this specific question cannot be answered, it can be addressed whether learning in general was dependent on or additionally benefited by sleep. Statistical analyses revealed a marginally significant increase in performance from Pretest to Post-test one, no difference between Post-test one and Post-test two, and a significant increase in performance at Post-test three after a night of sleep; therefore, it appears that sleep does benefit learning on this task given that no additional improvement was observed during the day and a large degree of improvement occurred after a night of sleep. One caveat is the variability in the data as a function of Group, which may suggest that learning depends more on the individual and the difficulty of the Trial Type or training material. For instance, when the data for Trained Participants is split by Group and Trial Type it is clear that performance

continues to improve on Not Trained Trials and Untrained Trials between Post-test one and Post-test two before sleep. In addition, those in Group B appear to steadily improve between all post-tests, while the other groups showed improvement at Post-test one and after a night of sleep at Post-test three. While sleep seems to offer some benefit, performance on this task seems better predicted by time and additional practice. A limitation of the current study is that there were not varied start times at which participants began the study. Fenn et al. (2003) used several groups, each which began testing in the morning or at night and then retested after a night of sleep or an equal period of time awake. It is possible that if the same design were adopted in the future that an effect of sleep may be observed, however, the data appear to suggest that learning to segregate speech sounds does not require sleep to consolidate, at least with the amount of trials used in the current study in a single 24 hour period.

Conclusions

In summary, the data from the current study allow several conclusions to be made. First, while training does not seem essential for improvement at the double vowel task, it is clear that training or practice through testing can help participants' performance. Behavioral performance and data from GMD analyses for Untrained trials at the first post-test may be evidence of generalization, although the significant testing effects that were observed during subsequent testing sessions leave room for doubt on this conclusion. Future studies should improve on the current paradigm by using unique stimuli in each testing session to limit rote learning and better address generalization. Next, sleep does not appear necessary to consolidate learning on this task. Instead, learning appears dependent on the difficulty of the training set, additional practice

through testing, and the number of years spent training as a musician. Those interested in creating training regimens could use this information to tailor training sessions to fit the experience level and need of a listener.

APPENDIX I: DEMOGRAPHIC QUESTIONNAIRE

Participant Code: _____

Screening Questions

Please answer the following questions completely and honestly.
All of your responses will remain confidential.

1. Birth Date ___/___/_____
   Month Day Year
2. Age? _____
3. What is the first language you learned? _____
4. What is your country of birth? _____
5. Have you lived outside of the U.S.? **Yes No**
   Where? _____

   For how long? _____
6. If English is not your first language, at what age did you begin learning English?

   _____
7. Gender **Male Female**
8. Are you left handed, right handed, or ambidextrous? **Left Right Ambidextrous**
9. Do you have any hearing loss (hearing aid)? **Yes No**
   What type of loss?_____

   For how long? _____
10. Do you have experience learning or playing music (including singing)? **Yes No**
    If **Yes**, list instruments, number of years for each, and number of years in formal training for each:_____

    _____

    _____
11. How many hours of music do you listen to every week? _____

12. What types of music do you listen to, typically?

_____

_____

_____

13. Have you ever had a head injury (e.g., automobile accident, fall, sports injury)?
Yes No
14. Have you ever or do you now have seizures? **Yes No**
15. Have you ever been unconscious? **Yes No** If so, for how long? _____
16. Do you have any neurological disorders? **Yes No** (*please describe*) _____

_____

_____

17. Have you ever had any kind of brain surgery? **Yes No** If yes, type: _____
18. Do you have any medical conditions (including substance abuse)? **Yes No**
(*please describe*) _____

_____

19. Have you been diagnosed with any mental or psychiatric disorder? **Yes No**
(*please describe*) _____

_____

APPENDIX II: SLEEP QUESTIONNAIRE

PARTICIPANT ID: _____ DATE:_____

REMINDERS:

- □ No Alcohol or Drug use starting night before
- □ No caffeinated beverages for 48 hours prior to the experiment
- □ No previous exposure to our vowel-pair test

1.) What time do you usually go to sleep?

2.) Do you fall asleep easily?

3.) How deeply do you sleep? (Light, Medium, Deep)

4.) What time do you usually wake up?

5.) Do you usually feel well rested?

6.) Have you ever sought medical attention for a sleep disorder?

7.) Do you have any disabilities that disrupt your sleep?

8.) Are you currently taking any medications to help you sleep?

9.) Are you taking any other medications? (excluding oral contraceptives)

10.) Do you have a history of substance abuse or diagnosed major mental illness?

11.) How many caffeinated beverages do you drink each day?

APPENDIX III: SLEEP LOG

Instructions

Please complete one page of this activity log each day. The Morning Sleep Log should be completed each morning right after you wake up.

Then, take this log with you and make entries to the Daily Activity Log as your day progresses. We ask you to keep this activity log with you during the course of the day so that you can make entries while your memories and experiences are still fresh.

The following is an alertness scale. When filling out the daily log, please use this scale to describe how alert you feel on the Daily Activity Log.

	Score
Feeling active, vital, alert, or wide awake	**1**
Functioning at high levels, but not at peak; able to concentrate	**2**
Awake, but relaxed; responsive but not fully alert	**3**
Somewhat foggy, let down	**4**
Foggy; losing interest in remaining awake; slowed down	**5**
Sleepy, woozy, fighting sleep; prefer to lie down	**6**
No longer fighting sleep, sleep onset soon; having dream-like thoughts	**7**
Asleep	**X**

The Daily Activity Log may cover parts of the day when you were asleep. After you wake up, mark those parts of the day as “asleep”.

You may complete this log by typing entries in on your computer, or printed with handwritten entries. If you choose to print this log on paper, please staple this instruction page to the rest of your log. You will need to refer to the above chart throughout the day.

If you have any questions, please don't hesitate to ask! You can email me at irsikv@unlv.nevada.edu or call 895-2951

If you have any questions about your rights in this study, please contact the Unlv's IRB office at 895-2794

Morning Sleep Log

- What time did you go to sleep last night? (circle the closest time)

  9:00p 9:30p 10:00p 10:30p 11:00p 11:30p 12:00a 12:30a other_____

- What time did you wake up this morning? (circle the closest time)

  6:00a 6:30a 7:00a 7:30a 8:00a 8:30a 9:00a 9:30a other_____

How restful was your sleep? (mark an X on the line below)

Not at all restful |----------| Very restful

0% restful 50% restful 100% restful

Daily Activity Log

	Where are you?	What are you doing?	How alert do you feel? (see 7-point scale in instructions)
Midnight-3:00am			
3:00am-6:00am			

6:00am-9:00am			
9:00am-Noon			
Noon - 3:00pm			
3:00pm-6:00pm			
6:00pm-9:00pm			
9:00pm-Midnight			

Did you drink anything **caffeinated** (ex: coffee, tea, energy drinks) or **alcoholic** (ex: beer, wine, liquor) today?

Kind of drink	Time	How much did you drink?

APPENDIX IV: IRB APPROVAL

UNLV
UNIVERSITY OF NEVADA LAS VEGAS

Biomedical IRB – Expedited Review
Approval Notice

NOTICE TO ALL RESEARCHERS:

Please be aware that a protocol violation (e.g., failure to submit a modification for any change) of an IRB approved protocol may result in mandatory remedial education, additional audits, re-consenting subjects, researcher probation, suspension of any research protocol at issue, suspension of additional existing research protocols, invalidation of all research conducted under the research protocol at issue, and further appropriate consequences as determined by the IRB and the Institutional Officer.

DATE: July12, 2012

TO: **Dr. Joel Snyder, Psychology**

FROM: Office of Research Integrity - Human Subjects

RE: Notification of IRB Action
Protocol Title: **The Role of Sleep in Generalization of Auditory Learning**
Protocol #: 1205-4163M
Expiration Date: July 11, 2013

This memorandum is notification that the project referenced above has been reviewed and approved by the UNLV Biomedical Institutional Review Board (IRB) as indicated in Federal regulatory statutes 45 CFR 46 and UNLV Human Research Policies and Procedures.

The protocol is approved for a period of one year and expires July 11, 2013. If the above-referenced project has not been completed by this date you must request renewal by submitting a Continuing Review Request form 30 days before the expiration date.

PLEASE NOTE:
Upon approval, the research team is responsible for conducting the research as stated in the protocol most recently reviewed and approved by the IRB, which shall include using the most recently submitted Informed Consent/Assent forms and recruitment materials. The official versions of these forms are indicated by footer which contains approval and expiration dates.

Should there be *any* change to the protocol, it will be necessary to submit a **Modification Form** through ORI - Human Subjects. No changes may be made to the existing protocol until modifications have been approved by the IRB. Modified versions of protocol materials must be

used upon review and approval. Unanticipated problems, deviations to protocols, and adverse events must be reported to the ORI – HS within 10 days of occurrence.

If you have questions or require any assistance, please contact the Office of Research Integrity - Human Subjects at IRB@unlv.edu or call 895-2794.

REFERENCES

Alain, C., Dyson, B.J., Snyder, J.S. (2006). Aging and the perceptual organization of sounds: A change of scene?. Handbook of Models for the Study of Human Aging, Elsevier Academic Press, Amsterdam, pp. 759–769.

Alain, C., Snyder, J. S., He, Y., & Reinke, K. S. (2007). Changes in auditory cortex parallel rapid perceptual learning. *Cerebral Cortex, 17*(5), 1074-1084.

Bamiou, D. E., Werring, D., Cox, K., Stevens, J., Musiek, F. E., Brown, M. M., & Luxon, L. M. (2012). Patient-reported auditory functions after stroke of the central auditory pathway. *Stroke, 43*(5), 1285-1289.

Bernstein, L. E., Auer, E. T., Jr., Eberhardt, S. P., & Jiang, J. (2013). Auditory Perceptual Learning for Speech Perception Can be Enhanced by Audiovisual Training. *Frontiers in Neuroscience, 7*, 34.

Beste, C., & Dinse, H., R. (2013). Learning without Training. *Current Biology*, 23(11), 489-499.

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. (1999). Training Japanese listeners to identify English /r/ and /l/: long-term retention of learning in perception and production. *Perception & Psychophysics, 61*(5), 977-985.

Bradlow, A. R., and Bent, T. (**2003**). Listener adaptation to foreign-accented English. *Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona, Spain, 2003*, 2881–2884

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual

learning on speech production. *Journal of the Acoustical Society of America, 101*(4), 2299-2310.

Bosnyak, D. J., Eaton, R. A., & Roberts, L. E. (2004). Distributed auditory cortical representations are modified when non-musicians are trained at pitch discrimination with 40 Hz amplitude modulated tones. *Cerebral Cortex, 14*(10), 1088-1099.

Conte, F., & Ficca, G. (2013). Caveats on psychological models of sleep and memory: a compass in an overgrown scenario. *Sleep Medicine Reviews, 17*(2), 105-121.

Demany, L. (1985). Perceptual-Learning in Frequency Discrimination. *Journal of the Acoustical Society of America, 78*(3), 1118-1120.

Diekelmann, S., Born, J., & Wagner, U. (2010). Sleep enhances false memories depending on general memory performance. *Behavioural Brain Research, 208*(2), 425-429.

Diekelmann, S., Landolt, H. P., Lahl, O., Born, J., & Wagner, U. (2008). Sleep loss produces false memories. *Plos One, 3*(10), e3512.

Diekelmann, S., Wilhelm, I., & Born, J. (2009). The whats and whens of sleep-dependent memory consolidation. *Sleep Medicine Reviews, 13*(5), 309-321.

Eisner, F., & McQueen, J. M. (2005). The specificity of perceptual learning in speech processing. *Perception & Psychophysics, 67*(2), 224-238.

Fenn, K. M., Gallo, D. A., Margoliash, D., Roediger, H. L., 3rd, & Nusbaum, H. C. (2009). Reduced false memory after sleep. *Learning & Memory, 16*(9), 509-513.

Fenn, K. M., Nusbaum, H. C., & Margoliash, D. (2003). Consolidation during sleep of perceptual learning of spoken language. *Nature, 425*(6958), 614-616. doi:

Francis, A. L., Baldwin, K., & Nusbaum, H. C. (2000). Effects of training on attention to acoustic cues. *Perception & Psychophysics, 62*(8), 1668-1680.

Gaab, N., Paetzold, M., Becker, M., Walker, M. P., & Schlaug, G. (2004). The influence of sleep on auditory learning: a behavioral study. *Neuroreport, 15*(4), 731-734.

Gaspar, J., G., Neider, M., B., Simons, D., J., McCarley, J.,S., Kramer, A.,F. (2013). Change detection: Training and transfer. *PLOS*, 8(6), 1-7.

Gilbert, C. D. (1994). Learning. Neuronal dynamics and perceptual learning. *Current Biology, 4*(7), 627-629.

Gottselig, J. M., Hofer-Tinguely, G., Borbely, A. A., Regel, S. J., Landolt, H. P., Retey, J. V., & Achermann, P. (2004). Sleep and rest facilitate auditory learning. *Neuroscience, 127*(3), 557-561.

Greenspan, S. L., Nusbaum, H. C., & Pisoni, D. B. (1988). Perceptual-Learning of Synthetic Speech Produced by Rule. *Journal of Experimental Psychology-Learning Memory and Cognition, 14*(3), 421-433.

Grieco-Calub, T. M., Saffran, J. R., & Litovsky, R. Y. (2009). Spoken Word Recognition in Toddlers Who Use Cochlear Implants. *Journal of Speech Language and Hearing Research, 52*(6), 1390-1400.

Hawkey, D. J., Amitay, S., & Moore, D. R. (2004). Early and rapid perceptual learning. *Nature Neuroscience, 7*(10), 1055-1056.

Huber, R., Ghilardi, M., F., Massimini, M., & Tononi, G. (2004). Local sleep and learning. *Nature*. 430(6995), 78-81.

Huber, R., Esser, S., K., Ferrarelli, F., Massimini, M., Peterson, M., J., Tononi, G. (2007). TMS induced cortical potentiation during wakefulness locally increases slow wave activity during sleep. *PLOS*. 3, 1-7.

Ingvalson, E. M., & Wong, P. C. (2013). Training to improve language outcomes in cochlear implant recipients. *Frontiers in Psychology*, *4*, 263.

Karni, A., & Bertini, G. (1997). Learning perceptual skills: behavioral probes into adult cortical plasticity. *Current Opinion Neurobiology, 7*(4), 530-535.

Klimacka, L., Patterson, A., & Patterson, R. (2001). Listening to deaf speech: Does experience count? *International Journal of Language &aCommunication Disorders, 36*(2), 210-215.

McGarr, N. S. (1983). The intelligibility of deaf speech to experienced and inexperienced listeners. *Journal of Speech Language and Hearing Research, 26*(3), 451-458.

Murray, M., M., Brunet, D., & Michel, C., M. (2008). Topographic ERP analyses: A step-by-step tutorial review. *Brain Topography*. 20, 249-264.

Mednick, S., Nakayama, K., & Stickgold, R. (2003). Sleep-dependent learning: a nap is as good as a night. *Nature Neuroscience, 6*(7), 697-698.

Mednick, S. C., Nakayama, K., Cantero, J. L., Atienza, M., Levin, A. A., Pathak, N., & Stickgold, R. (2002). The restorative effect of naps on perceptual deterioration. *Nature Neuroscience*, *5*(7), 677-681.

Molloy, K., Moore, D. R., Sohoglu, E., & Amitay, S. (2012). Less Is More: Latent Learning Is Maximized by Shorter Training Sessions in Auditory Perceptual Learning. *Plos One, 7*(5).

Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics, 60*(3), 355-376.

Obleser, J., Elbert, T., Lahiri, A. & Eulitz, C. (2003) Cortical representation of vowels reflects acoustic dissimilarity determined by formant frequencies. *Cognitive Brain Research*, 15, 207-213.

Ortiz, J. A., & Wright, B. A. (2010). Differential rates of consolidation of conceptual and stimulus learning following training on an auditory skill. *Experimental Brain Research, 201*(3), 441-451.

Parbery-Clark, A., Skoe, E., Lam, C., & Kraus, N. (2009). Musician enhancement for speech-in-noise. *Ear and Hear, 30*(6), 653-661.

Pasley, B. N., David, S. V., Mesgarani, N., Flinker, A., Shamma, S. A., Crone, N. E., et al. (2012). Reconstructing speech from human auditory cortex. *Plos Biology*, 10, e1001251.

Reinke, K. S., He, Y., Wang, C., & Alain, C. (2003). Perceptual learning modulates sensory evoked response during vowel segregation. *Cognitive Brain Research, 17*(3), 781-791.

Rubin, P., & Vatikiotis-Bateson, E. (1998). Measuring and modeling speech production. In S.L. Hopp, M.J. Owren, and C.S. Evans (Eds)., *Animal Acoustic Communication*. 250-290 Springer-Verlag.

Samuel, A. G., & Kraljic, T. (2009). Perceptual learning for speech. *Attention Perception & Psychophysics, 71*(6), 1207-1218.

Schwab, E. C., Nusbaum, H. C., & Pisoni, D. B. (1985). Some effects of training on the perception of synthetic speech. *Human Factors, 27*(4), 395-408.

Sebastian-Galles, N., Dupoux, E., Costa, A., & Mehler, J. (2000). Adaptation to time-compressed speech: phonological determinants. *Perception & Psychophysics, 62*(4), 834-842.

Shahin, A., Bosnyak, D. J., Trainor, L. J., & Roberts, L. E. (2003). Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. *Journal of Neuroscience, 23*(13), 5545-5552.

Stickgold, R., & Walker, M. P. (2007). Sleep-dependent memory consolidation and reconsolidation. *Sleep Medicine, 8*(4), 331-343.

Tervaniemi, M., Just, V., Koelsch, S., Widmann, A., & Schroger, E. (2005). Pitch discrimination accuracy in musicians vs nonmusicians: an event-related potential and behavioral study. *Experimental Brain Research, 161*(1), 1-10.

Tessier, C., Weill-Chounlamountry, A., Michelot, N., & Pradat-Diehl, P. (2007). Rehabilitation of word deafness due to auditory analysis disorder. *Brain Injury, 21*(11), 1165-1174.

Tong, Y. X., Melara, R. D., & Rao, A. (2009). P2 enhancement from auditory discrimination training is associated with improved reaction times. *Brain Research, 1297*, 80-88. doi: DOI 10.1016/j.brainres.2009.07.089

Tononi, G., & Cirelli, C. (2003). Sleep and synaptic homeostasis: a hypothesis. *Brain Research Bulletin, 62*(2), 143-150.

Tononi, G., & Cirelli, C. (2006). Steep function and synaptic homeostasis. *Sleep Medicine Reviews, 10*(1), 49-62.

Tononi, G., & Cirelli, C. (2012). Time to be SHY? Some comments on sleep and synaptic homeostasis. *Neural plasticity*, 2012, 1-12.

Tremblay K. L. (2007).Training-related changes in the brain: Evidence from human auditory evoked potentials. *Seminars in Hearing*. 28, 120-132.

Tremblay, K. L., Shahin, A. J., Picton, T., & Ross, B. (2009). Auditory training alters the physiological detection of stimulus-specific cues in humans. *Clinical Neurophysiology, 120*(1), 128-135.

Tucker, M. A., Tang, S. X., Uzoh, A., Morgan, A., & Stickgold, R. (2011). To Sleep, to Strive, or Both: How Best to Optimize Memory. *Plos One, 6*(7).

Van Engen, Kristin J. (2012). Speech-in-speech recognition: a training study. *Language and Cognitive Processes,* 27(7-8), 1089-1107.

Walker, M., P. (2005). A refined model of sleep and the time course of memory formation. *Behavioral and Brain Sciences,* 28, 51-104.

Zendel, B. R., & Alain, C. (2009). Concurrent sound segregation is enhanced in musicians. *Journal of Cognitive Neuroscience, 21*(8), 1488-1498.

CURRICULUM VITAE

Graduate College
University of Nevada, Las Vegas

Vanessa C. Irsik

Degrees:
Bachelor of Arts, Psychology, 2012
University of Nevada, Las Vegas

Special Honors and Awards:
Graduate & Professional Student Association Grant, 2014
ARO Student Travel Award, 2013

Publication:
Gregg, M., Irsik, V., Snyder, J.S. (2014). Change deafness and object encoding with recognizable and unrecognizable sounds. *Neuropsychologia*. 61, 19-30.

Thesis Title: The Effect of Sleep on Perceptual Learning and Memory Consolidation

Thesis Examination Committee:
Chairperson, Joel S. Snyder, Ph.D.
Committee Member, Erin Hannon, Ph.D.
Committee Member, Jefferson Kinney, Ph.D.
Graduate Faculty Representative, Gabriele Wulf, Ph.D.