Brain Mapping of the Mismatch Negativity Response to Vowel Variances of Natural and Synthetic Phonemes

Lyndsy Marie Smith

Brigham Young University - Provo

Follow this and additional works at: https://scholarsarchive.byu.edu/etd

Part of the Communication Sciences and Disorders Commons

BYU ScholarsArchive Citation

https://scholarsarchive.byu.edu/etd/3890

This Thesis is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in All Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu.
Brain Mapping of the Mismatch Negativity Response to Vowel Variances of Natural and Synthetic Phonemes

Lyndsy M. Smith

A thesis submitted to the faculty of Brigham Young University in partial fulfillment of the requirements for the degree of Master of Science

David L. McPherson, Chair
Shawn L. Nissen
Christopher Dromey

Department of Communication Disorders Brigham Young University
October 2013

Copyright © 2013 Lyndsy M. Smith
All Rights Reserved
ABSTRACT

Brain Mapping of the Mismatch Negativity Response to Vowel Variances of Natural and Synthetic Phonemes

Lyndsy M. Smith
Department of Communication Disorders, BYU
Master of Science

The mismatch negativity (MMN) is a specific event-related potential (ERP) component used frequently in the observation of auditory processing. The MMN is elicited by a deviant stimulus randomly presented in the presence of repeating stimuli. The current study utilized the MMN response in order to determine the temporal (timing) and linguistic processing of natural and synthetic vowel stimuli. It was hypothesized that a significant MMN response would be elicited by natural and synthetic vowel stimuli. Brain mapping of the MMN response was hypothesized to yield temporal resolution information, which would provide detail regarding the sequential processing differences between natural and synthetic vowel stimuli. It was also hypothesized that the location of dipoles within the cortex would provide information pertaining to differences in cortical localization of processing for natural and synthetic stimuli. Vowel stimuli were presented to twenty participants (10 females and 10 males between the ages of 18 and 26 years) in a three-forced-choice response paradigm. Data from behavioral responses, reaction times, and ERPs were recorded for each participant. Results demonstrated that there were differences in the behavioral and electrophysiological responses between natural and synthesized vowels presented to young, normal hearing adults. In addition, significant MMN responses were evoked by both natural and synthetic vowel stimuli. Greater reaction times existed for the synthetic vowel phonemes compared to the natural vowel phonemes. Electrophysiological differences were primarily seen in the processing of the synthetic /u/ stimuli. Scalp distribution of cognitive processing was essentially the same for naturally produced phonemes. Processing of synthetic phonemes also had similar scalp distributions; however, the synthetic /u/ phoneme required more complex processing compared to the synthetic /æ/ phoneme. The most significant processing localizations were located in the superior temporal gyrus, which is known for its role in linguistic processing. Continued processing in the frontal lobe was observed, suggesting continual evaluation of natural and synthetic phonemes throughout processing.

Keywords: brain mapping, dipole localization, electroencephalography, event-related potentials, mismatch negativity, natural speech stimuli, linguistic processing, speech perception, synthetic speech stimuli
ACKNOWLEDGEMENTS

I would like to express my gratitude to my thesis chair, Dr. McPherson for his guidance, advice, and constant encouragement during the completion of this project. I truly appreciate his mentoring and patience throughout this project and for helping me meet my goal. I would also like to thank my committee members, Dr. Dromey and Dr. Nissen for their advisement and suggestions. In addition, this project would not have been possible without the family members and friends who devoted their time and energy as participants and supporters of my project. Lastly, I would like to thank my loving husband Andrew who gave me unwavering support and encouragement throughout my graduate work and life. I would not have been able to do this without him and his consistent push towards helping me reach all of my dreams and goals!
# TABLE OF CONTENTS

Introduction ........................................................................................................................................ 1

Measuring Cerebral Processing ...................................................................................................... 1

Processing by the Auditory Cortex ................................................................................................. 4

Method .......................................................................................................................................... 13

Participants ..................................................................................................................................... 13

Instrumentation ............................................................................................................................. 14

Stimuli .......................................................................................................................................... 15

Procedure ..................................................................................................................................... 16

Data Analysis ................................................................................................................................. 17

Results .......................................................................................................................................... 18

Behavioral Data .............................................................................................................................. 18

Event Related Potential Data ......................................................................................................... 19

Repeated Measures ....................................................................................................................... 34

Discussion .................................................................................................................................... 37

Summary and Evaluation of Results .............................................................................................. 38

Recommendations for Future Research ......................................................................................... 43

References ..................................................................................................................................... 47
LIST OF TABLES

1. Descriptive statistics for reaction time, in ms, across participants for each stimulus condition ................................................................................................................................................................................................. 19

2. Descriptive statistics for latency, in ms, and amplitude, in µV, of the ERP components of the natural /æ/ stimuli across participants.......................................................................................................................... 21

3. Descriptive statistics for latency, in ms, and amplitude, in µV, of the ERP components of the synthetic /æ/ stimuli across participants. ................................................................. 22

4. Descriptive statistics for latency, in ms, and amplitude, in µV, of the ERP components of the natural /u/ stimuli across participants. ................................................................. 23

5. Descriptive statistics for latency, in ms, and amplitude, in µV, of the ERP components of the synthetic /u/ stimuli across participants. ................................................................. 23

6. Average latency, in ms, and amplitude, in µV, of the MMN for each vowel condition ... 24
LIST OF FIGURES

1. Number of correct vs. incorrect responses as a function of stimulus type. ................................. 20
2. Means plot of reaction time for stimulus type. .................................................................................. 21
3. Grand-averaged brain maps of the ERPs for the natural /æ/ stimulus........................................... 26
4. Grand-averaged brain maps of the ERPs for the synthetic /æ/ stimulus........................................... 27
5. Grand-averaged brain maps of the ERPs for the natural /u/ stimulus.............................................. 28
6. Grand-averaged brain maps of the ERPs for the synthetic /u/ stimulus........................................... 30
7. Grand-averaged z-score brain maps of the ERPs for the natural /æ/ and natural /u/ stimuli... 31
8. Grand-averaged t-score brain maps of the ERPs for the natural /æ/ and natural /u/ stimuli ... 32
9. Grand-averaged z-score brain maps of the ERPs for the synthetic /æ/ and synthetic /u/ stimuli ........................................................................................................................................... 33
10. Grand-averaged t-score brain maps of the ERPs for the synthetic /æ/ and synthetic /u/ stimuli ........................................................................................................................................... 34
11. Grand-averaged MMN brain maps of the ERPs for the natural /æ/ and synthetic /æ/ stimuli ........................................................................................................................................... 35
12. Grand-averaged MMN brain maps of the ERPs for the natural /u/ and synthetic /u/ stimuli ........................................................................................................................................... 36
LIST OF APPENDICES

Annotated Bibliography........................................................................................................... 51

Informed Consent to Act as a Human Research Subject ......................................................... 78
Introduction

Brain activity is responsible for the majority of all functions in the human body, including auditory processing of speech. Speech processing occurs when the brain perceives and identifies acoustic stimuli and then uses phonetic categories to interpret a spoken message (Benson et al., 2001). For example, children are born with an inherent capability to perceive phonetic categories (Näätänen, 2001). A child’s ability to distinguish phonetic categories matures and develops as the child grows (Benson et al., 2001; Näätänen, 2001). The inherent capability to discern phonetic categories allows individuals to process and understand the meaning of speech stimuli.

Current research suggests that nonspeech and speech stimuli may be processed differently by the brain; however, currently the exact differences between the processing of these two stimuli remains unclear (Benson et al., 2001; Blomert & Mitterer, 2004; Perry, 2012; Tampas, Harkrider, & Hedrick, 2005). Recent studies have focused on finding an accurate way to investigate auditory processing to determine the difference between nonspeech and speech processing. Developments in the use of electrophysiology in research, specifically through the use of auditory event-related potentials (ERPs), have shown that ERPs are an effective way to study cerebral auditory processing of both nonspeech and speech stimuli (Cook, O'Hara, Uijtdehaage, Mandelkern, & Leuchter, 1998; Kraus, McGee, Sharma, Carrell, & Nicol, 1992; Näätänen, 1995; Näätänen, 2001).

Measuring Cerebral Processing

Electrophysiology. Electrophysiology plays an integral role in the study of auditory processing and language (Bomba, Choly, & Pang, 2011; Csépe, Osman-Sági, Molnár, & Gósy, 2001; Horev, Most, & Pratt, 2007; Kasai et al., 2002; Oates, Kurtzberg, & Stapells, 2002; Perry,
2012; Tampas et al., 2005). Electrophysiology allows researchers to measure brain activity and localize neuronal populations of specific cortical regions within the central auditory nervous system (Kasai et al., 2002; Tampas et al., 2005). Through the identification of active neuronal populations and ERPs, researchers are able to map the activity of cerebral processing. ERPs are obtained through the use of electroencephalography (EEG). EEG measures brain electrical activity by placing electrodes at specific locations across the scalp. These electrodes record ionic current flow from large populations of neurons in the brain that are activated in response to extrinsic and intrinsic stimulation (Näätänen, 1995).

Auditory ERPs are one of several types of electrophysiological responses used in examining cerebral processing. ERPs are changes in the ongoing EEG that are elicited by cognitive, motor, or sensory stimuli (Horev et al., 2007). The ERP reflects activation of the auditory nerve, brain stem centers, auditory thalamo-cortical pathways, and higher linguistic and cognitive brain function. By recording the responses to auditory stimulation at a series of electrode sites across the scalp it is possible to observe activation of the auditory neural system for a variety of auditory tasks (Kraus et al., 1992; Tampas et al., 2005). In order for brain electrical potentials to occur, groups of neurons must be synchronized to the stimulus. The synchronization of neurons results in a sensory potential which can be recorded at the scalp (i.e., ERP). Consequently, the ERP represents the synchronous activity of many neurons in response to a sensory stimulus. Different types of stimulation will elicit different components of ERP. The classifications of ERPs allow researchers to investigate specific aspects of brain activity in response to different types of sensory stimulation (Picton, 2006).

Mismatch negativity. The Mismatch Negativity (MMN) is a specific ERP component used frequently in the observation of auditory processing. The MMN is obtained by presenting a
subject with a standard, or common, auditory stimulus. During the presentation of the standard stimuli, a deviant, *odd-ball* stimulus is randomly presented in place of the standard. This unexpected stimulus elicits a difference waveform, known as the MMN. The MMN is derived by subtracting the response elicited by the standard stimuli from the response elicited by the deviant stimuli. The MMN reflects a discrimination process that occurs in the auditory cortex (Näätänen, 1995; Picton, 2006).

The MMN response allows researchers to observe the sensory and short-term auditory memory components involved in auditory processing. Sensory memory has a role in recognizing deviant breaks in established patterns of auditory stimuli. As an auditory stimulus is perceived, the cortex momentarily stores the auditory pattern; therefore, when a deviant stimulus breaks the repetitive pattern, the cortex recognizes the difference and a MMN response occurs (Picton, 2006). The short-term memory that is created to the auditory patterns provides the means for auditory processing to occur each time a stimulus is presented (Näätänen, 1995).

The MMN is extensively used in research because it provides robust responses to changes in sensory stimulation. Stimuli differences effectively elicit MMN responses and provide an objective measure of discrimination independent of a participant’s ability to respond to stimuli. Also, the MMN is elicited regardless of a participant’s attention to the auditory task (Näätänen, 1995, 2001, 2008). Experiments have shown that natural and synthetic-derived speech as well as nonspeech stimuli will evoke a measurable MMN response (Kraus et al., 1992; Picton, 2011; Tampas et al., 2005). For example, Csépe et al. (2001) used the MMN to determine whether the MMN response to auditory stimuli was deviant in individuals with aphasia. Both natural and synthetic vowels were used to elicit the MMN. The results of the study concluded that the MMN can be elicited by natural and synthetic contrasts. In addition, the
study found that neurological impairment does result in a deviant MMN response when elicited by phonetic features. Kraus et al. (1992) also studied the MMN response elicited by natural and synthetic speech contrasts to assess central auditory functioning in typical adults and children. The study concluded that the MMN can be elicited by natural and synthetic speech stimuli. In addition, Kraus et al. (1992) concluded that the MMN is an accurate measure for assessing central auditory functioning in both adults and children.

Processing by the Auditory Cortex

Models of speech processing. Several models and theories have been proposed to describe the process of speech perception. One theory under consideration is the speech-parallel model. This model is based on the idea that acoustic and phonetic information is processed by the brain simultaneously in a parallel fashion. Work done by Tampas et al. (2005) found evidence which favored a speech-parallel model of processing instead of the more well-known, hierarchical model of speech. The theory of the hierarchical model of speech is that acoustic stimuli are processed and encoded into hierarchy of pattern-recognition schemes. Auditory stimuli that require simple encoding are processed by lower level neurons and more complex encoding schemas are processed by higher level neurons.

Tampas et al.’s (2005) study used synthetically generated consonant-vowel stimulus contrasts on a continuum (same phoneme with acoustical differences) in conjunction with nonspeech frequency glide contrasts. The stimuli were created to be as acoustically similar as possible. Participants were instructed to discriminate between stimuli using a button-push in a same/different discrimination task. The MMN measurements were collected for each individual. Behavioral results and the MMN measurements indicated that individuals exhibited better discrimination abilities with the nonspeech contrasts compared to the consonant-vowel
continuum. These results suggested that phonetic and acoustic characteristics are processed in a parallel fashion and occur as early as the level of the MMN generation in the central auditory system (about 200 ms). MMN generation is thought to originate at the N1-P2 dipoles near the temporal lobe and also in areas of the frontal lobe (Picton, 2011). Although parallel processing was observed, the authors did not reject the idea that hierarchical processing may also occur; however, the authors concluded that if hierarchical processing took place, it was only at the lower levels of the central auditory system (Tampas et al., 2005).

The hierarchical organization of processing model is a more commonly accepted speech processing model (Benson et al., 2001; Kraus et al., 1992; Näätänen, 2001). DeWitt and Rauschecker (2012) completed an analysis of 115 studies to demonstrate that the human cortex performs processing in a hierarchical fashion. In the hierarchical model, acoustic stimuli are received by the primary auditory cortex (Benson et al., 2001; DeWitt & Rauschecker, 2012; Kraus et al., 1992; Näätänen, 2001). The acoustic stimuli are then encoded into hierarchical pattern-recognition schemes and more complex stimuli are processed by higher level neurons (DeWitt & Rauschecker, 2012; Näätänen, 2001). Through a careful review of several studies, DeWitt & Rauschecker (2012) concluded that as hierarchical processing occurs, there is a subsequent increase in neuron receptive field size for more complex auditory tasks. In addition, higher levels of processing are required in word-form recognition tasks because temporal arrangement characteristics of phonemes must be processed. In contrast, only lower levels of processing were required in phoneme recognition tasks because processing of spectrotemporal features of phonemes is less complex (DeWitt & Rauschecker, 2012; Näätänen, 2001).

Similarly to the DeWitt & Rauschecker (2012) study, Näätänen (2001) also concluded that cerebral processing of speech sounds occurred in a hierarchically organized fashion.
Näätänen (2001) concluded that processing starts at the lowest level where phoneme traces are identified by the auditory system. These low level phoneme traces serve as recognition patterns which aid in the categorization of sounds. The phoneme traces are then combined to form syllables which are subsequently combined at a higher level to form words. Through this hierarchical processing, speech is recognized and interpreted by the auditory cortex (Näätänen, 2001).

**Categorical perception.** While the speech-parallel and hierarchical organization models explain speech processing differently, the models do not underestimate the importance of categorical perception and its role in speech perception. Categorical perception is the ability of a listener to perceptually differentiate between speech sounds. Speech sounds occur along a continuum of acoustic characteristics. The acoustic characteristics are either too discrete for the neural system to discern a difference, or a difference is noted by the neural system and the acoustic event is separated into a category. These acoustic differences mark phonemic boundaries and allow the neural system to discriminate speech sounds in order to place meaning to incoming stimuli (Tampas et al., 2005).

A study by Bomba et al. (2011) demonstrated how phoneme discrimination is dependent on linguistic knowledge. The experiment included eight native English speakers and eight native Japanese speakers, who also knew English. Each individual was presented with three phonemic contrasts while MMN measurements were collected from 26 electrodes placed across the scalp. The phonemic contrasts included vowels native to both languages, consonant-vowel syllables phonemic in both languages, and consonant-vowel syllables phonemic only in English. The results indicated that there was no difference in the MMN for the vowels native to both languages. However, the MMN was elicited in only 50 percent of the Japanese participants in
the English only consonant-vowel syllable task. These results suggested that Japanese speakers did not possess the distinct patterns of categorical discrimination that the native English speakers possessed. Through this study, the effect of language on categorical perception was demonstrated based on the Japanese participants’ inability to identify perceptual changes when presented with syllables that were nonphonemic in Japanese.

Näätänen (2001) reported similar findings to those of Bomba et al. (2011) regarding the effect of language on categorical perception. After a review of several studies, Näätänen proposed that perception of phonemes is based on language-specific phonetic traces, which arise in the auditory cortex. These traces, or neural pathways, are developed during the first few months of life and are used in the lower hierarchical levels to processes speech. Through categorical perception, listeners subconsciously attend to acoustic differences that alter word meaning and ignore differences that do not affect word meaning (Horev et al., 2007). Thus, through categorical perception, listeners place meaning to sounds.

Recent studies have found that categorization does not only occur with speech sounds, but the cortex also perceives some nonspeech sounds categorically (Horev et al., 2007; Kraus et al., 1992). For example, Horev et al. (2007) conducted a study in which behavioral measures and auditory ERPs were obtained from individuals as they identified and discriminated between two sets of speech stimuli. The stimuli included an analogous nonspeech continuum composed of synthesized formants and a voice-onset-time continuum of natural speech productions. Results from the ERP measures indicated that nonspeech and speech stimuli were both perceived in a categorical manner; however, it was also found that nonspeech and speech stimuli were processed differently, especially in the early stages of processing (Horev et al., 2007).
Localization of processing. Most researchers agree that the general location of speech processing is predominantly in the left hemisphere (Benson et al., 2001; DeWitt & Rauschecker, 2012; Näätänen et al., 1997; Neville et al., 1998; Perry, 2011; Wood, Goff, & Day, 1971). Näätänen et al. (1997) conducted an experiment to identify where in the cortex language processing occurs. Estonian and Finnish participants were instructed to press a button corresponding to a specific stimulus. The participants were presented with one standard phoneme and three deviant phonemes, one of which was only a prototype in Estonian. MMN measures were recorded for each individual. The results of the study indicated that the MMN in Finnish participants was enhanced when the individual was presented with a Finish prototype compared to when an Estonian prototype was presented. In conclusion, the study made two significant observations. First, language processing occurred predominantly in the left hemisphere. Secondly, cortical memory traces, which play a role in categorical perception, were also found in the left hemisphere.

Processing localization of nonspeech stimuli has also been studied. Benson et al. (2001) conducted an investigation with the use of functional magnetic resonance imaging (fMRI) to identify where processing took place in the cortex. Participants were presented with alternating trials of nonspeech and speech stimuli. Nonspeech stimuli were tonal in nature and increased in complexity from notes, to chords, to chord progressions. The speech stimuli consisted of nonsense syllables varying in phonetic complexity from steady-state vowels, to consonant-vowel syllables, to consonant-vowel-consonant syllables. During the presentation of these stimuli, fMRI was used to observe brain localization of speech processing. The fMRI showed that the processing of both nonspeech and speech stimuli occurred predominantly in the left hemisphere with different neural events occurring in specific areas of the left hemisphere. The presence of
these different neural events depended on whether nonspeech or speech processing occurred. The authors also observed right hemisphere activation; however, the study concluded that this observation was insignificant and that the majority of nonspeech and speech processing occurred in the left hemisphere.

Wood et al. (1971) conducted a study to attempt to determine where and why nonspeech and speech processing differences occur in the brain. Participants were fitted with silver disk electrodes placed across the left and right hemispheres of the scalp. Neural responses were measured as the participants completed two auditory discrimination tasks. One was a linguistic task, which used two stop consonants (/ba/ and /da/). The consonants were created to be identical in duration, initial fundamental frequency, frequency contour, and intensity contour. The other task was a fundamental frequency task, which used a stop consonant (/ba/) with different fundamental frequencies (/ba/-low and /ba/-high). Similar to the Benson et al. (2001) study, results indicated that neural events occurring in the right hemisphere were identical in both tasks; however, neural events in the left hemisphere differed between the two tasks. These differences were seen in the ERPs measured during the tasks. Wood et al. (1971) concluded that nonlinguistic and linguistic stimuli are processed differently and that different neural events occur in the left hemisphere depending on the stimuli being processed. Furthermore, linguistic processing requires a specialized unilateral neural mechanism, which is not necessary in the processing of nonlinguistic stimuli.

**Vowel processing.** Previous research in vowel processing has attempted to locate the neural areas responsible for the processing of specific acoustic events and properties. Much of this research utilized vowels as phoneme stimuli to target specific areas of processing. Vowels are differentiated through unique vowel formants. Formants are bands of energy which peakat
specific frequencies and convey linguistic information (Poeppel et al., 1997). Formants are categorically perceived by the auditory cortex, which is tonotopically organized into neuronal populations that respond to specific spectral bands of energy (Makela, Alku, & Tiitinen, 2003). The spectral bands of energy most important in the categorical perception of vowels are the formants F1 and F2, which are the two lowest formant frequencies of vowels. The acoustic distances between F1 and F2 provide information to the auditory cortex, which allows it to decode and categorize vowels (Sadeghian, Dajani, & Chan, 2011; Vihla, Lounasmaa, & Salmelin, 2000).

Vowels are a common type of stimuli used in research on language processing because they are easily alterable. By manipulating different characteristics of vowels it is possible to observe how the brain responds and what areas respond to different acoustic characteristics. In several studies vowels were used to compare the processing of natural versus synthetic forms of speech. One particular experiment performed by Eulitz, Diesch, Pantev, Hampson, and Elbert (1995) examined the effects of processing localization in synthetic vowels using magnetoencephalography (MEG) and EEG. The study used five synthetic German vowels (/a/, /i/, /u/, /æ/, and /oe/) of two different durations (45 msec and 600 msec). In addition to vowels, a 1000 Hz tone was also used as a stimulus. The long and short synthetic vowels and the 1000 Hz tone were randomly presented to eleven normal hearing individuals. Response data were collected for each participant using a 37-channel neuromagnetometer. A neuromagnetometer is used in MEG data collection and detects subtle magnetic fields resulting from cerebral electrical processing activities. The electrodes were used to measure the response of the N100 ERP wave, which is a negative electrical wave that occurs in response to the onset of a stimulus (Picton, 2006). The data were used to analyze the localization of dipoles in the processing of the five
synthetic vowels. Eulitz et al. (2007) concluded from the MEG and EEG data that the amplitudes of the MEG and the EEG signals were different depending on whether the signals were elicited by synthetic vowels or a 1000 Hz tone. The MEG and the EEG signal amplitudes were larger when elicited by synthetic vowels compared to the amplitudes elicited by a 1000 Hz tone. In addition, the MEG measurements indicated that the left hemisphere had greater dipole strength for the vowel condition as opposed to the tone condition. However, the study failed to show different dipole locations for the two conditions.

Another vowel experiment was completed by Makela et al. (2003), who observed cortical source locations during the presentation of sustained vowels. Participants in the study were presented with three semi-synthetic Finnish sustained vowels (/a/, /o/, and /u/). The vowels were presented in pseudorandom order in separate sequences to ten participants. Each participant was fitted with a 306-channel MEG in order to observe the response of the N1m (“m” designates the component was derived using MEG) wave. The N1m wave is a magnetic wave elicited by different vowels and corresponds to the N100. Data from the study showed that the left hemisphere is highly specialized to process speech. In addition, different vowels were found to activate distinctly separate areas in the auditory cortex of the left hemisphere. Vowels that had similar formants were processed closer to each other in the auditory cortex compared to vowels that had greater acoustic differences between F1 and F2. These findings led Makela et al. (2003) to conclude that categorical perception of vowels can be observed through mapping the location of F1 and F2 in the auditory cortex.

Eulitz et al. (2007), Makela et al. (2003), and other similar studies (Vihla et al., 2000; Yrttiaho, Tiitinen, Alku, Miettinen, & May, 2010) used MEG to identify cortical areas that are active during vowel processing. While the above studies provide important information, MEG
only observes spatial resolution of processing; therefore, questions remain regarding temporal processing (temporal resolution) of vowel information. Temporal resolution information, or information about the precise timing of processing, provides information on the sequential processing of vowel stimuli.

A recent study by Perry (2012) used natural speech sounds to evoke the MMN response and examined the processing required to identify vowels with altered formant bands. The study revealed that the MMN can be evoked by speech sounds and that preattentive formant-based processing occurs in the identification of vowels. Perry suggested that further research is needed in order to understand differences in the timing and cortical processing of natural speech and synthetic speech. The current study addressed these issues raised by Perry (2012).

The present study had two main purposes. The first was to evaluate differences in behavioral and electrophysiological responses between natural and synthesized vowels presented to young, normal hearing adults. The second purpose was to evaluate whether there were distinct differences in spatial-temporal processing in the cerebral cortex, using the above two types of stimuli. Behavioral measures of reaction time (RT) were included in the present study. Electrophysiological measures included amplitude and latency of the MMN, scalp distribution of the MMN, and estimates of source localization. It was hypothesized that the ERP could be used to measure a significant MMN response, which was elicited by all deviant stimuli, consisting of natural and synthesized vowel stimuli. Localization of dipoles within the cortex was hypothesized to yield temporal resolution information, which would explain the distinct sequential processing differences that occurred in natural and synthesized vowel processing.
Method

Participants

Twenty individuals (10 males and 10 females) between the ages of 18 and 26 years ($M = 23$) participated in this study. Both males and females were used in the current study because research has shown there is no effect of gender on the amplitude, latency, or topography of the MMN in normal adults (Kasai et al., 2002). All participants were required to be native English speakers (Bomba et al., 2011; Neville et al., 1998) and have no reported history of neurological, cognitive, or learning impairments (Csépe et al., 2001). The participants were also required to be right handed (Knecht et al., 2000) and to pass an initial hearing screening demonstrating that hearing was within normal limits, bilaterally. Hearing screening included otoscopy, tympanometry, pure tone testing, speech recognition thresholds, and word recognition scores. Hearing screenings met the specifications set forth by the American Speech-Language Hearing Association (ASHA, 1990). This included clear, healthy tympanic membranes bilaterally, bilateral type A tympanograms with static acoustic admittance measures between 0.3 and 1.4 mmhos, and peak pressure between -100 and +50 daPa. Normal pure tone thresholds were defined as $\leq 15$ dB HL for octave intervals between 250-8000 Hz and threshold differences between ears $\leq 5$ dB HL. Speech recognition thresholds did not exceed the limits of $\leq 15$ dB HL and were within 2 dB of the pure tone average. Word recognition scores were $\geq 98\%$ bilaterally (ASHA, 1990).

Each participant read and signed an informed consent document approved by the Institutional Review Board at Brigham Young University (see Appendix B) before participating in the study. In addition to meeting the ethical requirements set by Brigham Young University, this study also met the ethical requirements as stated in the Declaration of Helsinki (World
Medical Association, 2008). The participants received compensation ($10.00 USD) for their participation in the study. Three participants from the study were chosen at random to be tested a second time in order to compare data and verify test-retest reliability.

**Instrumentation**

**Stimulus preparation.** Stimuli were recorded in a sound-isolated chamber using a low impedance dynamic microphone (DPA 4011). During recording, the microphone was placed approximately 15 cm from the speaker’s mouth. The recordings were made at 44.1 kHz and digitized with an A/D converter by Apogee Systems. Adobe Audition 3.0 was used to edit the recordings and prepare them to interface with the NeuroScan 4.5 software. The recordings were down sampled to 22.1 kHz, 16 bits in order to be used on the NeuroStim system.

**Instrumentation for initial hearing screening.** Instrumentation used for the hearing screening included a Welch Allyn otoscope for otoscopy, a handheld Grason-Stadler model GSI-33 impedance meter for tympanometry, and a Grason-Stadler model GSI-1761 audiometer with Etymotic EA-3 insert phones for the auditory testing. Also, during data acquisition the test stimuli were presented to the participant via a Grason-Stadler audiometer.

Actual hearing tests were conducted in a double-walled, sound-treated test booth. Noise levels were within the limits as specified by ANSI S3.1-1999 R2008 Maximum Permissible Ambient Noise Levels for Audiometric Test Rooms (ASHA, 2008).

**EEG data collection.** Participants were fitted with an electrode cap (ElectroCap International, Inc., 2003) having 32 silver-silver chloride electrodes resting against the scalp and distributed according to the 10-20 International System (Jurcak, Tsuzuki, & Dan, 2007). In addition to the scalp electrodes, six electrodes were placed on the right and left mastoid process (linked-mastoid references), the outer cantha of the right and left eyes, and one above and below
the supraorbital foramen of the left eye. These additional six electrodes were placed to monitor activity and movement of the eye and facial muscles. Electrode impedances of the cap did not exceed 3000 ohms.

Specific software programs were used to record electroencephalographic data as well as to present the stimuli. Compumedics software (2008) was used for EEG data collection and initial analysis (NeuroScan 4.5). NeuroScan Stim 2 software was used for stimulus presentation. In addition, CURRY 6 (Compumedics Neuroscan, 2008) software was used for cortical localization of the electrophysiological responses, post hoc.

Stimuli

Two corner vowels (/æ/ and /u/) were chosen as the stimuli for the study based on the observation that they are more acoustically identifiable than other vowels (Peterson & Barney, 1952). The first set of stimuli consisted of two natural vowels (/æ/ and /u/). A female native English speaker produced the stimuli. The speaker sustained the /æ/ vowel for 378 ms and the /u/ vowel for 447 ms. One male and two female students, unfamiliar with the study, correctly identified the phonemes and judged the recordings to be of equal loudness level. The evaluators correctly identified the recordings as /æ/ and /u/ and estimated the loudness level of the recordings to be equal according to the method described by Perry (2012).

The final /æ/ recording was 278 ms and the /u/ recording was 347 ms. To maximize vowel identification, the middle portion of the vowel was preserved. These final recordings were presented to participants (Perry, 2012). Durational differences between the vowels were not considered in this study; however, duration within the phoneme pairs was equal.

A second set of stimuli using the same vowels was constructed for this study by the Articulatory Synthesis project at Haskins Laboratories, New Haven, Connecticut. All final
recordings were high-pass filtered to eliminate noise below 65 Hz. Each stimulus intensity was ramped with a filter to eliminate audible clicks in the first and last 10 ms of the recordings to remove artifacts. The synthesized phonemes were the same duration as their respective natural phonemes, as noted above.

**Procedure**

**Initial screening.** Participants were presented with the two natural vowel recordings (/æ/ and /u/) and were asked to identify the vowels in order to determine whether they could recognize the vowels. Participants unable to make the discrimination of the natural vowels were excluded from the study.

**Behavioral data acquisition.** Participants listened to four different stimulus blocks presented in an oddball paradigm. All stimuli were presented bilaterally at 70 dB HL for a series 884 trails. Participants responded manually to all stimuli with a button push. Prior to the presentation of the stimuli, participants were read the following instructions:

You will hear many speech sounds. Each sound will be presented as one sound followed by another. You must press a button after each sound. Press button one if the sound was /æ/. Press button two if the sound was /u/. Press button three if the sound was neither /æ/ nor /u/. If you wish to discontinue the test at any time, you may say, “I want to stop now.” Are there any questions? We will start the test.

Following the instructions, participants performed a training task to familiarize them with the procedure. Forty stimuli were presented to the participant in the training task. Testing began following the training activity.
Each of the two blocks contained 221 trials: 80% natural stimuli (standard) and 20% oddball stimuli (deviant) interspersed randomly throughout the block. The standard stimuli consisted of the /æ/ and /u/ natural vowels. The deviant stimuli consisted of the /æ/ and /u/ synthesized vowels. In order to ensure randomization, one deviant stimulus could not directly follow the presentation of another deviant stimulus. The blocks were presented to the participants in a randomized order.

**Event-related potential acquisition.** Participants sat quietly in an audiometric test room during the acquisition of the data. In order to maintain alertness, the participants were permitted to stretch (without removing or repositioning the electrode cap) after the presentation of the first block of stimuli. During stimulus presentation, the ongoing EEG was streamed onto the computer using Neuroscan 4.5 software and saved for post hoc averaging and processing of the recordings. Participants’ responses, response times, and EEG were recorded and stored on a secure computer.

**Data Analysis**

**Behavioral data.** The behavioral data were analyzed using a signal detection paradigm where hits, misses, correct rejections, and false alarms were obtained. Reaction times were also recorded and descriptive statistics were used to show the results. Only responses with RT less than 2000 ms were collected. The study had two independent variables and two dependent variables. The independent variables were vowel and vowel type (natural or synthesized). The dependent variable was RT. An ANOVA for RT was performed to determine whether the independent variables had an influence on RT. In addition, post-hoc t-tests were completed on the comparison stimuli pairs to determine whether differences from the ANOVA were significant.
**Event-related potential data.** Recordings and epochs were individually examined. Epochs were created from the raw EEG data. Prior to averaging the epochs, the Neuroscan 4.5 software was used to remove artifacts such as eye and jaw movement (Compumedics Neuroscan, 2008). Averages of the ERP data were calculated for each block of stimuli for each participant. Further averaging of individual ERP files for each stimulus was completed for a total of four grand averages of the ERPs. MMN response analysis was computed by subtracting the deviant ERP average from the common ERP average for each stimulus block (Picton, 2006).

Dipoles, cortical source sites of electrical activity, were identified using CURRY 6.0.12 software for all individual averaged ERP files and for the grand averaged ERP files (Compumedics Neuroscan, 2008; Näätänen, 2008). Locations of each dipole were compared between individuals and the grand average for all deviant and standard responses within each block of stimuli. The dipole from the grand averaged ERP file was used to determine the electrode site at which to compute EEG results. Peak latency and amplitude of the MMN were measured. Mean and standard deviations of the MMN measures were determined for all participants.

**Results**

**Behavioral Data**

Descriptive statistics were computed for RT of the four different stimulus conditions (Table 1). One particular statistic of interest was the longer RT seen for the synthetic /u/ stimulus compared to the other stimuli. An ANOVA for RT by stimulus condition showed significant $F$ values ($F(3, 19963) = 82.581$, $p < .001$). Post-hoc $t$-tests were completed on the comparison stimulus pairs. $T$-tests were used to make comparisons between conditions as follows: the comparison of the natural /æ/ versus synthetic /æ/ stimuli ($t(9973) = -6.736, p <$
.001), the comparison of natural /u/ versus synthetic /u/ stimuli ($t(9990) = -10.772, \ p < .001$), the comparison of natural /æ/ versus natural /u/ stimuli ($t(9984) = -2.947, \ p = 0.003$), and the comparison of synthetic /æ/ versus synthetic /u/ stimuli ($t(9979) = -7.401, \ p < .001$). All pairwise comparison $t$-tests showed significant differences.

Table 1

*Descriptive statistics for reaction time, in ms, across participants for each stimulus condition*

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>$N$</th>
<th>$M$</th>
<th>$SD$</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Natural /æ/</td>
<td>4991</td>
<td>673.63</td>
<td>892.36</td>
<td>156</td>
<td>31546</td>
</tr>
<tr>
<td>Natural /u/</td>
<td>4995</td>
<td>720.48</td>
<td>682.64</td>
<td>106</td>
<td>15016</td>
</tr>
<tr>
<td>Synthetic /æ/</td>
<td>4984</td>
<td>797.04</td>
<td>936.95</td>
<td>124</td>
<td>23660</td>
</tr>
<tr>
<td>Synthetic /u/</td>
<td>4997</td>
<td>1001.88</td>
<td>1715.46</td>
<td>106</td>
<td>48114</td>
</tr>
</tbody>
</table>

The ambiguity observed for the synthetic /u/ stimulus, which is demonstrated by an increase in the number of errors seen for this condition, is illustrated in Figure 1. This figure demonstrates that there were greater errors for the synthetic /u/ stimulus than for any other stimuli. Likewise, it is evident that the RT for the synthetic /u/ stimulus was greater than the other three stimulus conditions when both correct and incorrect responses were combined (Figure 2 inset). However, synthetic /æ/ showed the longest RT for the incorrect response (Figure 2).

**Event-related Potential Data**

*Descriptive statistics.* Descriptive statistics were computed for the three components of the ERP (P1, P2, and N1) for each condition (natural and synthetic) and phoneme (/æ/ and /u/). Tables 2 and 3 include the descriptive statistics for the three components of the ERP for the natural /æ/ stimuli (Table 2) and for the synthetic /æ/ stimuli (Table 3). The amplitude and
latencies for both the natural /æ/ phoneme and the synthetic /æ/ phoneme were within one standard deviation of each other. Likewise, the minima and the maxima were similar for the natural /æ/ and synthetic /æ/ conditions. An ANOVA failed to show significant differences between the amplitude and latencies of the ERP components between the natural and synthetic /æ/ conditions.

Figure 1. Number of correct vs. incorrect responses as a function of stimulus type.
Figure 2. Means plot of reaction time for stimulus type. The inset shows the RTs when the correct and incorrect responses are collapsed.

Table 2

Descriptive statistics for latency, in ms, and amplitude, in µV, of the ERP components of the natural /æ/ stimuli across participants.

<table>
<thead>
<tr>
<th>ERP component</th>
<th>M</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1 latency</td>
<td>91.68</td>
<td>18.70</td>
<td>74.40</td>
<td>147.20</td>
</tr>
<tr>
<td>P1 amplitude</td>
<td>1.18</td>
<td>0.85</td>
<td>-0.66</td>
<td>2.87</td>
</tr>
<tr>
<td>N1 latency</td>
<td>136.82</td>
<td>20.50</td>
<td>93.20</td>
<td>174.40</td>
</tr>
<tr>
<td>N1 amplitude</td>
<td>-2.67</td>
<td>1.80</td>
<td>-6.02</td>
<td>0.87</td>
</tr>
<tr>
<td>P2 latency</td>
<td>181.95</td>
<td>31.97</td>
<td>120.20</td>
<td>228.40</td>
</tr>
<tr>
<td>P2 amplitude</td>
<td>1.92</td>
<td>3.01</td>
<td>-2.03</td>
<td>11.39</td>
</tr>
</tbody>
</table>
Table 3

*Descriptive statistics for latency, in ms, and amplitude, in µV, of the ERP components of the synthetic /æ/ stimuli across participants.*

<table>
<thead>
<tr>
<th>ERP component</th>
<th>$M$</th>
<th>$SD$</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1 latency</td>
<td>87.63</td>
<td>13.71</td>
<td>70.60</td>
<td>125.60</td>
</tr>
<tr>
<td>P1 amplitude</td>
<td>0.72</td>
<td>0.81</td>
<td>-0.91</td>
<td>2.41</td>
</tr>
<tr>
<td>N1 latency</td>
<td>138.21</td>
<td>20.36</td>
<td>91.80</td>
<td>160.80</td>
</tr>
<tr>
<td>N1 amplitude</td>
<td>-3.08</td>
<td>1.54</td>
<td>-6.47</td>
<td>-0.33</td>
</tr>
<tr>
<td>P2 latency</td>
<td>181.92</td>
<td>32.12</td>
<td>120.20</td>
<td>217.60</td>
</tr>
<tr>
<td>P2 amplitude</td>
<td>1.26</td>
<td>1.95</td>
<td>-1.18</td>
<td>6.17</td>
</tr>
</tbody>
</table>

Tables 4 and 5 contain the descriptive statistics for the three components of the ERP for the natural /u/ (Table 4) and the synthetic /u/ phonemes (Table 5). Similar to the /æ/ phoneme, the mean amplitudes of the P1, N1, and P2 components for the natural /u/ and synthetic /u/ conditions were less than one standard deviation apart from each other. In addition, the standard deviations, minima, and maxima did not show large differences between the natural and synthetic /u/ ERP components. An ANOVA failed to show significant differences between the amplitude and latencies of the ERP components between the natural and synthetic /u/ conditions. When comparing the ERP components between the /æ/ and /u/ phonemes, the mean amplitudes were less than one standard deviation for the natural and synthetic conditions for all three ERP components. An ANOVA failed to show significant differences between the natural and synthetic conditions and the /æ/ and /u/ phonemes for the ERP components.
Table 4

Descriptive statistics for latency, in ms, and amplitude, in µV, of the ERP components of the natural /u/ stimuli across participants.

<table>
<thead>
<tr>
<th>ERP component</th>
<th>M</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1 latency</td>
<td>102.39</td>
<td>22.22</td>
<td>77.80</td>
<td>143.20</td>
</tr>
<tr>
<td>P1 amplitude</td>
<td>1.21</td>
<td>1.17</td>
<td>-0.50</td>
<td>3.73</td>
</tr>
<tr>
<td>N1 latency</td>
<td>143.12</td>
<td>22.25</td>
<td>96.00</td>
<td>179.80</td>
</tr>
<tr>
<td>N1 amplitude</td>
<td>-2.57</td>
<td>1.87</td>
<td>-6.24</td>
<td>1.06</td>
</tr>
<tr>
<td>P2 latency</td>
<td>181.53</td>
<td>32.40</td>
<td>120.20</td>
<td>224.40</td>
</tr>
<tr>
<td>P2 amplitude</td>
<td>1.48</td>
<td>2.22</td>
<td>-2.00</td>
<td>7.57</td>
</tr>
</tbody>
</table>

Table 5

Descriptive statistics for latency, in ms, and amplitude, in µV, of the ERP components of the synthetic /u/ stimuli across participants.

<table>
<thead>
<tr>
<th>ERP component</th>
<th>M</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1 latency</td>
<td>97.23</td>
<td>19.96</td>
<td>68.00</td>
<td>133.80</td>
</tr>
<tr>
<td>P1 amplitude</td>
<td>0.73</td>
<td>0.74</td>
<td>-0.35</td>
<td>2.43</td>
</tr>
<tr>
<td>N1 latency</td>
<td>145.58</td>
<td>21.88</td>
<td>100.00</td>
<td>185.20</td>
</tr>
<tr>
<td>N1 amplitude</td>
<td>-3.08</td>
<td>1.91</td>
<td>-7.11</td>
<td>-0.12</td>
</tr>
<tr>
<td>P2 latency</td>
<td>188.65</td>
<td>32.84</td>
<td>123.00</td>
<td>224.40</td>
</tr>
<tr>
<td>P2 amplitude</td>
<td>1.56</td>
<td>2.08</td>
<td>-1.31</td>
<td>5.97</td>
</tr>
</tbody>
</table>
Table 6 shows the MMN for the natural and synthetic conditions. The MMN was present for each phoneme condition. However, an ANOVA failed to show significant differences in the MMN amplitudes between the /æ/ and /u/ phonemes across the natural and synthetic conditions.

Table 6

*Average latency, in ms, and amplitude, in µV, of the MMN for each vowel condition*

<table>
<thead>
<tr>
<th>Measure</th>
<th>/æ/</th>
<th>/u/</th>
<th>/æ/</th>
<th>/u/</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>201.33</td>
<td>177.43</td>
<td>-1.01</td>
<td>-1.18</td>
</tr>
<tr>
<td>SD</td>
<td>56.91</td>
<td>60.92</td>
<td>1.36</td>
<td>1.24</td>
</tr>
<tr>
<td>Minimum</td>
<td>120.20</td>
<td>103.2</td>
<td>-4.51</td>
<td>-4.03</td>
</tr>
<tr>
<td>Maximum</td>
<td>291.80</td>
<td>263.6</td>
<td>0.79</td>
<td>0.51</td>
</tr>
</tbody>
</table>

**Temporal data analysis.** Two-dimensional brain maps were created for the ERP grand-averaged files (Figures 3 through 6). The brain maps represent areas of processing activity across the surface of the scalp.

Figure 3d (natural /æ/ stimulus) shows some initial processing from 150 to 200 ms over the left mid-temporal area. Between 200 to 250 ms primary processing occurred over the right frontal-temporal areas and spread to the mid-temporal areas (Figure 3e). This was followed by processing in the left mid-temporal areas from 250 to 300 ms (Figure 3f). This processing seemed to diminish by 400 ms (Figure 3i), with late processing that occurred at about 600 ms over the right frontal areas (Figure 3m), maximized at 700 to 750 ms (Figure 2o), and dissipated at 750 to 800 ms (Figure 3p).

Figure 4 (synthetic /æ/ stimulus) shows a similar distribution as seen in Figure 3 (natural /æ/ stimulus). Some relatively focused activity was seen over the left mid-temporal area from
150 to 200 ms (Figure 4d), the right mid-temporal area from 200 to 250 ms (Figure 4e), and again over the left mid-temporal area from 250 to 300 ms (Figure 4f).

The processing seen in Figure 5 (natural /u/ stimulus), although somewhat similar to the processing seen in Figures 3 (natural /æ/ stimulus) and 4 (synthetic /æ/ stimulus), shows significantly greater bilateral activation in the left and right mid-temporal regions from 150 to 200 ms (Figure 5d). This was approximately 50 ms earlier than what was seen in the processing for the natural /æ/ stimulus. In addition, there appeared to be a greater demand for processing in the 150 and 200 ms time frame given the increase in amplitude (Figure 5d). Increased demand for processing was also seen in the mid-temporal areas from 250 to 300 ms (Figure 5f).

Although the processing in the mid-temporal areas dissipated by about 400 ms (Figure 5i), the processing in that area reappeared with more anterior processing in the central temporal areas from 450 to 500 ms (Figure 5j) and dissipated by 550 ms (Figure 5k). Unlike the processing in natural stimuli, there was late processing that began at about 650 ms (Figure 5n). This processing maximized from 700 to 750 ms (Figure 5o) and dissipated at about 800 ms (Figure 5p) in the right frontal areas.

The processing seen in Figure 6 (synthetic /u/ stimulus) was very similar to the processing seen in Figure 5 (natural /u/ stimulus) except earlier mid-temporal processing that occurred from 100 to 150 ms (Figure 6c) for the synthetic /u/ stimulus as opposed to 150 to 200 ms (Figure 6d) for the synthetic /æ/ stimulus. Furthermore, there was an extremely high level of activity in the right anteriorfrontal region for the synthetic /u/ stimulus from 200 to 250 ms (Figure 6e). This high level processing activity dissipated by approximately 250 ms (Figure 6f) and reappeared again from 750 to 800 ms (Figure 6p).
Figure 3. Depiction of the grand-averaged brain maps of the ERPs for the natural /æ/ stimulus. Orange/red indicates areas of greatest positivity ($\geq 2.0 \mu V$) and blue indicates areas of greatest negativity ($\leq -2.0 \mu V$).
Figure 4. Depiction of the grand-averaged brain maps of the ERPs for the synthetic /æ/ stimulus. Orange/red indicates areas of greatest positivity (≥ 2.0 µV) and blue indicates areas of greatest negativity (≤ -2.0 µV).
Figure 5. Depiction of the grand-averaged brain maps of the ERPs for the natural /u/ stimulus. Orange/red indicates areas of greatest positivity (≥ 2.0 µV) and blue indicates areas of greatest negativity (≤ -2.0 µV).
Figure 7 shows the $z$-score distribution for the comparison of the natural /æ/ and natural /u/ stimuli. The most significant processing differences appeared to occur in the frontal areas from 150 to 200 ms (Figure 7d), as well as from 200 to 250 ms (Figure 7e) in the left temporal area and in the right anterior temporal areas. Activity was also seen in the right posterior parietal areas during this same time frame. The $t$-score map for the comparison of the natural /æ/ and natural /u/ stimuli (Figure 8) would suggest that the same time frames and distribution, from 150 to 200 ms (Figure 8e) as well as from 200 to 250 ms (Figure 8f), had the highest $t$-score values.

Figure 9 shows the $z$-score distribution for the comparison of the synthetic /æ/ and synthetic /u/ stimuli. From 20 to 100 ms, greater distribution differences existed in the right frontal areas (Figure 9b). These processing differences extended to about 150 ms (Figure 9c). Greater distribution differences were also located in the right posterior frontal areas from 250 to 300 ms (Figure 9f). Additional distribution differences were also located over the left posterior parietal areas from 400 to 500 ms (Figure 9i, j). Also, late processing differences were seen in the frontal areas from 700 to 800 ms (Figure 9o, p). The $t$-score map (Figure 10) for the same stimulus comparison, as seen in Figure 9, would indicate significant differences between 250 and 300 ms in the right frontal areas (Figure 10f) and in the left parietal areas between 400 and 500 ms (Figure 10i, j).

Figures 11 and 12 show the distribution for the MMN for the natural versus synthetic phonemes within each phonemic representation (i.e., /æ/ and /u/). There was a generalized processing seen across essentially all time periods over the left frontal areas as well as across the left anterior temporal areas for Figure 12.
Figure 6. Depiction of the grand-averaged brain maps of the ERPs for the synthetic /u/ stimulus. Orange/red indicates areas of greatest positivity (≥ 2.0 µV) and blue indicates areas of greatest negativity (≤ -2.0 µV).
Figure 7. Depiction of the grand-averaged z-score brain maps of the ERPs for the natural /æ/ and natural /u/ stimuli. Orange/red indicates areas of greatest positivity (≥ 2.0 µV) and blue indicates areas of greatest negativity (≤ -2.0 µV).
Figure 8. Depiction of the grand-averaged t-score brain maps of the ERPs for the natural /æ/ and natural /u/ stimuli. Orange/red indicates areas of greatest positivity ($\geq 2.0 \, \mu V$) and blue indicates areas of greatest negativity ($\leq -2.0 \, \mu V$).
Figure 9. Depiction of the grand-averaged z-score brain maps of the ERPs for the synthetic /æ/ and synthetic /u/ stimuli. Orange/red indicates areas of greatest positivity (≥ 2.0 µV) and blue indicates areas of greatest negativity (≤ -2.0 µV).
Figure 10. Depiction of the grand-averaged $t$-score brain maps of the ERPs for the synthetic /æ/ and synthetic /u/ stimuli. Orange/red indicates areas of greatest positivity ($\geq 2.0 \mu V$) and blue indicates areas of greatest negativity ($\leq -2.0 \mu V$).
Figure 11. Depiction of the grand-averaged MMN brain maps of the ERPs for the natural /æ/ and synthetic /æ/ stimuli. Orange/red indicates areas of greatest positivity (≥ 2.0 µV) and blue indicates areas of greatest negativity (≤ -2.0 µV).
Figure 12. Depiction of the grand-averaged MMN brain maps of the ERPs for the natural /u/ and synthetic /u/ stimuli. Orange/red indicates areas of greatest positivity (≥ 2.0 µV) and blue indicates areas of greatest negativity (≤ -2.0 µV).
Repeated Measures

Behavioral data and ERPs were acquired a second time from three participants following initial testing to determine test-retest reliability. Individual and grand averaged ERP waveform files were created for each test-retest participant. A $t$-test was conducted on the grand averaged ERP waveform files. No significant differences ($t(3) > 3.18, p > 0.05$) were found in any of the three participants for repeated measures which indicates that an acceptable level of test-retest reliability was established.

Discussion

The purpose of the current study was to evaluate differences in behavioral and electrophysiological responses between natural and synthesized vowels presented to young, normal hearing adults. In addition, the study aimed to determine whether there were distinct differences in spatial-temporal processing in the cerebral cortex using these two types of vowels. The results of the current investigation support the idea that processing differences do occur between nonspeech (artificially synthesized) and speech tasks (natural speech; Benson et al., 2001; DeWitt & Rauschecker, 2012; Horev et al., 2007; Kraus et al., 1992; Tampas et al., 2005). Processing differences were observed in behavioral and electrophysiological responses, which included RTs and ERP amplitudes and latencies. Similar to previous investigations, the MMN was obtained in the current study for both natural and synthetic stimuli and was used to identify differences in cortical spatial-temporal processing of these two types of stimuli (Kraus et al., 1992; Picton, 2011; Tampas et al., 2005). Furthermore, the current study supports the idea of a hierarchical model of speech in which more complex or unfamiliar acoustic stimuli place higher demands of the neural processing network when compared to more simplistic or familiar
Summary and Evaluation of Results

Behavioral differences. Behavioral observations of RTs and error rates made within the current study indicated that there were differences in behavioral responses between natural and synthetic vowel recognition. Table 1 and Figure 2 (inset) demonstrate that RTs are greater for synthetic vowel representations than for natural speech vowels. However, figure 2 also indicated that the greater RTs may be a result of a larger error rate for the synthesized vowels. Reaction time is known to be an indicator of cortical processing time. More specifically, reaction time is an indicator of cognition (Zhang & Li, 2012). There were no significant RT differences in the comparison of the natural /æ/ and natural /u/ stimuli. This is consistent with the idea that cognitive processing of natural vowels occurs somewhat equally between the two phonemes. When synthetic vowels were correctly identified, the cognitive processing time was similar to that seen for natural vowels. However, overall cognitive processing time of synthetic vowels increased as the number of errors increased. Therefore, based upon the increased number of errors seen for the synthetic /u/ condition, greater ambiguity existed when attempting to place synthetic phonemes into categories in which natural phonemes occurred.

Electrophysiological differences. Electrophysiological differences between natural and synthetic vowel processing were also observed in the present study. These differences were primarily seen in the processing of the synthetic /u/ stimuli. P1, N1, and P2 components are primarily influenced by auditory perception (McPherson & Starr, 1993; Picton, 2011). As noted in Tables 2 and 3, ERP peak latencies showed little difference between the natural /æ/ and synthetic /æ/ stimuli. This suggests that auditory perception between the two conditions was
quite similar; however, this suggestion is not seen for the /u/ phoneme. The latencies of the P1, N1, and P2 complex showed increased latency for the synthetic /u/ phoneme. Given the increased RT and larger number of errors observed for the synthetic /u/ phoneme compared to the synthetic /æ/ phoneme, the processing of synthetic phonemes may be dependent upon the phoneme itself.

The scalp distribution of the naturally produced phonemes was essentially the same (Figures 3 and 5). In general, the greatest processing for the natural phonemes occurred between 150 and 300 ms. This observation is consistent with other studies that have used imaging techniques in the study of phonemes and found the greatest processing to occur between 150 and 250 ms (Kasai et al., 2002; Näätänen, 2001; Picton, 2006). The natural /u/ phoneme showed more intense processing between 150 to 200 ms (Figure 5d) and between 250 to 300 ms (Figure 5f) compared to the natural /æ/ phoneme as indicated by their amplitude differences. That is, the amplitude of the natural /u/ phoneme increased between 150 to 200 ms (Figure 5d) and between 250 to 300 ms (Figure 5f); whereas, the natural /æ/ phoneme did not show an increase in amplitude. This increase in amplitude is an indication of greater cognitive demands on the neural network in discriminating the acoustic signal (Horev et al., 2007). The increased amplitude demonstrates that a definite acoustic difference was recognized and that a greater demand was required in order to accurately discriminate the phoneme (Näätänen, 2001; Näätänen, 2008). This observation further suggests that greater demands were placed on the processing of natural /u/ phoneme compared to the natural /æ/ phoneme. A greater processing trend was also seen in the processing of the synthetic /u/ phoneme (Figure 6) compared to the processing of the synthetic /æ/ phoneme (Figure 4).
The two synthetic phonemes also had similar scalp distributions (Figures 4 and 6). The areas of processing were generally similar; however, the demand or complexity of processing was different. The synthetic /u/ phoneme required more complex processing compared to the synthetic /æ/ phoneme. That is, the increase in processing represented by large amplitude for the synthetic /u/ phoneme, especially during the 100 to 150 ms and 200 to 250 ms time frame (Figure 6c, f), may have represented an increase in cognitive demands on the neural network needed in the discrimination of this phoneme (Horev et al.; Näätänen, 2001; Näätänen, 2008).

There were similar scalp distributions for the natural and synthetic phonemes. However, the use of z-score maps and t-score maps helped to identify processing differences. Specifically, z-score maps were used to help identify more specific processing differences and t-score maps were used to define whether the z-score differences were significant (Waragai, Yamada, & Matsuda, 2007). Results from these statistical mapping procedures indicated that there were specific processing differences between natural and synthetic phonemes.

Relative to the phoneme /æ/, significant differences between the natural and synthetic phonemes were observed early in the first 100 ms of presentation. This early processing difference corresponded to the P1 wave and was seen in the right frontal area, suggesting that there were increased processing demands relative to the acoustic differences between the two phonemes. Later processing differences were seen between 250 and 300 ms also in the frontal areas (Figure 11f). The presence of frontal area activation suggests that this particular area may play a role in language processing. This area is commonly responsible for the processing of the MMN and cognitive decision-making (Picton, 2011). Therefore, the frontal lobe activation seen in the processing of the natural and synthetic /æ/ phonemes suggests that a distinct decision making process was required in detecting a difference between these two stimuli. During the
processing of these two stimuli, the frontal regions were alerted to the fact that a deviance to a pattern had occurred and that a decision needed to be made regarding the deviant (Picton, 2006).

Similar to the processing of the /æ/ phoneme, in which the natural and synthetic counterparts were processed differently, significant processing differences also existed between the natural and synthetic /u/ phonemes. These significant processing differences between the natural and synthetic /u/ phonemes were primarily seen between 250 and 300 ms in the frontal areas (Figure 12f). However, unlike the natural and synthetic /æ/ phonemes, significant processing differences were also seen in the right posterior temporal area for both /u/ phoneme conditions, within the first 20 ms (Figure 12a) and suggests some higher brainstem or lower thalamic processing (McPherson & Starr, 1993; Picton, 2011). Although there was some activity in the same area for the phoneme /æ/ within this same time epoch, it did not reach the same level of significance as the /u/ phoneme. Lastly, significant differences between 450 and 500 ms were seen for the /u/ phoneme in the posterior parietal area (Figure 12j). This late processing was not seen in the processing of the natural /u/ stimuli.

In general, the z-score and t-score maps give evidence that synthetic phonemes require additional processing in the frontal areas between 250 and 300 ms (Figures 11f and 12f). This additional processing was represented by an increase in the amplitude of the responses in that area, suggesting greater demands in making decisions regarding those phonemes. Likewise, the two synthetic phonemes were processed somewhat differently. The phoneme /u/ was a more complex stimulus item and therefore required greater processing demand in the cognitive decision-making process. Processing in the auditory system is highly influenced by linguistic content and complexity (Peeva et al., 2010; Siok, Jin, Fletcher, & Tan, 2003). When synthetic phonemes are placed in a linguistic content, the ambiguities of synthetic identification noted in
this study may be influenced by surrounding phonemes in syllables and words. Surrounding phonemes may give further information to listeners regarding the identity of ambiguous phonemes; therefore, synthetic phonemes placed in a linguistic content may show to have more similar processing demands to that of natural phonemes in similar linguistic contents.

Latencies between 400 and 800 ms are known to represent linguistic processing (Picton, 2011). No significant processing was seen beyond 500 ms for either the natural or synthetic phonemes; therefore, there did not appear to be any linguistic processing demands made by these phonemic recognitions. Speech perception relies heavily on large linguistic units, such as syllables and words. These syllables and words have unique recognition patterns that are stored as memory traces in the cortex (Näätänen, 2001). Processing of these large linguistic units requires late, more complex processing because of the need to recognize and recall those memory traces (Näätänen, 2001; Neville et al., 1998). Therefore, the absence of significant latencies between 400 and 800 ms suggests that there were no required linguistic processing using isolated vowel phonemes.

The MMN is a processing negativity that results from acoustic and perceptual differences (Picton, 2011; Tampas et al., 2005). This negativity originates in the temporal and frontal lobes and occurs between 150 and 300 ms (Picton, 2011). Figures 11 and 12 depict the difference waveforms used to identify the MMN. These figures give evidence to two important findings. First, the left central frontal lobe was activated differently throughout the entire 800 ms time frame. The continued processing in the frontal lobe, which was responsible for executive decision-making, would strongly suggest that the natural and synthetic phonemes are continually being evaluated (DeWitt & Rauschecker, 2012; Näätänen, 2001). This continuous evaluation was reflected in the increased latency and RT seen in the processing of the synthetic vowels.
Furthermore, additional processing was seen in the anterior central temporal lobe for the phoneme /u/, which is consistent with known processing that occurs in the superior temporal gyrus. The superior temporal gyrus plays a role in linguistic and phonological processing (Benson et al., 2001; DeWitt & Rauschecker, 2012). Increased speech complexity has been found to be associated with increasing activation of the left superior temporal gyrus (Benson et al., 2001). Therefore, the processing of the phoneme /u/ required greater cognitive activation because of the acoustical complexity resulting from the ambiguity between the natural and synthetic counterparts of this phoneme. This greater ambiguity, and thus greater cognitive activation, resulted in greater activation of the left temporal lobe.

In conclusion, this study demonstrated that processing differences do exist between natural and synthetic vowels (Benson et al., 2001; DeWitt & Rauschecker, 2012; Horev et al., 2007; Kraus et al., 1992; Tampas et al., 2005). Synthetic vowel perception requires increased activation within the neural system. However, synthetic vowel perception essentially maintains the same neural processing network as the natural vowel processing. This supports the idea behind the hierarchical model of speech which indicates that as the processing of a phoneme becomes more complex, higher levels of neural activity are required in order for accurate processing to occur (Benson et al., 2001; DeWitt & Rauschecker, 2012; Kraus et al., 1992; Näätänen, 2001). In addition, the study demonstrated the necessity of continuous frontal lobe involvement in the perceptual and cognitive processing of vowel stimuli. Finally, the study found no evidence of linguistic processing using isolated vowel phonemes.

**Recommendations for Future Research**

The current study expanded on previous studies relative to the processing of natural and synthetic speech by providing spatial-temporal information in addition to behavioral and
electrophysiological data (Benson et al., 2001; DeWitt & Rauschecker, 2012; Horev et al., 2007; Kraus et al., 1992; Tampas et al., 2005). The spatial-temporal information found in the current study provides additional theoretical information about the distinct processing of natural and synthetic speech. While the current study adds valuable information to research regarding nonspeech and speech processing, the study has limitations which call for further investigation. First, the present work utilized natural and synthetic phonemes in isolation, which is not consistent with the linguistic contexts used by individuals on a daily basis. Second, this study did not assess the preattentive or late linguistic processing differences of natural and synthetic phonemes that may have occurred. The study also only utilized one specific ERP component (MMN); therefore, analysis of additional ERP components could provide further information regarding nonspeech and speech processing. Lastly, the current study only utilized native English listeners and did not consider whether listeners of a different primary language would process natural and synthetic phonemes differently as well.

Future researchers may wish to expand upon the current study by examining processing of natural and synthetic speech in more linguistically demanding contexts, such as in syllables or words. Comparison of natural and synthetic speech in larger linguistic units will provide greater information regarding the effects of speech naturalness on complex linguistic processing. This may yield additional processing information, which was absent in the current study because of the use of isolated vowel phonemes.

Future studies may wish to compare the effect natural and synthetic phonemes have on preattentive and late linguistic processing. Natural and synthetic phonemes can be placed in larger linguistic units and electrophysiological information could be gathered regarding processing using the P300 and N400 responses. These responses can be used to compare
processing that takes place preattentively to processing that occurs later. This information could be compared to the results of the current study and its use of the MMN to determine spatial-temporal processing of natural and synthetic phonemes.

Further investigation may also be warranted to determine whether native language plays a significant role in processing of natural and synthetic phonemes. Previous studies have determined that native language plays a role in categorical perception (Bomba et al., 2011; Horev et al., 2007; Näätänen, 2001). Therefore, presenting natural and synthetic phonemes to two different native language speakers may provide interesting information regarding the effect of synthetic speech on categorical perception in nonnative speakers. This information could be compared to previous studies to determine whether synthetic phonemes, that are nonphonemic in a particular language, were processed differently by nonnative language speakers compared to native speakers.

The current study provides further insight into the processing of natural and synthetic speech by providing temporal timing and cortical processing information. The study found that processing differences do exist between natural and synthetic vowels and that these processing differences were observed in both behavioral and electrophysiological responses. In addition, the study concluded that synthetic vowels resulted in increased processing demands within the neural system compared to the processing of natural vowels. However, the study also concluded that there were no late linguistic processing demands using either natural or synthetic isolated vowel phonemes. Lastly, the study demonstrated the importance of continuous frontal lobe involvement in cognitive processing of vowel stimuli. While the current study builds on previous processing studies of natural and synthetic phonemes, certain questions remain unanswered including: Does placing synthetic phonemes in linguistically relevant content have
an effect on processing? Does linguistically relevant content using synthetic phonemes make processing more similar to processing of natural speech? Does linguistic processing of synthetic stimuli result in greater demands of preattentive and late linguistic processing? Future investigations are recommended for resolving these questions and for furthering the general information of linguistic processing.
References


Näätänen, R. (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology, 38*(1), 1-21. doi: 10.1111/1469-8986.3810001


Waragai, M., Yamada, T., & Matsuda, H. (2007). Evaluation of brain perfusion SPECT using an easy Z-score imaging system (eZIS) as an adjunct to early-diagnosis of
neurodegenerative diseases. *Journal of the Neurological Sciences, 260*(1–2), 57-64. doi: 10.1016/j.jns.2007.03.027


Appendix A

Annotated Bibliography


**Objective:** The purpose of this article was to explain the Signal Detection Theory (SDT) and how it can be used to analyze data in research. **Conclusions:** The SDT can be used to analyze data in research when data was generated by a known process or by chance. In the SDT Model there are four different types of responses. The first response is a “hit.” A response is considered a “hit” if there was a signal present and a participant responds yes, a signal was present. A response is considered a “miss” if there was a signal present and the participant responds no, there was not a signal. A response is considered a “false alarm” if a signal was absent and a participant responds yes, a signal was present. Lastly, a response is considered a “correct rejection” if a signal was absent and a participant responds no, there was not a signal. There are two parameters that affect the SDT. The first parameter deals with the difficulty of the task. An easier task will have a larger proportion of “hits” and a smaller proportion of “false alarms.” The second parameter is the strategy of the participant. A participant can be considered liberal or conservative. A liberal participant tends to respond yes to a task. A conservative participant tends to respond no. Overall, the SDT Model attempts to estimate the effect of these two parameters on experimental tasks. **Relevance to current work:** The SDT model was used in the current work to categorize participants’ responses in different tasks. The four responses of the SDT were also used to determine the sensitivity and bias of participant performance in the current work. **Level of evidence**¹: N/A.


**Objective:** The American Speech-Language Hearing Association (ASHA) has published specific guidelines regarding screening and assessing individuals for hearing impairments and disorders. These guidelines are set forth to ensure ethical practice in conducting hearing screenings. In addition, these guidelines ensure that results of hearing screenings are interrupted the same nationwide. **Relevance to current work:** Each participant in the current study had a hearing screening in order to be considered for additional EEG investigation. The guidelines set forth by ASHA were followed in the participants’ initial hearing screenings. **Level of evidence:** N/A.

---

¹ Each article reviewed was given a ranking based on the strength and design of the study. This is referred to as the Level of Evidence. Supplementary information is available from the following sources: ASHA. (2013). Introduction to Evidence-Based Practice. ASHA.org. Retrieved August 26, 2013, from http://www.asha.org/Members/ebp/intro.html and Straus, S. E., Glasziou, P., Richardson, W. S., & Haynes, R. B. (2011). Evidenced-Based Medicine (Fourth ed.). London: Chruchill Livingstone (Elsevier).

**Objective:** The purpose of the study was to identify brain activation regions during speech perception tasks using functional magnetic resonance imaging (fMRI). **Methods:** Auditory stimuli was presented to 12 typical adults (six male and six female between ages 21-45 years) while fMRI scanning was conducted. Each participant listened to ten trials of stimuli. The trials consisted of stimuli that alternated between three characteristics. First, the stimuli was either speech or nonspeech. The speech stimuli consisted of nonsense syllables varying in phonetic complexity from steady-state vowels, to consonant-vowel syllables, to consonant-vowel-consonant syllables. The nonspeech stimuli was tonal and increased in complexity from notes, to chords, to chord progressions. Second, the stimuli was either natural or synthetic in nature. The synthetic speech was identical to natural speech up to 3500 Hz. Last, the stimuli varied between three levels of complexity. Analysis was performed on the fMRI scans to determine regions of interest. **Results:** Bilateral posterior superior temporal regions, which were highly left lateralized superiorly in the supramarginal gyrus (SMG), were highly activated in response to speech and chord progression stimuli. This suggests that both phonetic and acoustic processing may occur within these regions. These areas were also activated with nonspeech stimuli; however, activation strength was not as great for the nonspeech stimuli. Differences in activation strength were also seen in the natural and synthetic speech stimuli comparison. The left dorsolateral precentral sulcus was more active for natural speech compared to synthetic speech. Also, the right secondary auditory cortex was more active for natural speech compared to synthetic stimuli. These two regions could be involved in processing paraphonetic attributes that are only found in natural speech. **Conclusions:** The study identified seven spatially distributed brain regions which were strongly associated with speech processing. These regions included the following: the left posterior superior temporal gyrus, the right posterior superior temporal gyrus, the left ventral occipitotemporal cortex, the left inferior SMG, the left middle frontal gyrus, the right superior SMG, and the right anterior insula. The posterior superior temporal gyrus, or Wernicke’s area, was most selective in speech and phonetic complexity tasks. The ventral occipitotemporal and right insula regions were strongly associated with speech complexity dimension. This indicated a possible speech or phonetic procession function for these two reasons. Lastly, the left middle frontal gyrus had a role in differentiating between natural and synthetic speech. **Relevance to the current work:** This study played an important role in the current work by demonstrating that different areas of the brain are involved in different acoustic listening tasks. Different areas of the brain activated in response to natural speech compared to synthetic speech; however, this research does not address the temporal aspect differences of natural and synthetic speech processing. The purpose of the current study was to address the lack of information on temporal processing differences between natural and synthetic speech as well as verify dipole localization processing differences between natural and synthetic speech. **Level of evidence:** Level IIIa.
Objective: The purpose of this study was to compare the performance differences of dyslexic and control participants in natural and synthetic speech listening tasks. The study also examined whether a categorical-perception deficit was present in dyslexics and whether this deficit would generalize to a natural-speech continuum. Methods: Ten children, with previous diagnosed dyslexia, and 12 typically reading children were recruited for the study. Two speech-sound continuums were created for testing purposes. Each continuum consisted of seven steps between the two stimuli syllables. One continuum ranged from /ta/ to /ka/ and was a semi-synthetic utterance spoken by a female speaker of standard Dutch. The other continuum ranged from /da/ to /ba/ and was created using a synthetic source-filter model. The participants responded to the stimuli in a two alternative forced choice paradigm. During the first task, the semi-synthetic /ta/-/ka/ stimuli was presented to the participants. During the second task, task was repeated using the synthetic /ba/-/da/ stimuli. Results: The results were coded in percentage responses in the synthetic and natural stimulus condition. These measures were then analyzed according to stimulus type (natural vs. synthetic) and continuum (the seven steps of each continuum) within the subject variables and groups. The analysis revealed a significant main effect of continuum but no significant effect of group. In addition, there was not a significant interaction between group and continuum. The normal-reading group perceived the first two steps of the continuum and the last three steps of the continuum more consistently than the dyslexic group. There was also a significant effect in regards to stimulus type. The main effect of stimulus type showed that the synthetic continuum was perceived more categorically than the natural continuum. Conclusions: In the experiment it was found that the natural-speech continuum was more difficult to categorize for both groups than the synthetic speech continuum. This occurred because the residual formant transitions indicating the original place of articulation conflicted with the manipulated formant transitions, which was altered on the natural utterance. Because of this conflict, the natural-speech continuum was more difficult to perceive. On the other hand, the well-controlled, one-cue-varying synthetic-speech continuum was easier to categorize for both groups. It was also concluded that deficits in the perception of synthetic speech do not generalize to the perception of more naturally sounding speech. Relevance to the current work: The overall conclusion of this study was that deficits in perceiving synthetic speech did not generalize to the perception of natural speech. This evidence reinforces the fact that processing of natural and synthetic speech is independent of one another and likely involves different processes entirely. The purpose of the current study was to verify the temporal differences in processing between natural and synthetic speech stimuli. Level of Evidence: Level IIIb.


Objective: The purpose of this study was to examine MMN differences in certain phonemes between English and Japanese speakers. Second, the study compared MMN of glides, liquids, and vowels in native and non-native English speakers. By examining these two components, the study overall examined how different types of vowels and consonant-vowel phonemes are processed in the brain. Methods: Sixteen adults participated in the study including eight native
English speakers and eight native Japanese speakers who learned English after the age of 12 years. The stimuli consisted of vowels and consonant-vowel syllables presented in separate sequences. For all sequences, 1000 stimuli were presented randomly and consisted of a standard stimulus and a deviant stimulus. The standard stimuli were presented 85% or the time and the deviant stimuli 15% of the time. The Standard English vowel was /iy/ and the deviant vowel was /i/. The consonant-vowel syllables consisted of a standard set (/da/ and /ra/) and a deviant set (/wa/ and /la/). EEG was recorded for each participant from 26 electrodes. Results: When analyzing the vowel set there was no significant differences between native and non-native speakers on MMN latency and amplitude. For the /i/ versus /iy/ condition, there was a clear MMN along the frontal chain (F3, Fz, F4) and vertex (Cz) for both native and non-native speakers with a latency at about 200 ms. For the /da/ versus /wa/ syllables, MMN latency showed no significant differences between native and non-native speakers. For the native English speakers in the /da/ versus /wa/ condition, there was a MMN response along the frontal chain (F3, Fz, F4) and vertex (Cz) electrodes with a latency just greater than 200 ms. For the non-native speakers the MMN was smaller in amplitude along the frontal chain and almost impossible to identify at the Cz electrode. The /ra/ versus /la/ condition showed low amplitude MMN compared to the other stimulus conditions in native speakers. A clear MMN response was seen in 75% of the native participants with 25% showing poor MMN-like responses. For the non-native speakers, the MMN was absent in 50% of the subjects. In 38% of the subjects the MMN had extremely low amplitudes. Conclusions: It was found that native English speakers had larger amplitude MMNs than the non-native speakers in the consonant-vowel syllable conditions. However, there was no difference in the MMNs in the vowel condition. This data showed that vowels and consonants are processed differently in the brain as measured by MMN. The differences found in the MMNs of the consonant-vowel syllable conditions between the two groups showed that neural differences exist in phonemic processing between speakers of different languages depending on the level of exposure of a particular phoneme. Relevance to Current Work: The previous study indicated that phonemes are perceived differently based on a person’s native language. This evidence suggested that participants in the current study had to be native English speakers in order to avoid differences that may arise in the data due to language knowledge and memory. Level of evidence: Level IIIa.


Objective: The purpose of this study was to examine the effects of synthetic speech processing of a listener and observe the listener’s ability to segregate competing noise in an acoustic mixture. Methods: The study had eight normal-hearing participants (three males, five females ranging from 21 to 55 years). The speech materials used were synthesized recordings from the Coordinate Response Measure (CRM) corpus. This corpus had a total of 208 different phrases spoken by both male and female speakers. Each phrase in the corpus was digitally preprocessed to produce three additional kinds of speech signals. The three signals are the following: sine-wave speech (SWS), modulated sine-band speech (MSB), and modulated noise-band speech (MNB). Each stimulus presented to the participants consisted of two simultaneous phrases from the CRM corpus. A target sentence was generated along with a masker sentence. The masker sentence was either spoken by the same talker as the target phrase (TT), a different same-sex talker (TS), or a different-sex talker (TD). A normal speech control condition was also included
which paired the target and masker sentences from the original unprocessed CRM corpus. Each condition was produced with five different signal-to-noise ratios (SNR). The participants were presented with the stimuli through headphones and identification of the stimuli was a forced-choice identification task. Forty-eight trials per subject at each SNR in each condition were presented to the participants. Results: The experiment identified that the performance of the participants was consistently the best with normal speech. The analysis indicated that all three types of synthetic speech signals resulted in significantly poorer performance than normal speech in the presence of negative SNR values. The synthetic speech measurements indicated that listeners were not able to selectively attend to the quieter talker when synthetic speech stimuli were used instead of natural speech. Performance in the three synthetic speech conditions was dependent on the SNR value. At positive SNR values, performance was slightly better in the MNB and MSB conditions than in the SWS condition. At negative SNR values, the SWS condition was somewhat better than the MNB and MSB conditions. No difference in performance was found between the MSB and MNB conditions at any SNR value. The study also found that there were significant differences in performance across the TT, TS, and TD conditions for SWS and normal speech, but not for MNB and MSB speech. Conclusions: It was concluded that normal speech conditions were the most optimal for listening performance. The experiment also found that speaker-dependent voice characteristics that listeners use to segregate competing speech signals, specifically those that help listeners distinguish between male and female voices, were more faithfully preserved in SWS speech than in MSB or MNB speech. The sine-wave speech retained enough acoustic information that the listeners were able to distinguish male and female speakers and distinguish between male and female talkers. The most prominent outcome of the study was that listeners could easily use level differences to selectively attend to the quieter of two same-sex or same-talker natural speech signals but were not able to use level cues to segregate the quieter of two synthetic voices. This possibly occurred due to a spectral overlap which increased the effect of masking on the signal. Overall the study concluded that natural speech had the best performance and allowed the listener to most effectively use segregation cues in order to decipher differences in stimuli. Out of the three synthetic forms, listeners identified phrases in the SWS the best, but the performance in the SWS listening task was still significantly low compared to natural speech. Relevance to the current work: Performance differences between synthetic and natural speech give evidence that these speech stimuli are processed differently. This supported the current study’s hypothesis that natural and synthetic speech are processed differently. Level of evidence: Level IIIa.


Objective: The purpose of this study was to conduct an identification accuracy experiment with synthetic vowels with time-varying characteristics that are representative of natural speech. The synthetic vowels were based on vocal tract area functions and enhanced by allowing change in the area function to change over the duration of each vowel. Methods: Time-varying formant frequencies were obtained from 11 American English vowels. The first two formant frequencies of the vowels were estimated over time and manually adjusted so that the formant contours of F1 and F2 were aligned with the centers of their respective format bands. Also, recordings were used from an earlier study done by Bunton and Story in 2009. Coefficient-to-formant mapping was used to generate sets of area functions for the stimuli. In total there were 88 time-dependent
area functions generated. A synthetic vowel sample was generated for each time-varying area function with a voice source model along with a wave-reflection model of the trachea and vocal tract. There were ten listeners (5 male and 5 female) who listened to the vowel stimuli presented. Participants used a computer mouse to select one of the 11 English vowels on the computer screen following the presentation of the target vowel. Each listener heard five repetitions of each vowel sample blocked by speaker sex in random order. **Results:** The mean identification accuracy across all vowels for individual speakers ranged from 79% to 87%. A confusion matrix was created for each listener and complied across speakers to form a composite confusion matrix. Vowel confusions usually occurred in adjacent vowel categories in the vowel space. The results were then compared to the static vowels used in the Bunton and Story (2009) study. It was found that seven of the vowels synthesized with a time-varying vocal tract shape were identified more accurately than the static vowels used in the Bunton and Story study. The increase in accuracy of the time-varying vowels over the static cases was 50% or more for most of the vowels. The corner vowels [i, æ, u] were less accurately identified with the time-varying vowel than when it was static. **Conclusions:** The results of the study conclude that time-varying area functions produce vowels that are more identifiable than those produced with static area functions. This goes along with the idea that time-varying formant transitions and vowel duration are important cues for identification accuracy. Second, the study concluded that a model of the vocal tract area function can be useful in creating synthetic speech sounds by helping create sounds with more natural-like characteristics. **Relevance to current work:** The above mentioned study demonstrated that the closer a stimulus’s features are to natural speech; the more accurately that stimulus is identified. This was relevant to the current study because it suggested that different stimuli are processed differently by the brain. Also, stimuli constructed similarly to natural speech are perceived more accurately than deviant stimuli. Like this study, identification performance was also observed in the current study. **Level of evidence:** Level IIIa.


*Objective:* CURRY 6 software was created to reconstruct sources of electrical cerebral activity by overlaying temporal electrophysiological data onto brain images. **Relevance to current work:** CURRY 6 software was used in the current study to locate areas of electrical activity and localization.


*Objective:* The NeuroScan 4.5 software was designed to record EEG data as it is collected. This software also has the ability to analyze, filter, and edit data after collection. **Relevance to current work:** NeuroScan 4.5 software was used during EEG data collection in the current study.


*Objective:* The purpose of Stim 2 software was to stream auditory stimuli from the computer to the participant. **Relevance to current work:** Stim 2 software was used in the current study to present the stimulus to the participant.
Objective: The purpose of this study was to determine the accuracy of topographic electroencephalographic (EEG) maps for assessing local cerebral function. Methods: Six right-handed male subjects (age 22 to 30 years) with no history of neurological or psychiatric disease were recruited for the study. The study collected EEG data, via an ElectroCap with 35 scalp electrodes, while simultaneously performing a positron emission tomography (PET) scan. The EEG data was processed using three different montages. The montages examined were a linked ears reference montage, a source-derived montage incorporating only immediately neighboring electrodes, and a bipolar electrode montage. Along with the three montages, two EEG power measures (absolute and relative power) were examined. Each subject was studied in resting conditions and in activated states. To examine activation states, the subjects squeezed a foam ball for a period of two minutes. Data was collected with eyes both closed and opened as well as the ball being squeezed in both the right and left hand. Results: It was found that cerebral perfusion and absolute power had a much weaker association than the relationship between perfusion and relative power. The linked ears reference montage had the weakest perfusion relationship compared to the other two montages. The strongest relationship between power and perfusion was seen for the reattribution method of averaged bipolar power montage. Conclusions: The findings of the experiment led the authors to two conclusions. First, topographic EEG mapping can accurately reflect local brain function. Second, the choice of EEG measure and montage may have a significant influence on the degree in which maps reflect local activity and function. Relevance to current work: The study indicated that EEG is an accurate way to measure cerebral function. This finding was of significant relevance to the current study because EEG was used in mapping cerebral responses to the different stimuli.

Level of evidence: Level IIIa.


Objective: The purpose of this study was to evaluate whether the MMN response to auditory stimuli (speech and non-speech) was deviant in individuals with aphasia. Also, the study aimed to determine whether impairment was due to a deficit in phonemic processing or related to phonetic features. Overall, the study evaluated how aphasic individuals processed language and how it deviated from typical. Methods: The study involved four diagnosed aphasic patients and four neurologically unimpaired, control subjects. Three different types of stimuli were presented to the participants. The stimuli consisted of pure tones, front vowels, and consonant-vowel (CV) syllables. Event-related potentials were recorded for each individual with 21 electrodes using Neuroscan software. The ERPs were collected as the participants were presented auditory stimuli. While the stimulus was being presented, the participants were distracted by a visuo-motor. Results: In all control subjects a reliable MMN was collected for all stimuli. The four aphasic participants all had MMN abnormalities. Specifically, the MMN elicited by pitch deviations was not significant enough to distinguish between the aphasic patients and the control
group. The MMN elicited by consonant contrasts proved to show the most significant difference in aphasic patients in comparison with the control group. Lastly, a significant difference was seen in the MMN elicited by voicing and place of articulation. Aphasic participants showed great anomalies in this MMN compared to the control group. The MMN collected from the aphasic participants was either limitedly distributed, distorted, or completely missing.

**Conclusions:** This study concluded that the MMN elicited by contrasting features reflects deficient processes due to damaged or disconnected regions of the language-processing network seen in those with aphasia. MMN responses collected in individuals with aphasia were clearly deviant compared to those with unimpaired neurological systems; thus, demonstrating the affect that neurological damage has in brain processing. **Relevance to current work:** This research demonstrated the difference in brain processing that occurred when an individual was affected by neurological impairment. This demonstration supported the current study’s exclusion of individuals with known neurological, cognitive, or learning impairments because it would have an impact on the accuracy of data. **Level of evidence:** Level IIIb.


**Objective:** The objective of this study was to demonstrate that preference for complex sounds emerges in the human auditory ventral stream in a hierarchical fashion. **Methods:** This was a unique experiment because instead of performing an actual experiment, the authors researched experiments that were done in the past. The authors searched for publications of experiments in peer-reviewed journals which dealt with topics such as phonetics, speech sounds, and auditory words. After meeting specific criteria, the authors analyzed 115 studies. For determining comprehension sensitivity (CS), the authors compared activation of speech stimuli with activation of matched, nonnaturalistic control stimuli (i.e., tones, noise, and complex artificial nonspeech stimuli). For analyzing repletion suppression (RS), the authors compared activation of repeated and nonrepeated stimuli. For invariant representation (IR) analysis, the authors looked at studies where the RS had a design in which it repeated stimuli varied acoustically but not phonetically. For areal specialization (AS) analysis, a comparison was made of the activation to speech stimuli to other naturalistic stimuli (i.e., music, animal calls, tools sounds). Lastly, foci concordance was assessed by the method of activation likelihood estimation (ALE) in a random-effects implementation that controlled for within-experiment effects. **Results:** It was found that when dealing with CS, spatial concordance was greatest in the superior temporal gyrus (STG) and the superior temporal sulcus (STS) across each phonetic length-based analysis. Phoneme-length effects were discovered to be most strongly associated with left mid-STG and word-length effects with left anterior STG. The study also found that there is a possible processing gradient for phonetic stimuli that progresses from mid-STG to anterior STG and then into STS. The effects of phoneme- and word-length stimuli were also observed. It was determined that for both of these stimuli, concordant foci were observed to be strictly left-lateralized and exclusively within STG. Peak concordance for phoneme-length stimuli was seen in mid-STG and word-length stimuli peak concordance was seen in anterior STG.

Colocalization for CS, IR, and AS was also analyzed. Robust CS effects were observed in the STG and STS and significant concordance was also found in the left frontal cortex. In the IR analysis, localizing occurred to left mid- to anterior STG. The AS analysis found that foci were concordant in the STG and STS and a left-lateralized ventral prefrontal result was observed.
**Conclusions:** The study concluded that there was a left-hemisphere optimization for speech and an anterior-directed processing gradient. Elemental phonetic processing was most strongly associated with mid-STG. Auditory word-form processing was most strongly associated with anterior STG. Phrasal processing was most strongly associated with anterior STS. It was also concluded that colocalization in mid- to anterior STG occurs for CS, IR, and AS. All findings of the study supported the presence of an anterior-directed ventral-stream pattern-recognition pathway. **Relevance to current work:** Differences in speech processing were expected to be observed in the current study between the different stimuli. **Level of evidence:** Level I.


**Objective:** An Electro-Cap is a 32-channel cap that measures electrical activity (in µV) at specific locations around the scalp. **Relevance to current work:** Each participant in the current study was fitted with an Electro-Cap in order to collect electrophysiological data.


**Objective:** The study examined the processing of tones and vowels on the evoked magnetic field and its corresponding event-related potential. **Methods:** Eleven normal hearing, right-handed individuals participated in the study. Synthetic German vowels (/a/, /i/, /u/, /æ/, and /oe/) of two different durations (600 msec and 45 msec) were created using a Klatt synthesizer. In addition to these vowels, a 1000 Hz tone frequency was also used as a stimulus. Participants listened to the randomly presented long and short synthetic vowels and tone while lying on their side. The participants were asked to count the occurrence of the long and short vowel /oe/. Each participant was fitted with a 37-channel neuromagnetometer to collect neuromagnetic data in response to the different stimuli. Recorded data were selectively averaged for each stimulus. In addition, epochs were extracted from the data and equivalent current dipoles were analyzed. Amplitudes and latencies for waveforms were also calculated. **Results:** Dipole strength was noted to be greater in the left hemisphere than in the right hemisphere for processing of the tones and the vowels. The strength of the dipole in the vowel condition was found to be stronger than in the tone condition. **Conclusions:** From the results of the study, the authors concluded that speech-related processing predominantly occurs in the left hemisphere. The difference in the strength of processing between the vowel and tone stimuli could be related to the physical characteristics of the stimuli or the amount of attention the participant paid on the stimuli. Lastly, brain response during tone and vowel stimulation was sensitive to features of speech. **Relevance to current work:** This study demonstrated that although research has been done that has observed the processing of synthetic vowels; more research is needed to look at the temporal processing of vowels. The temporal processing of vowels was measured in the current study. **Level of evidence:** Level IIIa.
Brain Mapping of the MMN to Vowel Variances


**Objective:** This study was designed to investigate whether voicing perception is influenced by linguistic experience or innate temporal sensitivity. To determine this, the study used behavioral and EEG measures of speech categorical perception to compare the perception of a Voice-Onset-Time (VOT) continuum with the perception of an analogous nonspeech continuum. **Methods:** There were fourteen (7 mean and 7 women), healthy, right-handed native Hebrew speakers who participated in this study. Each participant completed two tasks involving identification and discrimination of stimuli which included a behavioral portion and a concurrent behavioral and electrophysiological portion. During the task, participants were instructed to respond as quickly and as accurately as possible by pressing a button on a response box. Stimulus type, behavioral response, and the EEG data (using a 19 electrode Electro-Cap) were collected for each individual during the experiment. During the study, two sets of stimuli were presented to the participants. These stimuli include speech stimuli (forming a natural VOT continuum) and nonspeech stimuli (forming an analogous synthesize continuum [Formant-Onset-Time]). The syllables /ba/ and /pa/ were recorded by a male Hewbrew speaker. The VOT continuum was created by editing natural productions of /ba/ and /pa/ and the Formant-Onset-Time (FOT) continuum was created by using a source-filter synthesis function. **Results:** The behavioral identification data collected for the VOT and FOT were similar. However, differences in the two stimuli were seen in the reaction times and within-category discrimination. The results of the EEG showed that there were significant differences in the VOT and FOT. N1 latency and N1 and P2 amplitudes were affected by the VOT continuum and P2 was affected by FOT continuum. Another important result found was that P3 was present in all speech stimuli but was not present in nonspeech stimuli. **Conclusions:** The study concluded that because VOT discrimination performance was best near phonemic boundaries, then categorization must be based on language-specific phonemic knowledge and experience. The differences in the EEG data between VOT and FOT suggested that brain processing of nonspeech and speech signals differed in the early stages of processing. **Relevance to current work:** The results found from the above study supports the idea that nonspeech and speech stimuli are processed differently by the brain. These differences were demonstrated in the study in the early stages of identification and discrimination. The work completed was relevant to the current study as it aimed to verify whether brain processes do differ in response to different stimuli. **Level of evidence:** Level IIIa.


**Objective:** The study was done in order to clarify the role of gender differences in auditory MMN by comparing the amplitude, latency, and topography of tonal and phonetic MMN. **Methods:** The experiment included 18 male participants and ten female participants, all of which spoke Japanese as their first language. Auditory ERPs were the index used to measure the MMN. The participants were presented with auditory stimulus sequences consisting of standard and deviant stimuli delivered randomly. The exception to the random pattern was that each deviant stimulus was preceded by at least one standard stimulus. The subjects were instructed to
watch a silent film and were encouraged to ignore the stimuli. After the film, the subjects were required to report on the content of the film to ensure their attention was on the film. They also were required to report the characteristics of the stimulus sequence to ensure they behaviorally perceived the duration of tones and the phoneme boundaries. The experiment looked at two conditions. First, it looked at the MMN in response to a duration change of pure-tone stimuli. Second, it looked at the MMN in response to an across-category vowel change. The EEG recording was done via a 128-electrode cap. The MMNs were measured using the difference waveforms obtained by subtracting the ERPs of standard stimuli from those of deviant stimuli. **Results:** The mean global field power peak latencies of the male and female groups were 162 ms for the pure-tone MMN and 156 and 170 ms for the phonetic MMN. These results indicated that there is no significant effect of gender on either pure-tone or phonetic MMN amplitude. The MMN topography also indicated that there were no differences between genders but there were differences between conditions. The latency of the MMN also did not show difference between genders but showed a difference between conditions. The latencies of the pure-tone MMN were significantly longer compared to the phonetic MMN, but this was found in both genders. After the experiment, all participants reported they could concentrate on the film and reported the content of the film correctly. All of the participants also reported correct information about the stimuli they heard while watching the film. **Conclusions:** The experiment concluded that there is no effect of gender on the amplitude, latency, or topography of tonal and phonetic MMN in normal adults using EEG. The conclusion reached allows researchers to know that combining males and females in experiments will not have obscure effects. The study also concluded that the pure-tone MMN was generated from Heschl’s gyrus and the phonetic MMN was generated from the planum temporal. 

Relevance to current work: The current study used both male and female participants. The study summarized above observed that there is no gender difference in MMN measurements. This was important information to the current work because it confirmed that using both males and females in the study would not affect the collection or accuracy of the data. **Level of evidence:** Level IIIa.


**Objective:** The purpose of this study was to confirm whether an association of language lateralization and handedness in healthy subjects exists. **Methods:** Three hundred twenty-six (198 females and 128 males) subjects participated in this study. All subjects were assessed for handedness using the Edinburgh Inventory and for hemispheric language dominance using functional transcranial Doppler ultrasonography (fTCD) during a word generation task. The fTCD was validated and tested for reproducibility by comparing results with intracarotid amobarbital injection and fMRI. **Results:** The study found that the distribution of hemispheric language dominance varied with the degree of handedness. The more right-handed subjects were on the Edinburgh Inventory, the lower the relative incidence of right-hemisphere language dominance, and vice versa. Left-handers demonstrated a higher incidence of right-hemisphere language dominance compared to strong right-handers. **Conclusions:** This study was the first study to examine the effects of language dominance in healthy individuals. The study concluded that left-handedness was neither a precondition nor a necessary consequence of right-hemisphere language dominance. However, the degree of handedness was linearly and highly significantly related to the side of language dominance; therefore, left-handedness increased the likelihood of
right-hemisphere language dominance. **Relevance to current work:** Only right-handed individuals were included in the current study. This requirement was supported by the above study because the study demonstrated that left-handed individuals may possibly have right-hemispheric language dominance. If left-handed individuals participated in the current study, it may have had a negative effect on the data because of the difference in cerebral language dominance. **Level of evidence:** Level IIIa.


**Objective:** The research performed was for the purpose of investigating the role of thalamocortical pathways in guinea pigs by eliciting MMN with speech-like signals. **Methods:** Fifteen, anesthetized guinea pigs were used for the experiment. Each guinea pig was fitted with epidural silver bead electrodes in order to record the surface auditory evoked potentials. The stimuli used to elicit the MMN were synthesized speech contrasts /ba/-/wa/ and /ga/-/da/. This speech stimulus was delivered monaurally through insert phones to the guinea pigs in a modified oddball paradigm. **Results:** It was observed that the /ba/-/wa/ contrast had the most significant changes in MMN in the caudomedial portion of the medial geniculate nucleus. Neither stimuli contrast elicited an MMN from the primary portion of the medial geniculate or the surface over the temporal lobe. In contrast, the midline surface had an elicited MMN in response to both stimuli contrasts. **Conclusions:** The findings of the study suggested that there is a hierarchy processing for different acoustic parameters and that speech features are processed differently. Lastly, the findings also concluded that the MMN was produced in a response to change, not repetition. Also, MMN responses were associated with non-primary areas of the auditory pathways. **Relevance to current work:** The current work was supported by the findings found in the above study, specifically that speech features are processed differently and that there is a hierarchical processing of speech processing. Also, the current work used MMN responses as an event-related potential measure; therefore, it is significant to know that MMN responses are associated with non-primary areas of auditory pathways. **Level of evidence:** N/A.


**Objective:** This study was designed for the purpose of determining whether MMN of speech stimuli can be used as a clinical measure for assessing central auditory function in normal adults and school-age children. **Methods:** Ten young adults (17-29 years) and 10 children (7-11 years) participated in the study. All individuals were healthy and had normal hearing. The speech stimuli used to elicit the MMN were synthesized speech (/da/ and /ga/) which had the second and third formants varied systematically from /da/ to /ga/. Participants were instructed to identify stimuli along a nine-item continuum in an oddball paradigm. Each participant had electrodes placed on their head to measure MMN. Placement of a supraorbital electrode was also used in order to detect eye movement. **Results:** MMN activity was found in all of the adult and children subjects when they were presented with deviant stimuli. In contrast, the MMN was absent when the same stimulus was presented alone. It is also important to note that MMN latency and amplitude measures were not significantly different between the adults and children.
Conclusions: The study concluded that MMN can be used as a clinical measure for assessing central auditory function in both adults and school-aged children. Also, the study showed that the MMN is an automatic response and can be elicited by speech stimuli. Relevance to current work: The research done in this study played an important role in the current study because it concluded that MMN is an appropriate and accurate electrophysiological response to examine in research studies; therefore, MMN was used in the current study to investigate processing differences between different stimuli. Level of evidence: Level IIIa.


Objective: The purpose of the chapter in this reference book is to discuss the aspects of the Signal Detection Theory (SDT) and how it can be used in experimentation. Conclusions: The SDT is often used to analyze data in situations where the role of decision processes needs to be made clear. The SDT provides a distinction between sensitivity and response bias and allows the comparison of data across tasks. In a yes-no design a participant in an experiment will be encouraged to always say “yes” when a signal is present and “no” when it is not. Although the participant only has two choices to respond, there are a total of four possible outcomes in any given trial. These four outcomes are a hit, miss, correct rejection, or false alarm. An important point to note is that the SDT assumes that the repeated presentations of a stimulus will give rise to an equal-variance Gaussian, or bell curve, distribution. Relevance to current work: The information above verified that the use of the SDT in the current work as a valid and an appropriate theory to use in the measurement and analysis of performance data. Level of evidence: N/A.


Objective: The objective of this study was to observe the cortical correlates of the perception of sustained vowels using magnetoencephalography (MEG). Methods: Subjects in the study were presented with three Finnish sustained vowels (/a/, /o/, and /u/) produced with the Semi-synthetic Speech Generation method. This method was used to create synthesized natural sounding stimuli. The stimuli were presented in pseudorandom order in separate sequences to the ten participants. Each participant was fitted with a 306-channel MEG measurement device in a magnetically shielded room. The N1m response, collected from the MEG, was studied by determining its amplitude, latency, and source location. Results: The N1m responses of all three vowels were equal in amplitudes. The equivalent current dipoles were located in the auditory cortex and the locations were more anterior in the right hemisphere than in the left. Furthermore, in the left hemisphere it was found that the centers of gravity of activation elicited by the vowels were located in three distinctly different areas. Conclusions: The investigation found that left-hemisphere specialization takes place in speech processing. In addition, different vowels were found to activate distinctly separate areas in the auditory cortex of the left hemisphere. The study found that vowels that were located close to each other in the formant space (/a/ and /o/) were located closer to each other in the auditory cortex than vowels that had a larger distance in the formant space (/u/ and /a/). It can be concluded from this information that categorical
perception of vowels might be explained by the mapping of the vowels’ formants in tonotopically organized cortical areas. **Relevance to current work:** This study demonstrated where localization of different vowels most likely takes place. It played an important role in the current study because the study observed where different vowels were processed in the left hemisphere. Also, the above study used MEG which provided spatial resolution; however, there is no temporal information. The current study furthered the research by observing the temporal processing differences in speech versus nonspeech stimuli. **Level of Evidence:** Level IIIa.


**Objective:** This article described receiver operating characteristic analysis and how they provide a quantitative evaluation of performance. **Conclusions:** Receiver operating characteristic (ROC) analysis provides the most comprehensive description of diagnostic accuracy because it estimates and reports all of the combinations of sensitivity and specificity of a test. An ROC analysis is able to describe all of the tradeoffs that a decision maker can achieve among true-positive, true-negative, false-positive, and false-negative decisions in any 2-group classification task. The graphical plot the ROC curve creates shows the performance on a 2-group classification tasks as its discrimination threshold is varied. **Relevance to current work:** Receiver operating characteristic curves play an important role in the analysis of the signal detection theory; as such, ROC curves played an important role in the analysis and demonstrating of data in the current work. **Level of evidence:** N/A.


**Objective:** This article explained how the ERP component, specifically the MMN, can be used in understanding auditory function and forms of its pathology. Also, the article discussed how MMN can be used as an accurate objective measure in research. **Conclusions:** MMN is elicited when an acoustically deviant stimulus replaces a standard stimulus. The deviant stimulus creates a difference way that is negative. This negativity created is generated by a change-discrimination process that occurs in the auditory cortex. The reason why MMN is a good measurement in research is that it is easy to elicit, provides an objective measure of discrimination ability, is elicited without attention, central auditory representations are involved in MMN generation, provides a representation of speech processing, and reflects auditory sensory memory. The usefulness and properties of MMN is supported by numerous amounts of studies that the author has cited in the article. The author also proposed that MMN is elicited by two intracranial generators, one in the auditory cortex and the other in frontal areas. This proposal is also in harmony with other studies that have discovered the same thing. **Relevance to current work:** The information obtained in this article provided a thorough definition of MMN. MMN played a major role in the collection of data in the current study. This article, along with the many other studies that are cited within it, supported the use and analysis of MMN in the current research. **Level of evidence:** N/A.
Brain Mapping of the MMN to Vowel Variances

Näätänen, R. (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology, 38*(1), 1-21. doi: 10.1111/1469-8986.3810001

**Objective:** This article reviewed the importance of MMN and its significant contribution to the understanding of speech sound processing. Suggestions and conclusions presented in this article are supported by research studies. **Conclusions:** MMN is an auditory ERP that does not require an individual to pay attention to a task. This ERP is a frontocentrally negative component that occurs when a deviant change in an auditory stimulus interrupts repetitive stimuli. The MMN provides researchers with information about the memory and cerebral areas involved in perceiving speech. Researchers have used the MMN to identify phonetic traces located in the left hemisphere. This finding has suggested that processing language is a left hemisphere dominant activity. Another property of MMN that helps researchers understand sound processing is its ability to demonstrate behavioral discrimination accuracy and when the auditory cortex makes that discrimination. In fact, an MMN wave can be present, even when a minor sound change occurs, and the individual might not even be consciously aware of a change. MMN can also be used in the discrimination of linguistic stimuli such as phonemes. It has been found that the MMN amplitude is larger with easier discrimination or improved discrimination ability. The article also gave insight to the effects of neurological impairment on the MMN. The way MMN is elicited can be affected by neurological impairment; however, every healthy adult with no neurological impairment can expect to have a normal MMN. **Relevance to current work:** This article played a vital role in the understanding of MMN in the current study. The MMN has several uses and is clearly a well-accepted and accurate objective measure used in many studies. Mismatch negativity measures were recorded for each individual in the study and analyzed to discover how the brain processes speech and sound deviations. Also, it is important to note that a normal MMN can be elicited in all adults with normal hearing and no neurological impairment; therefore, theses guidelines were followed in recruitment of participants in the current study. **Level of evidence:** N/A.


**Objective:** The article reviewed the purpose and uses of MMN as well as its potential use as an index of general brain plasticity. **Conclusions:** The MMN is an automatic electrophysiological brain response which discerns changes in auditory stimuli. When the change of a stimulus is near-threshold, then the MMN elicited is low in amplitude; in contrast, when magnitude of a stimulus increases than the amplitude of the MMN increases and the peak latency decreases. This behavior demonstrated by MMN allows researchers to use it as a measure of discrimination accuracy. MMN also gives insight regarding the memory-dependent response generated by auditory stimuli. The MMN occurs because of the existence of sensory-memory traces of preceding sound stimuli that exists at the moment of an occurrence of a deviant stimulus. It has been studied that discrimination of sounds can be taught or improved; for example, when an individual learns a new language they are taught to discriminate differences in sounds in that language. Upon being taught how to discriminate sounds, it has been found that an individual’s MMN will change and increase in amplitude because they are now able to discriminate sounds. Before being taught these discriminations, the MMN amplitude would be low or indiscernible
because the individual is unaware that the sounds are different. This phenomenon suggests that MMN can actually be used to observe brain plasticity. **Relevance to current work:** This article, along with many articles published by Näätänen, once again gave vital insight into the use of MMN, especially in research studies. The paper summarized how MMN can be used to demonstrate an individual’s ability to discriminate auditory sounds. The current study used MMN to observe subjects and their ability to discriminate between different speech stimuli. 

**Level of evidence:** N/A.


**Objective:** The objective of the study was to determine whether language-dependent memory traces activated by speech processing exist and where those traces are located in the brain.

**Methods:** There were two parts to this study. First, 13 normal-hearing Finnish participants (18-29 years of age, right-handed, four females and nine males) were presented with the phoneme /e/ as the standard stimulus. The standard stimulus was randomly replaced by /ö/ (a prototype in Finnish) or /õ/ (a prototype in Estonian) as a deviant stimulus. The next part included 11 normal-hearing Estonian participants (19-31 years of age, right-handed, five males and six females) who were put through the same task as the Finnish participants. An electroencephalogram was collected for each individual and the Fz electrode was observed for the MMN amplitude.

**Results:** In both groups of participants, the MMN amplitude increased as a function of the frequency deviance. In Finnish participants the deviant /ö/ (prototype in Finnish) elicited a larger MMN than the deviant /õ/ (non-prototype in Finnish) even though /ö/ deviated acoustically less from /e/ than did /õ/. This fact led the authors to conclude that MMN was enhanced in size for prototype deviants rather than just acoustic differences. In the Estonian participants, the MMN amplitude drop from /õ/ to /ö/ did not occur. This can be explained by the fact that /õ/ is a prototype in Estonian but not Finnish. In summary, MMN amplitude was enhanced when the deviant stimulus was a prototype as opposed to an equally complex non-prototype. Also, MMN responses were recorded as significantly larger in the left hemisphere than in the right when a deviant stimulus was presented. 

**Conclusions:** The results led the authors to conclude that neural traces of language-specific phoneme representations exist. Cortical memory traces of speech sounds were found and are suggested to serve as recognition patterns which allow the correct perception of phonemes in a given language. Lastly, it was concluded that the left-hemisphere is the dominating part of the brain for phoneme discrimination because neural traces were located there. 

**Relevance to current work:** The study found that memory traces involved in speech processing exist and are language-dependent and that these traces can be found predominantly in the left hemisphere. Localization of MMN was expected to be discovered in the left hemisphere of individuals in the current study. Also, this article supported many other findings that MMN can be used to accurately observe discrimination of speech and sound stimuli; therefore, MMN was used in the current experiment. 

**Level of evidence:** Level IIIa.
Objective: The purpose of this study is to examine cerebral organization using fMRI in three groups of individuals with different language experiences. The three groups consisted of the following individuals: (1) normally hearing, monolingual, native English speakers who do not know American Sign Language (ASL), (2) congenitally, genetically deaf individuals who use ASL as their first language and learned English later on in life with no auditory input, and (3) normally hearing, bilingual subject with both ASL and English as native languages. Methods: All subjects from the aforementioned groups were right-handed, healthy adults. Each group of individuals was scanned using fMRI while processing sentences in English and ASL. The English sentences were presented on a screen and the ASL sentences consisted of a film of a signer producing the sentences. The materials were presented in four different runs, two English runs and 2 ASL runs. At the end of each run, participants were asked yes/no recognition questions on the stimuli to ensure attention. Images were collected for both the left and right hemispheres and comparisons of the images were made across hemispheres and languages. Also, regions of activation were observed and evaluated. Results: When normally hearing subjects read English sentences, activation was observed in the left hemisphere in areas including Broca’s area, Wernicke’s area, and the angular gyrus. Weak and variable activation was seen in the right hemisphere. On the other hand, deaf subjects did not display left hemisphere dominance when reading English and instead displayed activation in the middle and posterior temporal-parietal structures in the right hemisphere. When the monolingual hearing individuals viewed the ASL they displayed no significant activation. However, when deaf subjects processed ASL there was a significant activation area in the left hemisphere within Broca’s and Wernicke’s area. Also, significant activation was identified in the right hemisphere. These results were also seen in the hearing individuals who also knew ASL as they viewed the ASL film. Conclusions: The study concluded that the processing of a person’s native language occurs predominantly in the left hemisphere. This suggests that there are strong biological constraints that render particular areas in the left hemisphere of the brain to be designed to process linguistic information. Relevance to current work: This study found that native language is predominantly processed in the left hemisphere. The stimuli that were used in the current study consisted of phonemes from the English language. Because English was being used, it was vital that all participants were native English speakers because the above study suggests that the brain processes secondary languages differently. Level of evidence: Level IIIa.


Objective: The purpose of this study was to evaluate the effects of sensorineural hearing loss on the detection, amplitude, and latency of cortical ERPs and the effects on associated behavioral measures when speech sound stimuli was used. Methods: There were 40 subjects in the study, 20 with normal hearing and 20 with sensorineural hearing impairments. The participants with sensorineural hearing impairments were broken down into severity groups including those with mild, moderate, and severe/profound losses. Consonant-vowel speech stimuli (/ba/ and /da/)
were used in this study. Stimuli were presented to the subjects via loudspeaker at both 70 and 85 dB HL. Each participant was fitted with eight gold cup electrodes in order to record EEG data. Behavioral measures of the stimuli included reaction time and percent correct discrimination. During the actual procedure participants were tested under active and passive conditions. In the active condition, subjects pressed a button when they heard a deviant stimulus. During the passive condition, subjects were asked to sit quietly, ignore the stimuli, and concentrate on reading a book of their choice. Results: After analysis of the data, it was discovered that prolongations in the ERP and behavioral latencies were seen in individuals with sensorineural hearing loss compared to those with normal hearing. Also, ERP amplitudes and behavioral discrimination scores were lower for the participants with hearing loss that exceeded 60 dB HL at 1000 to 200 Hz. It was also discovered that as the degree of hearing impairment increases, there is a significant decrease in the amplitude and increase in the latency of the ERP components in response to the speech stimuli. Conclusions: This study demonstrated that sensorineural hearing loss has a significant impact on behavioral measures and the timing, strength, and location of the cortical brain processes that play a role in the detection and discrimination of speech stimuli. Relevance to current work: The Oates et al. (2002) study revealed that hearing loss can significantly impact ERP measurements and other behavioral measures. Brain processes are impacted by hearing loss and prove to be deviant compared to those with normal hearing. For these reasons, the current study only included participants with normal hearing. Because of the importance of this, all participants were screened before collection of EEG to ensure that data was being collected on normal hearing adults. Level of evidence: Level IIIb.


Objective: The article described the definition of a receiver operator characteristic (ROC) curve and how it can be used in experimentation and data analysis. Conclusions: The ROC curve plots test sensitivity versus a 1-specificity or false positive rate. These curves help evaluate the quality of tests. In dealing with ROC curves, sensitivity and specificity are defined as the number of true positive decisions and the number of true negative decisions in a test. Sensitivity and specificity are the basic measures of performance on tests. The most commonly used ROC curve is the fitted, or smooth, ROC curve which assumes a binormal, or Gaussian, distribution. The smooth ROC curve uses a statistical method called maximum likelihood estimation. ROC curves can provide an accurate and reliable way of analysis in experiments. Relevance to current work: The current experiment incorporated the signal detection theory; therefore, the ROC curve played an important role in the analysis of the experimental data. Because ROC curves are frequently used in experiments and data analysis, it was determined that these curves provided a reliable way to illustrate data. Level of evidence: N/A.

**Objective:** The objective of this study was to determine the speech network and regions that are responsible for phonemic, syllabic, and supra-syllabic sequences. **Methods:** Eighteen native French speakers participated in the study. The participants laid inside a functional magnetic resonance imaging (fMRI) scanner while reading aloud bisyllabic pseudowords. The pseudowords involved four conditions. The first was the identical condition where the same pseudoword was repeated. The second condition was the recorded condition where syllables were rearranged while preserving the syllable consonant-vowel order. The third condition was the resyllabified condition where the words were constructed by resyllabification with the same set of phonemes. The last condition was the variable condition with six different phonetically unrelated pseudowords. A fMRI adaptation paradigm and analysis was used to determine the functional organization of motor and speech processes. **Results:** When participants read the pseudowords, activation was seen in the bilateral pre- and post-central gyri, the bilateral superior temporal gyrus, the medial premotor cortex, and the left insula. These areas are commonly activated in speech production. More specifically, phoneme-level processes were found in the left supplementary motor area, pallidum, posterior superior temporal gyrus, and superior lateral cerebellum. Syllable-level processes were found in the left ventral premotor cortex. Supra-syllabic processes were found in the right superior lateral cerebellum. **Conclusions:** The study concluded that specific areas are activated in relation to phonemic, syllabic, and supra-syllabic sequences. Based on the activation information, the authors of the study concluded that there is a separate representation in the supplementary motor area that activates for each specific phoneme in a sequence. In general, the study concluded that the medial and lateral premotor regions of the left hemisphere process phonemes and syllables. **Relevance to current work:** The study verified that speech processing predominantly takes place in the left hemisphere and that phonemic processing takes place predominantly in the left superior temporal gyrus. The current study came to a similar conclusion that the superior temporal gyrus plays a crucial role in the processing of phonemic information. In addition, activation of the superior temporal gyrus increases as complexity increases. **Level of evidence:** Level IIIa.


**Objective:** The purpose of this study was to determine to what extent natural speech sounds would evoke the MMN response. Also, the study examined whether deviant stimuli would evoke an increase in amplitude of the MMN based upon the spectral content of the stimuli. **Methods:** Twenty participants (10 females and 10 males; ages 18 to 26 years) participated in the study. The participants were native English speakers, right handed, had normal hearing, and had no history of neurological, cognitive, or learning impairments. All participants were fitted with a 32 channel Electrocap to record EEG data. The stimuli of the study consisted of two corner vowels, /æ/ and /u/. Participants listened to six different stimulus blocks presented in a two-interval forced choice paradigm. Each block contained a total of 200 stimuli, with 160 unfiltered and 40 filtered stimuli dispersed randomly throughout the block. Filtered stimuli consisted of the /æ/ and /u/ vowels with the first and second formants removed. Random deviant stimuli were
created by filtering random bands of spectral energy from the two vowels. These filtered spectral bands were not important to vowel identification. The participants’ responses, response times, and EEG were recorded. Results: Significant differences were found for reaction time between groups for stimulus type. The shortest mean reaction times for the /u/ and /æ/ stimulus pairs were seen in the unfiltered stimulus conditions. Brain maps were created which depicted the areas of greatest positivity and negativity during vowel processing. The grand average of all stimulus pairs showed greatest negativity in the left frontal area from 174 to 223 ms and from 298 ms to 547 ms. The greatest positivity was found in the left frontal region at 124 ms and from 248 ms to 273 ms. Dipole localization was also examined and averaged. The grand average file of all stimulus pairs yielded a dipole in the cingulate gyrus. Conclusions: The study concluded that MMN response does occur following the presentation of natural speech sounds. This evidence supports a formant-based theory of linguistic processing which is consistently found in other literature. The author concluded that the MMN responses reflect early speech processing and that formant information contributes to speech processing. It can also be concluded that the cingulate gyrus plays a significant role in language processing. Lastly, it was found that amplitudes of the MMN were greatest for pairs of auditory stimuli in which spectral energy not corresponding to formant frequencies was removed. Relevance to current work: The work cited above concluded that MMN responses are found following the presentation of natural speech sounds. MMN responses can be measured to further examine speech processing. The current study used MMN to examine the processing of natural and synthetic vowel stimuli. Perry’s (2012) study validated that MMN would provide relevant information for analysis. This work also provided suggestions for further research such as comparing natural and synthesized speech, which was the focus of the current work. Level of evidence: Level IIIa.


Objective: The objective of this study was to investigate the relation between the vowel phoneme intended by a speaker and the vowel phoneme identified by a listener. This relation was then used to investigate the acoustic measurements of the phoneme and its energy formant. Methods: Ten monosyllabic words were recorded by different speakers (76 total speakers). Each word began with /h/, ended with /d/, and had a different vowel in between. Ten different vowels were used. The words were randomized and played to different listeners (70 total listeners). A total of 1520 words were presented to the listeners across a series of eight sessions. The 1520 words were also analyzed using sound spectrograph. Results: The results indicated that the listeners more commonly agreed on the identification of the vowels /i/, /ɛ/, /æ/, and /u/. Listeners had the most difficulty identifying the vowel /ɑ/. When observers disagreed with the classification of a vowel, the two classifications were nearly always similar in either the frequency of the first or second formant. First and second formants were plotted for each vowel, which created a vowel loop graph. Vowels with similar first and second formants had overlapping vowel loops on the graph. The following vowels had overlapping vowel loops: /ɛ/ and /ɛ/, /ɛ/ and /ɔ/, /ɔ/ and /u/, and /ɑ/ and /o/. The results also indicated that children’s vowel formants are highest in frequency, followed by women’s, and men’s were the lowest in frequency. Conclusions: The authors concluded that sound spectrograph could be used to investigate the details of speech sounds. It was also found that these spectrographs are not random, but are instead indicative of a vowel sound. Lastly, the data showed that certain vowels are better understood than others. Relevance to current work: The Peterson and Barney study
provided information on how vowels are perceived by listeners. It also provided information regarding which vowels are more easily identified compared to others. The results from this study were used in the current study to determine which vowels should be used as stimuli and which vowels would be easiest for listeners to discriminate. \textit{Level of Evidence: Level IIIa.}


Objective: The purpose of this article was to describe the importance of auditory ERPs, to define the different types of ERPs, and to explain how ERPs are evoked. Conclusions: Event-related potentials are evoked by acoustic stimuli and can be measured by electrodes placed on the human scalp. There are several ways to classify ERPs including transient potentials, steady state potentials, or sustained potentials. Transient responses are evoked by rapid changes in stimuli. Steady state responses are evoked by regularly changing stimuli. Sustained potentials are evoked and recorded throughout the duration of a stimulus. One specific ERP response is the MMN. The MMN is a type of steady state response which is evoked when a deviant stimuli interrupts a repeated auditory stimuli sequence. When a change in a stimulus is heard by an individual, the MMN occurs approximately 150 ms afterwards. The MMN is frequently used in research because it is related to sensory memory and is evoked whether a person is attending to stimuli or not. \textit{Relevance to current work:} The current research recorded ERPs for each participant. Participants were presented with stimuli which were frequently interspersed with deviant stimuli. The ERPs, specifically the MMN, were used to measure when the brain detected a change in stimuli. The MMN response was used to determine the temporal processing of natural and synthetic vowel stimuli. \textit{Level of Evidence: N/A.}


Objective: The purpose of this study was to evaluate evoked responses to vowels as a function of pitch, phonetic category (formants), and hemisphere using evoked neuromagnetic responses to evaluate the M100 wave. Methods: There were six participants in the study. The participants listened to three synthetic vowels (/ɑ/, /i/, and /u/) at two fundamental frequencies (100 and 200 Hz). Vowels were created using a Klatt formant synthesizer. Vowel amplitude of voicing and fundamental frequency was altered between the vowels. High resolution volumetric magnetic resonance images were recorded for each participant using a 37-channel biomagentometer. Results: For three subjects, the M100 peak was larger in the left hemisphere than in the right across all stimuli. For one individual, the left hemisphere peak was larger for five out of the six stimuli. One subject (left handed) had larger response amplitudes in the right hemisphere across all stimuli. According to statistical analysis, the main effect of hemisphere and fundamental frequency did not have a significant effect on the M100 amplitude. Latencies of the M100 were related to vowel type (/ɑ/ versus /i/ versus /u/). Lastly, it was found that all subjects had a significant M100 latency difference located in the left hemisphere. Conclusions: The study agreed with former studies that the left hemisphere is specialized for speech and language processing as well as for fast temporal processing. Also, the authors found that vowels are processed independently of their pitch. \textit{Relevance to current work:} The current work attempted
to observe the differences in processing that occurs when altered vowel stimuli was presented to participants. The above mentioned study found that vowel type does provoke some noticeable processing differences and that these differences are found predominantly in the left hemisphere. The current study hypothesized that alternating natural and synthetic vowel stimuli would result in processing differences. **Level of evidence:** Level IIIa.


**Objective:** The purpose of this study was to investigate whether speech evoked potentials (SpEPs) contain information that can be used to distinguish different speech stimuli. **Methods:** Speech evoked potentials are auditory brainstem responses to speech stimuli. In this experiment, SpEPs were measured using three gold-plated Grass electrodes placed on the participants’ vertex and earlobes. There were 4 male participants in this experiment. Five synthetic vowel stimuli (/a/, e/, /æ/, /i/, and /u/) were generated using a formant synthesizer and presented to the volunteers in 6 trials. Only the first three formants of each vowel were used in the vowel stimuli. Speech evoked potentials were used to calculate the envelope frequency response (EFR) and the frequency following response (FFR). The EFR reflects auditory neural phase-locking to the envelopes of the speech stimuli. The FFR reflects auditory neural phase-locking to the formants of the speech stimuli. **Results:** The results indicated that the EFR response appears to discern vowels better than the FFR response. **Conclusions:** The study concluded that it was possible to successfully use SpEPs to collect information about vowel processing in the auditory system. Brainstem neural response in the region of F0 and F1 contain valuable information for discriminating vowels. In addition, the study concluded that using the EFR response appears to give more accurate information regarding vowel discrimination compared to the FFR. **Relevance to current work:** This study found that SpEPs can be used to discern vowels. The current work also used evoked potentials to investigate whether processing differences occurred when altered vowels were presented to individuals. The use of evoked potentials in the current study was supported by the Sadeghian et al. study. **Level of Evidence:** Level IIIa.


**Objective:** The purpose of this study was to determine whether separate cortical regions are responsible for syllabic and phonemic processing. **Methods:** Eleven native Mandarin speakers (between 19 to 21 years) participated in the study. The participants completed two phonological tasks while receiving an functional magnetic resonance image. In the first task, participants judged whether two Chinese characters were homophones. This first task required syllable level phonology only. In the second task, participants were asked to determine whether a pair of characters shared the same beginning consonant. This second task required both syllabic and phonemic processing. Analysis was completed to determine the cortical areas that were similar and different in the tasks. **Results:** Results of the study indicated that in the homophone judgment task, peak activation was located in the left middle frontal cortex. The left superior frontal gyrus and cingulated gyrus were also activated. During the initial-consonant task, the left middle frontal cortex was activated along with the left inferior frontal cortex and left medial frontal gyrus. **Conclusions:** The study concluded that phonemic processing of word initial
consonants was located in the left inferior frontal gyrus. This area may be responsible for important processes such as retrieval, manipulation, or selection of phonological representations. When participants were required to use syllabic level phonological information, the left middle frontal cortex was activated. This suggests that that this area is responsible for syllables rather than phonemes. In addition, the left inferior frontal cortex was also activated. This indicates that more complex linguistic content requires greater processing, and redundant, processing demands. 

*Relevance to current work:* This particular study gives important information regarding the difference in processing of simple versus complex stimuli. The current study supports this finding that more simplistic processing requires more simple and less redundant processing.

*Level of evidence:* Level IIIa.


*Objective:* The purpose of the investigation was to examine whether MMN is influenced by phonetic characteristics of a stimulus and if the MMN reflects an acoustic level of processing. In addition, the study also used the MMN to gain further understanding of categorical perception in speech processing. *Methods:* Ten normal hearing participants (between 19-35 years old, 8 female and 2 male) participated in the study. Stimuli consisted of a nine-item synthetically generated continuum of consonant-vowels (CVs) or nonspeech frequency glides. All CV stimuli were generated using the /a/ vowel. Stimulus one was the most /b/-like and stimulus nine was the most /d/-like. Frequency glide stimuli were generated using a sampling rate of 50 kHz. The starting frequency for each glide was identical to the F2 onset frequency for the formant transitions in the CV stimuli continuum. During the experiment, listeners were asked to identify stimuli in a same/different discrimination task by pressing a button. The participants also completed an oddball discrimination task in which they would press a button upon hearing a deviant stimulus. During all listening tasks, MMN and P300 responses were measured using a two-channel electrode configuration. *Results:* All participants identified the /ba/ to /da/ continuum in a way that reflected the presence of a categorical boundary. Participants exhibited the best discrimination to the nonspeech stimuli compared to the CV speech in the same/different task and the oddball discrimination task. The results of the ERP responses were similar to the results of the behavioral labeling and discrimination tasks. An MMN response was elicited by the nonspeech stimuli, but in the speech stimulus contrast the MMN was absent. Results of the P300 analysis showed that there was a larger amplitude and earlier P300 response elicited by the nonspeech stimuli. The P300 response evoked by the speech stimuli had a smaller amplitude and longer latency. *Conclusions:* The results indicated that nonspeech stimuli were processed differently than the CV speech, even though the acoustic content was analogous. Results also concluded that categorical perception occurs as early as the level of the MMN generators. In addition, the results provided evidence that a parallel processing of acoustic and phonetic information occurs at the level of the MMN generators. *Relevance to current work:* The study summarized provided a different perspective about the processing model of speech. Many of the studies reviewed believed there is a hierarchical model of speech processing; however, this study disagreed with that model in favor of a parallel speech processing model. These conflicting ideas was addressed and considered in the current work. *Level of evidence:* Level IIIa.
Objective: The purpose of this work was to test whether spectral changes in successive stimuli are detected differently in nonspeech and speech sounds. Methods: The stimuli for the experiment consisted of natural vowels (/i/, /e/, and /a/) and two-frequency complex tones. The stimuli were presented in an oddball paradigm in four blocks. There were 15 participants in the study who were Finnish-speaking, right-handed volunteers with normal hearing. Magnetoencephalographic signals were recorded for each participant using a 306-channel cap. Responses were recorded and equivalent current dipoles (ECD) were localized. Results: The results indicated that the strongest responses to the stimuli occurred about 100 ms after the stimulus onset in the bilateral auditory cortices. Mismatch field amplitudes reflected acoustic deviations for complex tones when spectral differences between the complex tones were significant. Mismatch field amplitudes for the natural vowels were similar despite different spectral compositions. Conclusions: The study concluded that greater frequency deviation enhances the MMN elicited by simple tones. Processing differences in response to the acoustic differences of the natural vowels were reflected by mismatch field latencies. The MMN amplitudes reflected acoustic differences in the nonspeech, complex tones. This finding possibly explains that vowels display categorical, rather than acoustic discrimination. Overall, the study concluded that changes in spectral composition of successive stimuli are encoded differently in the cortex for natural vowels and two-frequency complex tones. Relevance to current work: The study hypothesized that differences in sound stimuli (natural vs. synthetic vowels) will evoke a noticeable MMN. Also, alternating vowel stimuli was hypothesized to evoke an MMN which would demonstrate that processing differences between natural and synthetic stimuli occurred. The current study expanded this study by examining the temoral processing differences that took place in processing vowel stimuli. Level of Evidence: Level IIIa.


Objective: The purpose of this study was to evaluate the use of the easy Z-score imaging system (eZIS) in early diagnosis (within one year from onset) of neurodegenerative diseases including dementia, Parkinson disease, ataxia, and motor neuron disease. The eZIS allows computer-assisted statistical analysis of single-photon emission computed tomography (SPECT) images by examining the regional cerebral blood flow objectively (rCBF). Methods: Participants were selected from patients with a clinical diagnosis of neurodegenerative disease whose duration of illness was less than one year. There were a total of 172 participants with a variety of neurodegenerative diseases and 61 control participants. Brain SPECT was performed for each participant. Afterwards, SPECT images for all patients with each disease were anatomically standardized using eZIS. An averaged SPECT image was created for each disease using the anatomically standardized images. A Z-score map for this averaged SPECT image for each disease was obtained by comparison with SPECT images of age-matched healthy volunteers. Each patient was followed by a neurologist for more than 3 years to confirm clinical diagnoses.
were made using medical history, laboratory investigations, serial magnetic resonance images, and specific criteria. Results: The eZIS analysis found significant findings in the rCBF. In patients with Alzheimer disease, a significant rCBF decrease was found in the posterior cingulated cortex to the precuneus and the right parietal cortex. In patients with dementia with Lewy bodies, a significant rCBF decrease was found in the occipital cortex, the precuneus gyrus, and the posterior cingulated cortex. In patients with frontotemporal dementia, a significant rCBF decrease was found in the right lateral frontal gyrus to the right insula. In patients with Parkinson disease, increased rCBF was found in the bilateral basal ganglia, thalamus, and the dentrate nucleus of the cerebellum. In patients with vascular Parkinsonism, a significant rCBF decrease was found in the bilateral basal ganglia, the cingulated cortex, and the brainstem. In patients with cortical cerebellar atrophy, a slight rCBF decrease was found in the cerebellum, right hemisphere and the frontal cortex. In patients with multiple systemic atrophy of the cerebellar type, a significant rCBF decrease was found in the cerebellum and the pons. In patients with amyotrophic lateral sclerosis, a diffuse decrease in rCBF was found in the frontoparietal cortex and around the bilateral central sulcus. Conclusions: The study concluded that eZIS analysis of SPECT images are useful for early and differential diagnosis of neurodegenerative disease by identifying abnormal rCBF patterns. In addition, the study found that z-score maps can be used to identify more specific processing differences and t-score maps can be used to define whether the z-score differences are significant. Relevance to current work: This study supports the current study’s use of t-score and z-score maps to determine differences in processing. The current study used t-score and z-score maps to determine significant processing differences. Level of evidence: Level IIIa.


**Objective:** The purpose of this study was to examine whether different neural events occur in the left hemisphere during linguistic processing versus processing of nonlinguistic stimuli. **Methods:** Ten right-handed subjects were recruited for participation in the study. Neural responses evoked by speech signals were recorded using silver disk electrodes placed in scalp locations over the left and right hemispheres. The subjects completed two auditory identification tasks. The first task was called the Stop Consonant task and it required the participants to process acoustic parameter that were important for making linguistic distinctions. In the task, subjects listened to two stimuli (/ba/ or /da/) and were required to make a decision regarding which stimuli they heard. The second task was called the Fundamental Frequency task. This task required participants to process an acoustic parameter which provided no linguistic information at the phoneme level. In this task, the stimuli had identical linguistic information (low frequency /ba/ and a high frequency /ba/) and only differed in fundamental frequency. EEG as well as identification responses and reaction times were recorded for each subject. **Results:** The recorded reaction times did not differ significantly between tasks. The EEG analysis revealed that evoked potentials in the right hemisphere were identical in the Stop Consonant and Fundamental Frequency tasks. In contrast, there were significant differences in evoked potentials occurring in the left hemisphere in the Stop Consonant versus the Fundamental Frequency tasks. **Conclusions:** The study concluded that different neural events do occur in the left hemisphere depending on whether linguistic processing was occurring or processing of acoustic characteristics. In addition, the study concluded that the right hemisphere had identical neural events regardless of the processing task. The research results support the idea that a
unilateral neural mechanism is specialized to perform linguistic processing tasks in speech perception. **Relevance to current work:** Wood et al.’s study verifies that different areas of the brain have different responsibilities during tasks of linguistic processing. This discovery supports the hypothesis of the current work which is that processing of natural and synthetic stimuli will take place in different areas of the cortex because of the unique vowel characteristics. The above study also specifies that the current study will most likely find these processing differences in the left hemisphere because this is where speech processing predominantly takes place. **Level of evidence:** Level IIIa.


**Objective:** This document was created by the World Medical Association (WMA) as a statement of ethical principles that should be followed for medical research involving human subjects. The principles also should be followed in research involving identifiable human material and data. **Relevance to current work:** The current study was done in an ethical manner in harmony with the principles stated in the Declaration of Helsinki. Along with being in accordance with the Declaration of Helsinki, the current research was also conducted under the ethical principles upheld by Brigham Young University’s Institutional Review Board (IRB). In addition, the current study was approved by the Brigham Young University IRB. **Level of evidence:** N/A.


**Objective:** The purpose of the study was to estimate temporal integration and cortical activation, underlying cortical sensitivity, of speech periodicity as stimulus duration increased. **Methods:** Fourteen participants were fitted with a 306-channel neuromagnetometer used to record brain activity elicited by acoustic stimuli. The periodic structure and duration of the speech vowel stimuli were manipulated. The /a/ vowel was manipulated and a periodic and aperiodic version was created. The duration of the vowel stimuli was also varied in units corresponding to the vowel fundamental period. The vowel stimuli elicited N1m responses, which were measured and localized. ANOVA statistics were calculated for the N1m response to determine the amplitude, latency, and source localization of brain activity. **Results:** Differences in N1m amplitudes and latencies between the periodic and aperiodic conditions were dependent on stimulus duration. The N1m amplitude in the periodic condition was overall larger than in the aperiodic condition. In the periodic vowel condition only, the N1m latency decreased with increased stimuli duration. Equivalent current dipoles (ECD) of the N1m in the periodic condition were located more anterior in the cortex compared to the ECD localization of the aperiodic condition. The right-hemisphere responses for the periodic condition were also located more anterior than the responses of the aperiodic condition. In addition, the ECDs of the periodic condition were slightly medial compared to those obtained in the aperiodic condition. **Conclusions:** The results agreed with previous studies that cortical sensitivity to sound periodicity can be observed in the amplitude, localization, and latency of event-related responses. The results indicated that specific regions in the cortex respond to different characteristics of vowel stimuli such as vowel periodicity and to the vowel spectral envelopes. Therefore, the
study concluded that the cerebral processing rate is rapid enough to encode these periodicity and spectral characteristics found in communicative speech. Findings also suggested that speech sounds are processed in the brain by the auditory cortex and by a higher hierarchical chain of processing. Lastly, the study agreed with past investigations that the localization of the processing takes place in the temporal areas of the cortex. 

Relevance to current work: This study provided much insight to the processing of vowel stimuli within the auditory cortex. The findings of this study showed that slight differences in vowels can have an effect on the listener and the processing that occurs in the cortex. The current study aimed to expand the knowledge of vowel processing by investigating whether differences in the naturalness of vowel stimuli would have an effect on auditory processing. 

Level of Evidence: Level IIIa.


Objective: The objective of this study was to determine the neural networks and the relationship between neural networks underlying cognitive control. 

Methods: Fifty-nine healthy adults completed a reaction time, “go” and “stop” task. In the task, a dot appeared on the screen. After a randomized time interval (between 1 and 5 seconds), the dot turned into a circle or an “X.” The participants were instructed to press a button if a circle appeared on the screen (“go” trial) and to not press the button if it was an “X” (“stop” trial). Participants were scanned with an fMRI during the task. The collected data was analyzed using an independent component analysis algorithm. This algorithm was used to identify spatial and temporal brain regions.

Results: Trials were sorted into three categories which include the go success, stop success, and the stop error trials. These trials were compared and contrasted to find differences in cerebral activation. During the go success trials, activation was seen in the motor cortical network including the left/right precentral gryus, left/right post central gyrus, and left/right supplementary motor area. During the stop success trials, the right fronto-parietal network accounted for the most activation. Lastly, during the stop error trials, the midline cortico-subcortical network had the most activation. 

Conclusions: The study concluded that distinct network of brain regions were identified during the task and each region played a specific role processing. For example, the motor cortical network was responsible for motor preparation and execution. The right fronto-parietal network was responsible for attentional monitoring. The left fronto-parietal network was responsible for response inhibition. The midline cortico-subcortical network was responsible for error processing. The cuneus-precuneus network was responsible for behavioral engagement. Although these networks are distinct, they work together and are all involved in cognitive control. 

Relevance to current work: This particular study provided new insight to the components of cognitive processing and to the fact that reaction time can be used to determine cortical and cognitive processing. This insight provided support of the current study’s use of reaction time to estimate the amount of cognition required in language processing tasks.

Level of evidence: Level IIIa.
Appendix B

Informed Consent to Act as a Human Research Subject

Brain Mapping of the Mismatch Negativity Response to Vowel Variances in Natural and Synthetic Phonemes

David L. McPherson, Ph.D.
Communication Science and Disorders
Brigham Young University
(801) 422-6458

Name of Participant: ______________________________________

Purpose of Study
The purpose of the proposed research project is to study whether specific locations of brain activity are influenced by natural and synthetic speech sounds using vowel stimuli. Also, the research project attempts to investigate the temporal resolution differences between natural and synthetic speech processing.

Procedures
You have been asked to participate in this study by Lyndsy Smith, B.S., a student conducting research under the direction of Dr. David L. McPherson. The study will be conducted in room 111 of the John Taylor Building on the campus of Brigham Young University. The testing will consist of one session including orientation and testing and will last for no more than 3 hours. You may ask for a break at any time during testing. Basic hearing tests will be administered during the first half-hour of the session.

Surface electrodes (metal discs about the size of a dime) will be used to record electrical activity of your brain. These discs will be applied to the surface of the skin with a liquid and are easily removed with water. Blunt needles will be used as a part of this study to help apply the electrode liquid. They will never be used to puncture the skin.

Acoustic and linguistic processing will be measured using an electrode cap, which simply measure the electrical activity of your brain and does not emit electricity; no electrical impulses will be applied to the brain. These measurements of the electrical activity are similar to what is known as an “EEG” or brain wave testing. These measurements are of normal, continuous electrical activity naturally found in the brain.

You will wear the electrode cap while you listen to different speech vowel sounds, during which time the electrical activity of your brain will be recorded on a computer. The sounds will be presented through insert earphones at a comfortable, but not loud listening level. You will be seated comfortably in a sound treated testing room. You will be asked to give responses during the hearing test and portions of the electrophysiological recording by pressing a series of buttons.
The procedures used to record the electrophysiological responses of the brain are standardized and have been used without incident in many previous investigations. The combination of sounds presented is experimental, but the recording procedure is not.

**Risks/Discomforts**
There are very few potential risks from this procedure, and these risks are minimal. The risks of this study include possible allergic reactions to the liquid used in applying the electrodes. Allergic reactions to the liquid are extremely rare. There is also a possibility for an allergic reaction to the electrodes. If any of these reactions occur, a rash would appear.

Treatment would include removing the electrodes and liquid and exposing the site to air, resulting in removal of the irritation. If there is an allergic reaction, testing procedures would be discontinued. Another unlikely risk is a small abrasion on the scalp when the blunt needle is used to place electrode gel. Treatment would also include removing the electrode and gel, exposing the site to air and testing procedures would be discontinued.

**Benefits**
You will receive a copy of your hearing assessment at no charge. You will be notified if any indications of hearing loss are found in this area. The information obtained from this study may help to further the understanding of language processing, which will be beneficial to professionals involved in treating speech and hearing disorders.

**Confidentiality**
All information obtained from testing is confidential and is protected under the laws governing privacy. All identifying references will be removed and replaced by control numbers. Data collected in this study will be stored in a secured area accessible only to personnel associated with the study. Data will be reported in aggregate form without individual identifying information.

**Compensation**
You will be given $10 compensation at each session you attend for this study; you will receive this compensation whether or not you complete the study.

**Participation**
Participation in this research study is voluntary. You have the right to withdraw at any time or refuse to participate entirely without affecting your standing with the University.

**Questions about the Research**
If there are any further questions or concerns regarding this study, you may ask the investigator or contact David McPherson, Ph.D, Communication Science and Disorders, at (801) 422-6458; Taylor Building Room 129, Brigham Young University, Provo, Utah 84602; e-mail: david_mcpherson@byu.edu.

**Questions about your Rights as a Research Participant**
If you have questions regarding your rights as a research participant, you may contact the BYU
Other Considerations

There are no charges incurred by you for participation in this study. There is no treatment or intervention involved in this study.

The procedures listed above have been explained to me by: _____________________________
in a satisfactory manner and any questions relation to such risks have been answered.

I understand what is involved in participating in this research study. My questions have been answered and I have been offered a copy of this form for my records. I understand that I may withdraw from participating at any time. I agree to participate in this study.

Printed Name:__________________________

Signature:_____________________________

Date:________________________________