Neuronal Adaptation And Formant Transition Direction In Vowels: An MMN Study

by

Nathanael A. Crawford

Submitted in partial fulfilment of the requirements
for the degree of Master of Science

at

Dalhousie University
Halifax, Nova Scotia
March 2014

DEDICATION PAGE

To my wife, Sarah – For your unfailing love and support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

While there are electrophysiological techniques that are currently used clinically to assess sound encoding without the need for behavioural feedback (e.g., ABR), they give no information about the capacity to discriminate speech sounds (Aiken & Picton, 2008). This information can be captured via event-related electroencephalographic responses (ERPs), using the Mismatch Negativity (MMN), a long-latency waveform produced by the auditory cortex in response to a sound pattern break (Martin, Tremblay & Korczak, 2008). The MMN has been shown to assess neural plasticity and recovery in brain-damaged adults (Kujala & Näätänen, 2010) and may be used clinically as a method to monitor the progress of language therapy or validate a hearing aid fitting, but more research is needed before this technique will be clinically viable.

Examined was whether the MMN varied predictably in response to changes in the direction of frequency-modulated tone glides and equivalent second formant transitions in vowels (e.g., /ɪ/ as in "bit" and /e/ as in "bate"). A novel stimulus presentation paradigm was designed to distinguish the MMN from the N1 component. 10 normal-hearing adults with no neurological diseases were recruited and presented stimuli via insert earphones while they watched a silent, subtitled movie. ERPs were recorded from 128 scalp electrodes. The MMN was successfully distinguished from the N1, marking participants' ability to discriminate vowel stimuli only. A significant early P2 component, which decreased in size with successive stimulus presentations, was also elicited for vowels only and is believed to reflect formant encoding. Discrepancies between vowel and tone results are discussed along with clinical implications and contributions to the fields of ERP and vowel research.

Keywords: adaptation, frequency modulation, N1, mismatch negativity, P2, speech perception, vowel-inherent spectral change

LIST OF ABBREVIATIONS USED

A1	Primary auditory cortex
ABR	Auditory brainstem response
AEP	Auditory-evoked potential
AM	Amplitude modulation/modulated
ASSR	Auditory steady-state response
BOLD	Blood-oxygen level dependent
CAPD	Central auditory processing disorder
CN VIII	Eighth cranial nerve
CNS	Central nervous system
CV	Consonant-vowel
Cz, Fz etc.	Central scalp electrode, fronto-central scalp electrode etc.
dB	Decibels
EEG	Electroencephalogram
ERF	Event-related field
ERP	Event-related potential
f_0	Fundamental frequency
f_1, f_2, etc.	First harmonic, second harmonic etc.
F_1, F_2… etc.	First formant, second formant, etc.
FFR	Frequency-following response
FM	Frequency modulation/modulated
fMRI	Functional magnetic resonance imaging
FWE	Family-wise error
HG	Heschl's gyrus
hVd	/h/-vowel-/d/
Hz	Hertz
IC	Inferior colliculus
ICA	Independent components analysis
ISI	Inter-stimulus interval
L1	First-language
L2	Second-language
LH	Left cerebral hemisphere
M	Mean
MEG	Magnetoencephalogram
MMF	Mismatch field
MMN	Mismatch negativity
MRI	Magnetic resonance imaging
nAmp	Nanoamperes
MOT	Multiple object tracking
ms	Milliseconds
N1, N2 etc.	First negative ERP, second negative ERP, etc.
N100 m	ERF negativity 100 ms post-stimulus onset
N100, N400 etc.	ERP negativity 100 ms post-stimulus onset, 400 ms post onset, etc.

P1-N1-P2	Complex composed of first positive, first negative, and second positive ERPs
P1, P2 etc.	First positive ERP, second positive ERP, etc.
P300, P600 etc.	ERP positivity 300 ms post-stimulus onset, 600 ms post onset, etc.
PCA	Principal components analysis
PET	Positron emission tomography
PSP	Post-synaptic potential
PT	Planum temporale
RH	Right cerebral hemisphere
RMS	Root-mean square
s	Seconds
SD	Standard deviation
SHCD	School of Human Communication Disorders
SLI	Specific Language Impairment
SNR	Signal-to-noise ratio
SOA	Stimulus-onset asynchrony
SPL	Sound-pressure level
SQUID	Superconductive quantum interference device
STG	Superior temporal gyrus
STS	Superior temporal sulcus
uV/μV	Microvolts
VC	Vowel-consonant
VISC	Vowel-inherent spectral change

Acknowledgements

First, I would like to convey my sincere gratitude to my supervisor, Dr. Aiken, for his guidance in all aspects of my thesis. I would like to thank him for taking me on as a student, spending the time to teach me new techniques and encouraging me to think creatively and apply my new knowledge and skill independently. Our discussions were always friendly, stimulating and positive.

Second, I would like to thank my committee. Thank you to Dr. Kiefte for helping with stimulus creation and sharing his knowledge on speech science. Thank you to Drs. Newman and Petley for sharing their expertise on electrophysiological techniques, stimulus presentation, data collection and analysis. Thank you to Dr. Billings for agreeing to be my external examiner and providing insight.

Third, I would like to thank several of my classmates at the School of Human Communication Disorders for their help, many of whom took the time to participate in my study. Thank you to fellow thesis-track student Caroline Jamison, in particular, for helping with participant preparation and setup for EEG data collection.

Fourth, I would like to thank my family for their patience and support in my academic endeavours. Thank you to my parents, Dale and Carol, for giving my wife and me a place to live throughout my graduate degree and encouraging me to pursue my interests. Thank you to my grandparents Wendy and Ken Dickson and my brother David for their support and encouragement.

Last, but certainly not least, I would like to extend a special thank you to my wife Sarah. Her efforts to remain patient, understanding and genuinely empathetic, even as she completed her own (post)graduate work, have been my mainstay for the past three years. I could not have done this without her.

CHAPTER 1 INTRODUCTION

When two people fail to communicate successfully with one another, there are numerous points at which the communication process could have broken down, from the inability of the speaker to produce the desired acoustic signal, to the inability of the listener to process the meaning of the message. In a medical setting, it is helpful to know at which precise point(s) patients experience such communication breakdowns when they occur, for a variety of reasons. For example, such knowledge may facilitate treatment planning for people who experience communication difficulties following neurological damage (Kujala & Näätänen, 2010). There are well-established behavioural tests for determining whether one's difficulties are due to peripheral hearing loss (e.g., damage to the cochlea) or poor understanding of language at a higher level. However, problems with speech perception are complex: speech perception may break down in someone with normal hearing sensitivity for a number of reasons, which are not easily distinguished via behavioural tests (Feng, Yin, Kiefte, & Wang, 2010). First, there may be problems with encoding the raw acoustic properties of speech sounds in the auditory cortex, despite normal peripheral sensitivity. Second, while these properties may be encoded properly, there may be problems with "mapping" them onto linguistically relevant (i.e., phonemic) representations, which are used to discriminate speech sounds (e.g., discriminating /ɪ/ as in 'bit' from /e/ as in 'bate'). Both encoding and "mapping" problems would result in similarly poor performance on speech discrimination tests, but it is important to be able to distinguish these problems from each other, since each might warrant different treatment. For instance, difficulties with encoding might be ameliorated by intensive auditory training (Tremblay, Shahin, Picton, & Ross, 2009; Wilson, Arnott, & Henning, 2013), whereas discrimination-specific difficulties with phonemic mapping would likely be better addressed by language-based training (Medwetsky, 2011; Näätänen et al., 1997).

Thesis description

The aim of this thesis was to investigate the use of electrophysiology as a window into the neural processing involved in two important aspects of speech-perception: the encoding and discrimination of rapid formant transitions. To this end, this study examined differences in cortical event-related potentials (ERPs) to rapid formant

transitions and similar frequency-modulated tonal sweeps. It was hypothesized that the different ERP's that reflect these two processes could be reliably distinguished, and would be similar for both formant transitions (vowels) and tones. We did this by rapidly presenting frequency-modulated tone glides and vowels with equivalent second-formant (F_2) glides to participants while their ERPs were recorded as they watched a silent, subtitled movie. The stimuli were presented using a novel presentation paradigm that allowed us to mathematically extract the ERP's associated with each process separately and compare them across stimuli and conditions.

Motivations and goals

This work is important for several reasons:

Motivation 1: First, it has been difficult to reliably isolate the ERPs associated with sound encoding (i.e., the N1 component) and discrimination (i.e., the mismatch negativity) in both vowels and tones. Being able to make this distinction between these ERPs may help to clarify the underlying neural mechanisms involved. This is crucial to interpreting what these ERPs tell us about speech perception. One motivation for this study was to better understand these ERPs by carefully controlling how we elicit them.

Goal 1: We created a novel stimulus presentation paradigm to separate these ERPs from one another. The goal was to test the appropriateness of this paradigm to ultimately add to the basic principles of ERP recording and analysis.

Motivation 2: Second, the link between tone (auditory) perception and speech perception in the brain is beginning to receive more attention, but it is still poorly understood and requires further study. Many studies pay close attention to the research principles and findings of one field or the other, but few consider them in equal depth. Understanding how the most basic sound perception processes in the brain are related to the perception of more complex speech sounds may give us insight into the development of various communication disorders, such as dyslexia, specific-language impairment (SLI) and central auditory processing disorder (CAPD).

Goal 2: This research aimed to bring together the two very important fields of general auditory perception and speech perception by measuring and comparing ERP responses to equivalent frequency-modulated tone and synthetic vowel stimuli. Research principles from both fields were carefully employed in paradigm design and stimulus

creation to determine whether or not there are similarities in neural populations that process tonal sweeps and formants that rise or fall within the same frequency range.

Motivation 3: Third, being able to measure the ability to encode and discriminate speech sounds without the need for behavioural feedback could have a wide variety of significant research and clinical applications. There are clinical tools already in use that assess neural processing at lower levels (e.g., Auditory Brainstem Response), but more basic research needs to be done to improve our ability to record and interpret the ERPs associated with the encoding and discrimination of speech sounds to be clinically viable.

Goal 3: This research aims to lay the groundwork for the creation of a clinical tool that can be used to assess the capacity of individuals to discriminate speech sounds without the need for them to respond behaviourally to those sounds. Such a tool would have a variety of applications such as helping distinguish auditory processing difficulties from cognitive and receptive language impairments that may affect traditional speech tests (see Humes et al., 2012, for a review). It could also help in testing speech perception in people who cannot provide reliable behavioural responses (e.g., individuals with severe motor-speech impairments or dementia).

Contributions

There are several contributions the results of our study make to the literature on electrophysiology and speech perception. Our study results show that:

1) Our stimulus presentation paradigm can be used to isolate the MMN from the N1 so that the MMN is a pure measure of participants' ability to discriminate formant transitions.
2) Our paradigm also can be used to distinguish P2 adaptation as a measure of the neural encoding of formant transitions in synthetic vowels.
3) Our paradigm is a tool that may be used in further electrophysiological studies of speech perception. Such research may optimize it for specific research questions (e.g., which properties of vowels are neural populations sensitive to) and for use in the clinical setting to test speech perception in "hard-to-test" clinical populations.

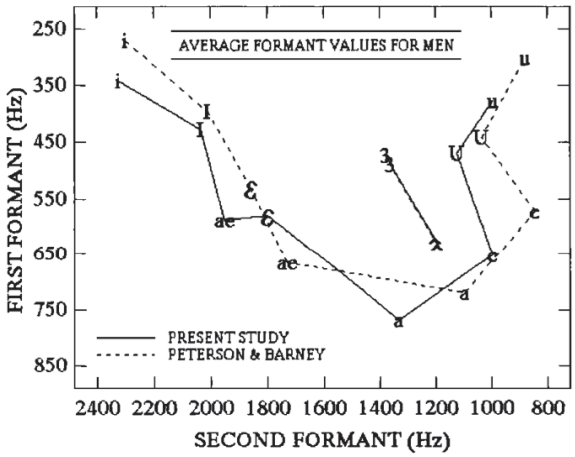
CHAPTER 2 BACKGROUND

Behavioural studies of speech perception

Formants and speech perception

For the brain to be able to perceive speech, it needs to be able to encode the acoustic properties of the speech stimulus. Voiced speech sounds have two key properties: fundamental frequency, denoted f_0, and formants (Klatt & Klatt, 1990). F_0 reflects a variety of factors, such as the rate at which the vocal folds vibrate and the length of the vocal tract of the speaker, which differs across people and genders (Barreda & Nearey, 2012). Speakers vary f_0 to convey emotional (e.g., mood) and some linguistic (e.g., making a statement or asking a question) information in English (Altmann & Gaese, 2013). There is also energy at integer multiples of the fundamental frequency (i.e., harmonics), denoted f_n, where 'n' is the integer multiple. Speakers move their articulators—the tongue, jaw, and lips—to change peaks in resonant sound energy to make different speech sounds. These peaks are called formants, denoted F_n, where 'n' is the formant number. For example, when switching from the high, front vowel /i/, to the low, back vowel /ɑ/, F_1 increases in frequency due to tongue lowering, and F_2 decreases in frequency due to tongue retraction. The most important formants for vowel discrimination are the lowest two, F_1 and F_2, with F_2 being most important (Hillenbrand, Getty, Clark, & Wheeler, 1995; Jin & Liu, 2013). All vowels can be plotted on a F_1 x F_2 quadrilateral (Hillenbrand et al., 1995; Peterson & Barney, 1952 - see Figure 1). Discrimination of consonants operates in a similar fashion (Kewley-Port, 1982; Korczak & Stapells, 2010).

Figure 1 Average formant F_1 and F_2 frequencies for adult male speakers (Hillenbrand et al., 1995, p. 3104; Petersen & Barney, 1952).



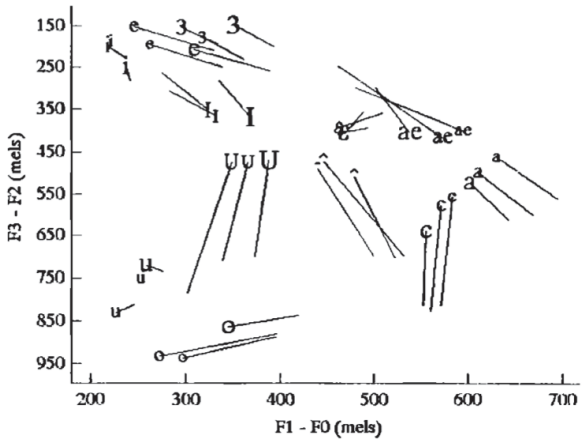
Vowel-inherent spectral change (VISC)

There can be a great deal of variability in the F_1 x F_2 vowel quadrilaterals, or "vowel spaces," of individual speakers, based on a variety of factors. Some speakers have smaller vowel spaces than others, independent of gender (Hillenbrand et al., 1995; Neel, 2008). The size of one's vowel space may affect his or her intelligibility in some cases, with decreased vowel space size being noted in people with motor-speech impairments (Kim, Hasegawa-Johnson, & Perlman, 2011; Neel, 2008). Conversely, studies have shown that mothers exaggerate their vowel spaces when speaking to their babies to facilitate speech-sound distinction as they learn language (Kuhl, 2004). Thus, the greater the "acoustic distance" between vowels in the vowel space (i.e., as characterized by the difference in their onset frequencies), the larger the formant transitions between them are, and the easier they are to distinguish from each other.

Not only are formant differences between vowels integral to determining vowel identity, so too are formant changes within the vowels themselves. This later phenomenon is known as vowel inherent spectral change (VISC) and has been increasingly studied over the past 20 years. In 1995, Hillenbrand and colleagues recreated a classic study on vowel perception by Peterson & Barney (1952), where acoustic measurements were taken of reliably identified English vowels produced in a /h/-vowel-/d/ (hVd) context. While Peterson & Barney (1952) only measured the steady-

state portions of vowels at one point in time, Hillenbrand and colleagues (1995) measured formant (i.e., spectral) changes at 20% and 80% vowel duration. Their results showed that some vowels (i.e., monophthongs) had significant changes in F_2 frequency between the two time points, some had changes in F_1, while others had both changes in F_1 and F_2 (see Figure 2).

Figure 2 Average formant F_1 and F_2 frequencies variation for male speakers; phonetic symbol represents 80% vowel duration, while line points to 20% vowel duration (Hillenbrand et al., 1995, p. 3105).



Hillenbrand and colleagues (1995) examined this further by adding and removing a variety of factors that may be important to vowel discrimination, asking participants to indicate which vowels they heard by choosing one of 12 alternatives. They found that beyond the steady-state, single-sample portions of F_1 and F_2 in vowels and vowel duration, the most important factor in vowel classification was within-vowel spectral change (i.e., VISC) over two time points (increasing classification accuracy by approximately 11%). By contrast, adding F_3 and or a third time-point did not appreciably improve classification accuracy. Furthermore, they showed that even vowel location on the F1 x F2 vowel space as determined by steady-state portions of some vowels do not predict accurate discrimination. For example, the vowels /æ/ and /ɛ/ are very close to each other in the vowel space, and plausibly more difficult to discriminate, but they were just as easily distinguished by listeners as other vowels that were further apart. Also, despite the fact that the vowels in Peterson and Barney's (1952) study had a wider F_1 x F_2 vowel space distribution, discriminability was just as accurate for Hillenbrand and colleagues' (1995) vowels. These results caused the authors to conclude, "…the vowels

of American English are more appropriately viewed not as points in phonetic space but rather as trajectories through phonetic space" (Hillenbrand et al., 1995, p. 3109). Thus, if speech perception relies more on dynamic trajectories, rather than static peaks of energy at certain sound harmonics, then it follows that smaller vowel space does not necessarily result in decreased intelligibility. This changed many assumptions long held about speech perception, and the authors called for more studies into "how listeners map spectral change onto perceived vowel quality," using synthesis methods (Hillenbrand et al., 1995, p. 3110).

It has been almost 20 years since the seminal work by Hillenbrand and colleagues (1995), and there is still debate over which specific acoustic aspects of VISC are most important to vowel discrimination. While, it is agreed that vowel onset is the most important acoustic feature, there are various theories that attempt to explain which aspects of VISC are most important to vowel identity after its onset. The "onset-offset" theory proposes that the frequencies of the formants at the end of a vowel's duration—the offset frequencies—in relation to their onset frequencies, are most important (i.e., at 80% vowel duration). The "onset-direction" theory proposes that the initial trajectories that the formants take, regardless of their offset frequencies, are most important. The "onset-slope" hypothesis proposes that the rate of spectral change in the formants is most important. The strongest support has been found for the "onset-offset" hypothesis (Chiddenton & Kiefte, 2013; Morrison & Nearey, 2007). For example, Morrison and Nearey (2007) tested all three theories, by presenting synthetically altered English vowels (monophthongs) in a "/b/-vowel-/p/-/ə/" (or "bVpa") context. Vowels were altered to have either straight formant transitions (onset-offset), "elbowed" transitions (onset-trajectory), or be shortened in duration (onset-slope). Participants classified the vowels presented in each context, and statistical models with parameters for each theory, systematically added and taken away, were used to explain their accuracy. The authors found that "onset-offset" parameters best explained the data.

Other factors have also been shown to be important in VISC, such as "spectral tilt" or the weighting of acoustic energy over time (e.g., ratio of hi- to low-frequency energy), the phonetic context of the vowels (e.g., "hVd" vs. "bVpa") and the amount of spectral change (Fox & Jacewicz, 2009; Kiefte & Kluender, 2008). Finally, not only do

differences in VISC have a direct impact on speech perception, VISC has also been shown to underlie differences in dialectical variations of English vowels in both native and non-native speakers (Fox & Jacewicz, 2009; Jacewicz & Fox, 2012; Jin & Liu, 2013). Thus, VISC is important to study because it is one of the most basic types of linguistically meaningful phonetic changes in speech perception.

Limitations of behavioural studies: Encoding of VISC in the brain

Though it is generally well understood that speech discrimination depends on formant changes over time, as in VISC, it is not well understood how formant changes are specifically encoded in the cerebral cortex. *Encoding* can be defined as the sensory registration of the raw acoustic properties of a stimulus in the central nervous system (CNS), *before* they are used in discrimination. It is important to know how these changes are encoded because encoding difficulties may underlie speech discrimination difficulties in some cases. The relationship of VISC to speech perception has primarily been studied behaviourally, however (e.g., Fox & Jacewicz, 2009; Hillenbrand & Nearey, 1999; Kiefte & Kluender, 2008), which makes the assessment of speech sound encoding difficult. Specifically, these behavioural measures do not always isolate the process of speech-sound encoding from discrimination, nor do they always accurately reflect the underlying speech perception processes occurring in the brain. For example, a person may score poorly on a behavioural speech-sound discrimination test because the stimuli cannot be properly encoded by the auditory cortex, despite having good discrimination abilities. Alternatively, they may have excellent encoding and discrimination abilities, but may respond incorrectly due to motor limitations or psychological factors. The behavioural results give no insight into these distinctions, and limit the conclusions that can be drawn about the nature of participants' speech discrimination and encoding abilities (Wilson et al., 2013). More direct measures of brain activity involved in speech perception may, therefore, be helpful for studying formant encoding apart from discrimination with regards to speech-sounds.

Neuroimaging techniques used to study auditory and speech perception

Techniques with high spatial resolution: PET, fMRI

Both auditory and speech perception have been studied using a wide variety of neuroimaging techniques, which both directly and indirectly assess the neural activity of various parts of the brain. Positron emission tomography (PET) has been used to study the brain activity of normally functioning humans as they perform auditory and speech perception tasks (Poeppel et al., 2004). In PET, a radioactively labeled biological molecule is injected into the bloodstream travelling to the brain. Sensors are placed around the head, which measure the location of this molecule as it decays and moves throughout the bloodstream. When a specific neural population is active (e.g., when processing speech-sounds), there is increased blood flow to that region of the brain, and thus a higher concentration of radioactive molecules is seen in that area (Crivello & Mazoyer, 1999; Luck, 2005).

Similar to PET are magnetic resonance imaging (MRI) and functional magnetic resonance imaging (fMRI), in which a large magnetic field is used to cause the water molecules of the brain to spin a certain manner (Frahm, Fransson, & Krüger, 1999). Radiofrequency pulses are then used to excite these molecules, changing their magnetic gradients. Again, sensors are placed around the head to measure the recovery of these molecules to their original state and decay of these magnetic changes created in them over time. Different tissues in the brain (e.g., white matter and gray matter) have different magnetic properties, allowing them to be distinguished anatomically in MRI. FMRI uses the magnetic properties of blood vessels in the brain over time, to measure localized changes in blood oxygenation associated with neural activity (i.e., the blood-oxygen level dependent, or "BOLD," response, see Binder et al., 2000, below).

The advantage of these techniques is their spatial resolution in the brain. Where they lack, however, is in temporal resolution, with both the radioactive decay in PET and the BOLD response in fMRI being time-lagged (e.g., on the order of seconds), indirect measures of brain activity (Martinez-Montes, Valdes-Sosa, Miwakeichi, Goldman, & Cohen, 2004). They are too slow to accurately measure activity related to rapidly changing speech sounds (as seen in VISC). MRI and PET are also very costly, and the MRI scanner makes too much noise to be practical for many studies of the auditory system. Though these factors make them impractical to be used clinically so assess

speech-sound discrimination, they have been used to provide a wealth of information regarding auditory and speech processing in general (see below for some examples).

Techniques with high temporal resolution: Electrophysiology

There are a number of electrophysiological techniques that measure brain activity in a more direct manner, which provide a high degree of temporal resolution (e.g., on the order of milliseconds; ms). For example, the activity of individual neurons can be measured directly in the brain via implanted electrodes. Individual neurons are sensitive to different stimuli, and "fire" maximally in response to said stimuli via action potentials. Whole arrays of electrodes have been implanted in large neural populations in the brain areas related to speech in epileptic patients pre-surgery (e.g., Bouchard, Mesgarani, Johnson, & Chang, 2013; Liégeois-Chauvel, de Graaf, Laguitton, & Chauvel, 1999, as cited in Altmann & Gaese, 2013). These "near-field" studies are too invasive, however, to be ethically acceptable with participants who are not already candidates for neurosurgery.

There are less invasive techniques that can be used, however, which are based on the same electrophysiological principles as near-field studies. Rather than measuring the (direct) activity of single neurons intracranially, these can measure the activity of whole neural populations at the level of the scalp (Seubert & Herman, 2012). Neural populations in the auditory system are organized tonotopically from the cochlea (periphery), where sound wave encoding begins to the primary auditory cortex (A1, Skoe & Kraus, 2010). There are a number of structures between the cochlea and A1, where neural populations perform various sound encoding operations, such as the eighth cranial nerve (CN VII), superior olivary complex, cochlear nuclei, lateral lemniscus, inferior colliculus (IC) and medial geniculate body. Specifically, populations of neurons involved in the same process fire together, with their summed electrical discharge being measurable by electrodes at the scalp in the form of a continuous electrical waveform. Positive and negative deflections in this waveform in response to sound stimuli are called auditory evoked potentials (AEPs). Changes in the size (amplitude) and timing (latency) of these AEPs can be used to test the integrity of the various structures in the auditory pathway (Seubert & Herman, 2012). Although AEPs may not reflect direct neural

activity per se (i.e., action potentials), they reflect this activity "more directly" than the responses obtained via the aforementioned neuroimaging techniques.

Auditory brainstem response (ABR)

The Auditory Brainstem Response (ABR) captures "short-latency" AEPs, which arise between the cochlea and the brainstem (below A1), with the use of only a few electrodes. These short-latency responses are positive deflections that occur within the first 10 ms post-stimulus onset. There are 7 of them, with I (CN VII), III (cochlear nucleus) and V (IC), being the most clinically useful. Mid-latency responses occur between 10 and 60 ms after stimulus onset, and reflect activity of the medial geniculate body and the A1. They are associated with wakefulness under general anaesthesia (Seubert & Herman, 2012). It is also possible to measure other responses from the brainstem, such as the Frequency-Following Response (FFR), in which neural populations discharge in a time-locked manner to a stimulus presented repeatedly at a fast, periodic rate. This has been used to assess the peripheral hearing sensitivity of children, without the need of conscious attention (Aiken & Picton, 2006; Aiken & Picton, 2008).

Electroencephalography (EEG)

There is evidence that despite the tonotopic mapping of the whole auditory system, the encoding processes at A1 that are precursors to speech discrimination are quite distinct from those that occur at the cochlea (Herrmann, Henry, Scharinger, & Obleser, 2013). Whereas auditory nerve and brainstem activity is measured by ABR (see above), the activity of the A1 is measured via event-related potentials (ERPs). ERPs are transient positive and negative deflections in the signal that occur on the order of 50 ms – 1000 ms post-stimulus onset. Stimulus salience is inferred from the amplitude and latency of the ERP (Martin, Tremblay, & Korczak, 2008). ERPs occur in the frequency range of 0.1 Hz – 30 Hz, and are believed to arise from specific neural populations in the brain engaged in the same process. The current "best" model of ERP generation is the "equivalent current dipole," which these neural populations "form." Dipoles are created by post-synaptic potentials (PSPs) of many single neurons oriented in the same direction and firing synchronously in response to the same event. These PSPs are believed to sum together to create an electric field pattern that approximates that of ERPs observed by

electrodes at the scalp (Luck, 2005). The advantages of ERPs are their temporal precision, which is excellent for measuring responses to the fast changes in speech-sounds, and their relatively low financial cost.

There are several disadvantages when recording ERPs, however, which need to be overcome. First, when trying to localize the sources of ERPs, a large number of electrodes are necessary because the electrical signal conducted by the soft tissue of the brain diffuses rapidly when it meets the hard tissue of the scalp. Indeed, ERP recording lacks the spatial precision of PET and MRI, meaning that only the general underlying brain areas that activity comes from can be identified. To further complicate matters, there may be multiple dipoles and configurations of dipoles that contribute to any one ERP recorded at the scalp, complicating source localization (i.e., the "inverse problem"). There are methods of model fitting, however, whereby various dipole configurations (or more general regions of activity) are estimated, with the signal they would predict matched against the actual signal recorded (i.e., the "forward problem"). These are called independent components analysis (ICA) and principal components analysis (PCA). These methods only lead to "best estimates" of the dipoles underlying EEG activity, however, and are still subject to the "inverse problem" (Luck, 2005).

There are also a variety of other EEG sources that also need to be removed from the EEG waveform to isolate ERPs. These include general electrical noise from the brain, participant factors related to attentiveness (e.g., alpha waves from drowsiness), and muscle movement (e.g., eye-blinks). Activity from these sources must be filtered out using a variety of methods (e.g., narrowing the frequency range of EEG waves included, recording eye activity using separate electrodes and subtracting it, using PCA to remove activity from brain regions not of interest). Indicators of the onset of stimulus presentation or "triggers" must also be time-locked to the EEG waveform so that ERPs can be extracted from it (e.g., in -100ms pre-signal to 800 ms post-signal time-windows) and averaged together to remove this "noise" from the signal. This means that hundreds of stimulus presentations are often necessary to obtain adequate amounts of data for analysis (Luck, 2005). This often results in increased time commitments for participants, unfortunately.

Magnetoencephalography (MEG)

In addition to the electrical current produced by dipoles (e.g., the EEG signal), there is a concomitant magnetic field created which radiates out from the dipole. This field can be measured using magnetoencephalography (MEG) via magnetic sensors placed around the scalp (i.e., superconductive quantum interference devices, or SQUIDs, Diekmann, Erné, & Becker, 1999). The magnetic equivalents to ERPs are event-related fields (ERFs). Although EEG and MEG have similar temporal resolution, MEG has higher spatial resolution than EEG. Additionally, MEG is more sensitive to activity from sulcal sources, whereas EEG is more sensitive to activity from gyral sources. Like fMRI, MEG is very expensive, although work is being done to optimize it for use as a clinical tool (e.g., cortical language maps, D'Arcy et al., 2012). Also, work has been done to specifically examine the timing and localization of neural populations sensitive to amplitude and frequency changes in sounds (Altmann et al., 2011; Mäkelä, Hari, & Linnankivi, 1987).

In sum, though EEG and MEG provide the appropriate level of temporal precision for assessing speech-sound encoding and discrimination, more research needs to be done optimize these techniques for studying speech perception clinically (Altmann & Gaese, 2013; Kraus, McGee, Sharma, Carrell, & Nicol, 1992; Kujala & Näätänen, 2010).

ERP components used to study speech perception

There are several different ERPs that have been identified and studied in auditory and speech perception (Luck, 2005; Martin et al., 2008). For example, the P600 is a positive deflection in the EEG waveform that occurs 600 ms post-stimulus onset and is believed to reflect syntactic incongruity (i.e., improper sentence placement) of linguistic stimuli. The N400, conversely, is a negative deflection at 400 ms post-stimulus onset that reflects semantic violations. These ERPs reflect cognitive/language processes at a much higher level than speech-sound perception, however (Poeppel, Idsardi, & van Wassenhove, 2008). One ERP that could be useful to assess speech-sound perception at the discrimination level is the P300, which is a positive deflection that occurs 300 ms post stimulus. It reflects recognition of a stimulus that breaks a train of repeating stimuli (e.g., /ga/-/ga/-/ga/-/da/). This ERP, requires conscious attention on the part of the

listener to be elicited, however, and can be modulated by a number of factors, such as age and certain psychological conditions (Martin et al., 2008). Thus, to assess speech-perception at the discrimination and encoding levels, earlier ERPs, such as the N1 and the MMN, may be more useful. Martin and colleagues (2008) provide an excellent review of these ERPs.

The N1 component

First is the N1 component, which is a well-established index of the *encoding* of the acoustic properties of sounds in the human brain. Before the inception of ABR, this "vertex potential" was the primary electrophysiological marker used to assess hearing thresholds (Davis & Zerlin, 1966; Martin et al., 2008). Specifically, it is a negative deflection that occurs approximately 100 ms after the onset of a new stimulus, and is part of a string of ERPs called the P1-N1-P2 complex, P1 being the last positive mid-latency response (at about 50 ms). The N1 has at least three neural generators in the vicinity of the A1. The first is the N1b, which is a fronto-central scalp negativity generated by vertically oriented dipoles in both of the superior temporal lobes of the brain. Thus, it is measured best at electrode Cz (center of the scalp). The second component is the T-complex, which emanates radially from the superior temporal gyrus (STG), being negative at about 70-80 ms post stimulus-onset, positive at about 100 ms and negative at 140-160 ms, and is measured best by mid-temporal electrodes (Tonnquist-Uhlen, Ponton, Eggermont, Kwong, & Don, 2003). There is a third negativity at 100 ms of unknown origin, which is also best measured at Cz (Martin et al., 2008). It is seen when the interstimulus interval (ISI)—the time between the offset of one stimulus and the onset of another—is long (e.g., > 4s), whereas the N1b is seen at shorter ISIs. In general, N1 amplitude is positively correlated with stimulus intensity and ISI. It can be elicited passively, while participants are engaged in an unrelated task, although amplitude increases have been shown with conscious attention. Additionally, decreases in the size of the N1 or "habituation effects" have been shown to occur over long periods of stimulus presentation (e.g., 10 minutes or more), especially for less spectrally complex stimuli (Näätänen & Picton, 1987).

Similar to habituation effects in the N1 is the concept of adaptation, which also affects the amplitude of this ERP. Whereas habituation occurs over a long period of

time, across all stimuli, adaptation occurs on a short-term basis between stimuli. Adaptation refers to a decrease in the amplitude of the N1 with each successive repetition of an auditory stimulus (e.g., AAAAA). A specific subset of neurons in a neural population is activated, and as this same subset is stimulated repeatedly by the same sound, this subset becomes increasingly "adapted" or "refractory," firing more rhythmically and less strongly. The N1 amplitude typically reduces in to about half its original size after the third or fourth stimulus repetition, at which point the response stabilizes(Näätänen & Picton, 1987). As with senses other than hearing, which are dynamic and respond primarily to stimulus change, this is presumably due to a lack of novel sensory information (Kiefte & Kluender, 2008). When a new sound is heard however, it activates a slightly different, "non-refractory" subset of the same neural population, resulting in a larger N1 to the new stimulus (i.e., a "release from refractoriness"). The more different the new sound is, the larger the subset of non-refractory or "fresh" neurons are activated and the larger the release from refractoriness will be. Näätänen & Picton (1987) reviewed 3 studies (Butler, 1968; Näätänen et al., 1988; Picton, Campbell, Baribeau-Braun, & Proulx, 1978) that demonstrate this phenomenon well using tone stimuli that differed only in frequency. Interestingly, by gradually increasing the frequency difference between "adaptor" and "test" tones, they showed that, in addition to firing maximally to a specific frequency, auditory neurons also fire less strongly to tones of different frequencies the further away they are from the maximal frequency for a specific neuron/population.

Finally, the N1 needs to be elicited at least about 100 times in a given study for averaging purposes to increase the signal-to-noise ratio (SNR) enough to obtain a reliable response for data analysis (Martin et al., 2008).

In sum, the N1 signals the detection of a particular acoustic feature (e.g., the onset of periodicity), but no direct inferences can be made from it concerning one's capacity to discriminate sounds (May & Tiitinen, 2010).

The mismatch negativity (MMN)

Second is the Mismatch Negativity (MMN), which, in contrast to the N1, is an index of the *discrimination* of acoustic change in the auditory system. It is an increased negativity that typically occurs approximately 100-250 ms after a stimulus that deviates

from a regular series or pattern of sounds. Specifically, it may reflect discrimination at either the perceptual and/or sub-perceptual level and is not always identifiable when discrimination at either of these levels occurs. Traditionally, stimuli are presented in an "oddball" paradigm, in which several "standard" stimuli are presented followed by a "deviant" stimulus (e.g., AAAABA). While most studies typically use a stimulus deviance rate of about 10%, the MMN can be elicited with as few as 2 or 3 standard stimuli making up a pattern (Bendixen & Schröger, 2008; Jacobsen & Schröger, 2001). The increase in negativity is revealed in the deviant waveform by subtracting the waveform from the standard stimulus (or pattern) from the response from it. This subtraction process is supposed to remove the earlier N1 component from both waveforms (i.e., the encoding), leaving only the MMN (i.e., the discrimination response) for analysis, but there may still be a residual N1 negativity. This negativity is known as the "N1 effect," and is due to an increase that is typically seen in the N1 any time a new stimulus is presented. One way that the N1 effect can be teased out from the MMN is to ensure that the difference between standards and deviants is very small, which increases the latency of the MMN, making it more distinct from the N1 (Martin et al., 2008).

The MMN can be elicited in the absence of attention, although, its amplitude has been shown to become larger with attention in some cases (Alain & Woods, 1997). The apparent MMN amplitude increase with attention noted in some studies, however, may simply be residual amplification of the attention-sensitive N2b component, which follows the P1-N1-P2 complex and occurs at roughly the same latency as the MMN. Importantly, the N2b has been shown to be distinct from the MMN because, while the MMN inverts at the mastoids, the N2b does not (Alain & Woods, 1997; Sculthorpe, Collin, & Campbell, 2008). Furthermore, while the amplitude of the MMN has been shown to correlate well with behavioural, "attentive" measures of discrimination, it is believed to arise from a pre-attentive sensory memory mechanism from neural populations in the STG, HG and planum temporale (PT, May & Tiitinen, 2010; Tiitinen, May, Reinikainen, & Näätänen, 1994).

Studies have been carefully designed to minimize attentional effects on the size of the MMN waveform. For example interference of the N2b component on MMN can be minimized by directing participants' attention to visual stimuli, while their auditory

discrimination is being assessed (Näätänen, Paavilainen, Rinne, & Alho, 2007). Additionally, while MMNs elicited by the oddball paradigm have been shown to be more susceptible to attention effects, MMNs elicited using "repetition deviants" have been shown to be less affected by attention. Repetition deviants are elicited when a stimulus breaks an alternating pattern (e.g., ABABABBA), giving rise to a "pattern MMN." For example, Sculthorpe and colleagues (2008) had participants play a visual tracking game while listening to tones that differed in frequency, which were presented using this paradigm. They found that, when more attention was necessary to succeed at the game, this did not affect the size of the pattern MMN.

Finally, unlike the N1, the amplitude of the MMN increases in size with decreased ISI (Alain, Woods, & Ogawa, 1994; Martin et al., 2008). This means that another way to separate the N1 from the MMN is to use are relative fast ISI (i.e., < 1 s). That being said, the MMN is smaller in amplitude than the N1, meaning that it has a lower SNR and therefore requires more data (~200 sweeps) for averaging to reveal a reliable response for data analysis (Martin et al., 2008).

The relationship of the MMN to the N1

There are many issues involved in collecting MMN data making its clinical utility questionable. Firstly, there is considerable variability in the presentation of the MMN across individuals in terms of its amplitude and latency compared to the N1, which is much more consistent (Martin et al., 2008). Secondly, since it is not as stable an ERP as the N1 and has a smaller amplitude, the increased amounts of data required for averaging the MMN (see above) often make data collection too long to be clinically feasible. There have been special stimulus presentation sequences, however, that attempt to speed up data collection (e.g., Näätänen, Pakarinen, Rinne, & Takegata, 2004). Beyond these issues, however, there is one final issue regarding the relationship of the MMN to the N1. There is a debate as to whether or not the MMN and the N1 reflect the same or different processes in the brain (Garrido, Kilner, Stephan, & Friston, 2009; May & Tiitinen, 2010).

Two models of MMN production

Presently, there are two key models of MMN production, the Memory Trace Model and the Adaptation Model, which differ squarely on the proposed relationship of

the MMN to the N1 (for reviews, see Jacobsen & Schröger, 2001; May & Tiitinen, 2010). The Memory Trace Model considers the MMN to be the result of more advanced cognitive processes related to sensory pattern memory that is generated by a different neural population than the N1 (Näätänen, Jacobsen, & Winkler, 2005). In contrast, the Adaptation Model considers the MMN to be merely an N1 effect, being the product of a release from refractoriness, as discussed above (May & Tiitinen, 2010).

Due to these two possibilities, the encoding versus discrimination distinction is difficult to assess with ERPs, but further research using an appropriate paradigm should be able to elucidate this distinction. There is evidence that both the Memory Trace and Adaptation models are at least partially correct (i.e., that adaptation occurs and there is a separate discrimination response), and the implications of both models need to be considered in tandem. Unfortunately many MMN studies have not controlled for differences in the magnitude of the N1 response that would reflect a release from refractoriness, and likely show N1 changes as opposed to true MMN responses. Indeed, many studies of speech perception (see below) tend to assume that one of the two MMN models is correct (e.g., Memory Trace) along with its implications, without controlling for these differences. There have been a small number of studies, however, carefully designed to show that a true MMN can be measured (i.e., that cannot be explained by differences in the N1 reflecting adaptation in neural populations), thereby supporting the Memory Trace Model (e.g., Alain & Woods, 1997; Alain et al., 1994; Jacobsen & Schröger, 2001; Martin et al., 2008; Sculthorpe et al., 2008; Sussman & Gumenyuk, 2005). It would, therefore, be helpful to have more studies that carefully control their stimuli in similar fashion to elucidate differences in N1 Adaptation and the Memory Trace Models of MMN generation, specifically with regard to speech perception.

In sum, if the N1 and the MMN can be reliably distinguished from one another, without either component contaminating the other, each can be separately used to measure speech-sound encoding (N1 adaptation) and discrimination (MMN). This would mean that peoples' ability to discriminate sounds could be assessed independently of their ability to encode the raw spectral properties of said sounds (and vice versa) in the same study. Conclusions about their abilities could be made more reliably made, perhaps

ultimately leading to a better quality of treatment provided by healthcare and academic professionals.

Electrophysiological studies of speech perception

Encoding (N1) and discrimination (MMN) of speech-sounds

The N1 and the MMN have been used to study speech-sounds (e.g., consonants and vowels) in a variety of ways. There is evidence that the N1 operates in a similar fashion in response to vowels and tones (Näätänen & Picton, 1987). Specifically, both the N1 and the MMN are sensitive to f_0 (Aaltonen, Eerola, Lang, Uusipaikka, & Tuomainen, 1994; Näätänen & Picton, 1987) and changes in F_1 and F_2 (Aaltonen et al., 1994; Deguchi et al., 2010). This suggests that the N1 could be used to index the neural encoding of these formants. By measuring an N1 and MMN, it might be possible to separately assess the encoding and discrimination of these important speech features. Indeed, the MMN has been shown to be a valid indicator of the development of categorical perception (i.e., language-specific phoneme distinction) in first- and second-language (L1 and L2 respectively) learning in normally developing babies as young as 11 months (Cheour et al., 1998; Conboy & Kuhl, 2011) as well as adults (Näätänen et al., 1997; Winkler et al., 1999). It has also been used to demonstrate poor speech-sound discrimination abilities in children with learning problems (Kraus et al., 1996).

More specifically, both the N1 and the MMN have been shown to be sensitive to formant changes in both natural consonants and vowels (Korczak & Stapells, 2010; Martin, Kurtzberg, & Stapells, 1999). Martin and colleagues (1999) studied N1 and MMN responses to deviant CV syllables using the oddball paradigm, which varied from standards in place of articulation only (i.e., /da/ and /ba/). They showed systematic decreases in N1 amplitudes to deviants, as high-pass cutoffs for masking noise were systematically decreased in frequency. By contrast, MMN amplitudes showed sharp decreases when the masking noise made behavioural discrimination noticeably more difficult (at a high-frequency-masker cutoff of 1000 Hz). The authors suggested that the N1 reflected the amount of sound energy encoded by the brain (in an absolute sense), while the MMN reflected behavioural discriminability.

While there is a general consensus that the N1 distinctly represents the neural encoding of the raw spectral properties of vowels, there is a discrepancy in the literature regarding nature of the discriminability reflected in the MMN elicited by vowels. It is generally accepted that speech discrimination operates in a categorical fashion at higher-level processing (Kuhl, 2004) but when measuring ERPs (i.e., the MMN) related to lower-level processing, this may not necessarily be the case.

Speech-sound discrimination (MMN): Bottom-up effect

One explanation is that the MMN is sensitive to the size of absolute spectral changes between standards and deviants, operating in a "bottom-up" fashion. Specifically, formants in the F_1 x F_2 vowel quadrilateral (see Figure 1) by which vowels are encoded may map tonotopically onto the auditory cortex, with greater movement across this quadrilateral resulting in a larger MMN. That is, populations of neurons in the cortex may be sensitive to absolute frequency changes in vowel sounds as reflected by not only the N1 (Obleser, Elbert, Lahiri, & Eulitz, 2003) but also the MMN in a linear fashion. Indeed, Deguchi and colleagues (2010) found that both N1 and MMN responses to deviant French vowels, as well as behavioural results (i.e., response speed and discriminability) increased with increases acoustic distance from standards. They concluded that vowels were classified pre-attentively in the brain based on absolute position in the vowel space.

There are several studies that further support the "absolute" approach to MMN generation for vowels. For example, Korczak and Stapells (2010) studied three different types of phonemic contrasts, place of articulation (e.g., /ba/ and /da/), vowel space (e.g., /bi/ and /bu/) and voicing (e.g., /da/ and /ta/) and found that the MMN and N1 were both sensitive to changes in vowel space only. They argued that the changes in the MMN were due to differences in the steady-state information present in the vowels (i.e., absolute vowel space locations) between standards and deviants, rather than perceptual discriminability. Furthermore, other studies have shown that MMNs to deviants change in size based, not on their categorical status, but absolute amounts of spectral content (Maiste, Wiens, Hunt, Scherg, & Picton, 1995). Of particular note is Aaltonen and colleagues' (1994) study examining changes in the size of the MMN in response to synthetic Finnish vowels, which differed in either f_0 or F_2 to varying degrees and pure-

tone controls, which corresponded to the f_0 and F_2 changes. They noted increases in MMN amplitude as a function of increasing frequency difference between standard and deviant in all conditions (f_0 and F_2). Furthermore, MMN changes were smaller in vowels than pure-tones, with the MMN process being "saturated" at 40% deviance (greatest deviance). MMN changes to f_0 were also larger than those to vowel differences in general. The authors took an "absolute" approach, concluding that, since vowel and pure-tone stimuli, had similar results for f_0 (phonetically irrelevant) and F_2 (phonetically relevant), that general principles of auditory perception underlie phoneme discrimination.

There is an important caveat to the above findings, in that they fail to consider the importance of VISC in vowel discrimination (see above) when interpreting their results. Thus, a second "absolute" explanation is that it is possible that specific formant trajectories (VISC; see Figure 2), rather than more general acoustic distances, are encoded by specific populations of neurons arranged in a tonotopic map in the A1. Indeed, the above studies (e.g., Aaltonen et al., 1994; Deguchi et al., 2010; Korczak & Stapells, 2010) only focus on discrimination as it relates to differences in the static, steady state portions of vowels (e.g., the vowel /i/ in "see"). The VISC studies (see above), however, show that the non-steady state portions of vowels (i.e., in /i/ and /u/) are important sources of spectral information used in discrimination. Thus, while VISC certainly contributes to the changes in the N1 and MMN observed in the above speech-perception studies, it largely remains un-discussed. This means that, if the A1 encodes and discriminates vowels in an absolute sense, it is difficult to know if this is due to vowel space (i.e., vowel quadrilateral; see Figure 1), "formant-trajectory space" (see Figure 2) or a combination of both. This problem could be addressed by completing a similar study to those above, while explicitly manipulating VISC only in synthetic vowels used as standards and deviants.

Speech-sound discrimination (MMN): Top-down effect

An alternative explanation is that the MMN is indeed sensitive primarily to "top-down," phonemic effects (i.e., categorical perception), rather than absolute position on the vowel space or formant trajectory space. Here, stimuli are not perceived as reflecting absolute locations on tonotopic maps, but as the closest phonemically relevant vowel in a given language. For example, Martin and colleagues (1999; see above) found that the

absolute sound energy encoded by their vowel stimuli, reflected in linear decrements in the N1 amplitude, did not correlate in a one-to-one fashion with the amplitude of the MMN. While stable with high-pass frequency noise cutoffs above 1000 Hz, at 1000 Hz cutoffs and below, where F_2 is partially masked, standard and deviant stimuli could no longer be discriminated, and the amplitude and latency of the MMN showed marked changes. This suggests that there is something more than just the raw spectral properties of vowels that are integral to vowel discrimination, such as categorical perception.

Studies of the development of categorical perception in infants further support the "top-down" explanation of MMN elicitation in vowels. Specifically, as infants learn a language, categorical perception develops to facilitate processing of only the sounds that are meaningful to that language, causing them to eventually ignore non-phonemically relevant distinctions (Kuhl, 2004). For example, Cheour et al., (1998) found that the amplitude of the MMN reflected the absolute acoustic distance between standard and deviant vowels on the F_2 axis of the vowel quadrilateral in 6-month old babies. 6 months later, however, their MMNs had reorganized to reflect the development of "language-specific memory traces" for phonemes in their L1. This study is important because it provides a clear example of what MMNs that are generated by encoding differences should look like. Furthermore, given that MMNs changed with age to reflect L1 phoneme classification, adult MMNs should mainly reflect this too. Thus, if stimuli are perceived as vowels by adult participants in discrimination tasks, the MMN will not necessarily reflect absolute position in the vowel quadrilateral, but relative position for a given language. Interestingly, in the same way vowels can be plotted by absolute frequency in the vowel quadrilateral, multidimensional scaling can be used to plot vowels in listener's "perceptual space," which reflects perceptual boundaries based on phonemes in their L1 (Iverson et al., 2003).

Studies have also shown that the more prototypical a standard stimulus is of a vowel in one's L1, the harder it is to discriminate other vowels that are closer to it in the vowel space from one another, a phenomenon known as the "perceptual magnet effect" (Iverson & Kuhl, 2000; Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992). This phenomenon operates purely on a relative basis, rather than absolute one, in a "top-down" manner similar to categorical perception. This may suggest that the positioning of

vowels on the vowel space (the absolute explanation) may be more relevant in phoneme distinction (and hence MMN generation) the further away standards and deviants are from their vowel prototypes. This would suggest that both the "top-down" and "bottom-up" explanations, might both be, at least in part, true.

Interestingly, despite explaining their results via an absolute perspective, Aaltonen and colleagues (1994; see above) suggested that there might be different neural populations in the brain, which are sensitive to different vowel (phoneme) features (Aaltonen et al., 1994). This suggests that more advanced neural networks (e.g., the long-term memory networks involved in categorical perception) may be involved. Indeed, this sentiment is echoed by Iverson and colleagues who state, "The changes in perception due to language experience are almost certainly speech-specific, and thus can be considered, by definition, to be phonetic rather than purely auditory" (Iverson et al., 2003, p. B49). They suggested that language exposure can affect auditory processing, although it is unclear at what precise level. Indeed, the degree to which higher-level language learning (e.g., vocabulary development) affects the development of auditory and speech perception is important, but this is a complex and controversial issue, which is beyond the scope of this thesis (for relevant discussions, see Cacace & McFarland, 2005; Wallach, 2011; Werker & Yeung, 2005). That said, Iverson and colleagues (2003) did discuss the need for studies to examine the boundary between "late auditory" and "early phonetic" perception, where these effects may operate (e.g., Cheour et al., 1998), which could be assessed at a lower-language level (i.e., the phoneme level).

Contrary to the above findings, Lidji and colleagues' (2009) findings suggested that pitch (f_0) and vowel change are processed preattentively in the same part of the brain (Lidji, Jolicœur, Moreau, Kolinsky, & Peretz, 2009). Specifically, they found differences in the size of the MMN elicited by changes in pitch and vowel identity, to sung deviant vowels, but that these effects were not additive. This could be taken to suggest that basic auditory features such as pitch and complex speech features, such as formant transitions, are discriminated by the same neural populations. These results only imply that some overlap, however, occurs between the neural populations engaged in pitch and vowel-change discrimination. If some of the same neurons are engaged, along with many separate neurons, the responses will still not be additive. Thus, again, more careful

research needs to be done to determine whether there are specific neural populations involved in general auditory and speech-sound perception. This may be done by using the N1 and MMN to examine the specific neural populations involved in vowel encoding and discrimination (respectively) separately at the "late auditory" and "early phonetic" level of perception. Indeed, many of the above studies only assessed the MMN without quantifying the N1 in appropriate levels of detail for comparison (e.g., Aaltonen et al., 1994; Deguchi et al., 2010).

Contextual effects on vowel classification

From the discussion thus, far, it appears that peoples' ability to classify stimuli as specific vowels from the language being tested bears on whether or not discrimination (and the MMN) is more dependent on their absolute or relative spectral differences. That is, if stimuli are not classified as vowels (based on their similarity to their prototypes), their vowel space position is more important determining factor in their discriminability. Studies have shown that the context in which stimuli are presented plays a large role on how people classify vowels and whether they perceive sounds as speech or non-speech. Mann and colleagues (1980) found that the consonant in CV syllables they presented was perceived differently depending on the spectral characteristics of the consonant in the preceding VC pair (Mann, 1980). Holt and colleagues (2006) further found that pure-tone sine waves that mimic individual formant frequencies (e.g., F_2) of VC syllables or isolated vowels have the same effect on the perception of the following stimulus as natural speech sounds. Speech sounds can also affect whether following pure-tones are perceived as speech. Furthermore, all of these factors interact with each other in various ways to affect speech perception, when simultaneously manipulated (Holt, 2006). Kiefte & Kluender (2008) found that when either the general energy distribution (spectral tilt) or F_2 was held constant across sentences and following vowels (monophthongs), the feature not held constant was used to discriminate the vowel. This suggests that the auditory system could use any type of contextual energy change to aid in discrimination (not just formant peaks). It has also been shown that synthetic, steady state monophthongs of short duration are not as easily perceived as speech as more dynamic diphthongs (Kiefte & Kluender, 2005) although F_2 frequency is still used in both cases for discrimination (suggesting that the dynamic, VISC component was key; Kiefte & Kluender, 2008).

The above findings are significant because they suggest that MMN's to deviant stimuli perceived as vowels do not necessarily reflect absolute changes in the vowel quadrilateral in adults, but rather a relative type of change, as seen in categorical perception. To determine whether the MMN to vowels reflects absolute (i.e., due to raw spectral energy differences) or relative changes (i.e., due to top-down effects), responses to tones can be compared with those to synthetic vowels of short duration with equivalent spectral change between standards and deviants. Again, the N1 can be used as a measure of neural encoding (absolute effects) and the MMN a measure of discriminability (absolute and/or relative effects).

Speech perception: Memory trace (top-down) v. adaptation (bottom-up)

Finally, it remains unknown if speech discrimination as reflected by the MMN is more in line with a Memory Trace account of MMN generation or an Adaptation account. MMNs that reflect differences in vowel space would seem to be better explained by the Adaptation Model, which involves a purely "bottom-up" process. MMNs that reflect differences in categorical perception would seem to be better explained by the Memory Trace Model, which involves a more "top-down" process. That being said, it is conceivable that a memory trace could be created by absolute differences in vowel stimuli regardless of the presence of N1 release from refractoriness. If adaptation is controlled for, as well as the spectral complexity versus phonemic status of vowels, then distinctions between these two models can be made.

Limitations of electrophysiological speech perception studies

There are several methodological limitations to the above speech perception studies, based on the basic principles of N1 and MMN collection (see above), which further make the distinction between encoding and discrimination in their results difficult. First, all of the above studies used the oddball paradigm to elicit MMNs, which may be contaminated by the N1 (measure of encoding). Second, many of these studies (e.g., Aaltonen et al., 1994; Deguchi et al., 2010) did not collect the appropriate number of samples for deviant MMN's (< 100) as per Martin and colleagues' (2008) recommendations. Third, many of the N1s and MMNs in these studies were produced to deviant stimuli that vary over a variety of spectral parameters simultaneously, rather than

manipulating one parameter only. For example, Deguchi and colleagues (2010) manipulate F_1 and F_2 together and Maiste and colleagues use /ba/ versus /da/ stimuli, which also vary along a number of spectral parameters. To hone in on neural encoding via the vowel quadrilateral or formant-transition space, only one parameter per stimulus should be manipulated (i.e., one formant transition).

All of the above issues can be overcome by following strict electrophysiological procedures (mentioned above) while measuring an N1 to synthetic monophthongs in which only one of their formant transitions are different (e.g., the trajectory of F_2). If the different formant transitions are encoded by unique neural populations (even partially), they should elicit larger N1s (i.e., less adapted), when following a train of differing stimuli verses trains of repeated stimuli (i.e., using an adaptation paradigm; see above). Tone controls, which have similar spectral content to the vowels, could be used to determine whether the findings are consistent with general auditory perception principles, broader categorical perception principles, or both (Holt, 2006; Iverson et al., 2003). In this way, the recovery of the N1 could be used to assess the degree to which the two formant trajectories are encoded by unique neural populations. This could then be used to describe any changes observed in discrimination as reflected in the MMN.

Evidence for feature-sensitive neural populations in the cortex

In support of the "absolute" approaches, there is evidence in cortical regions other than the A1 for feature-sensitive neural populations, which are mapped similarly to the tonotopic frequency mapping of the A1. For example, Bouchard and colleagues (2013) did intracranial recordings, using high-density multi-electrode (i.e., near-field) arrays, of the sensorimotor cortices of epilepsy surgery patients as they articulated various consonant-vowel (CV) syllables. They found specific neural populations that were organized by the various phonetic features of the syllables produced. There is also evidence that several areas involved in speech production are also involved in speech perception, and are co-activated with speech perception areas (Poeppel et al., 2008). It would be logical; therefore, that there would be neural populations in the auditory cortex, sensitive to phonetic features, analogous to those in the sensorimotor cortex. Vowel or formant-trajectory space, as seen in VISC, may constitute such features. As mentioned

above, however, near-field studies are too invasive for most clinical populations, so it is more practical to assess these neural populations using ERPs (e.g., N1 and MMN).

Furthermore, with regards to the formant-trajectory space maps postulated for VISC, the presence of direction- or trajectory-sensitive neural populations seems to be ubiquitous across the sensory systems of the human brain. For example, there is strong evidence from studies of the visual system that the visual cortex has specific neural populations sensitive to specific grating orientations of incoming two-dimensional (spatial dimensions) visual stimuli (Shamma, 2001). These are conceptually similar to one-dimensional (temporal dimension) formant trajectories and it is suggested that the A1 processes auditory stimuli similarly to the visual cortex (Altmann & Gaese, 2013; Shamma, 2001; Stilp, Alexander, Kiefte, & Kluender, 2010). If this were a general type of neural organization in the brain, it would make sense that there are neural populations sensitive to the orientation of formant trajectories in speech-sound encoding. More research needs to be done in this domain, however, in the realm of general auditory perception, particularly with the use of pure-tones that are similar in makeup to formant transitions.

Auditory perception with FM tones

Behavioural studies

In a recent review of the literature on frequency modulation (FM), Altmann & Gaese (2013), provided evidence of specific neural populations in the A1 that are sensitive to continuous changes in the frequencies of pure-tones over time or "frequency modulations." There are three types of frequency modulation: linear and logarithmic (i.e., a tone with a sustained rise or fall) and sinusoidal (i.e., a sustained tone that fluctuates sinusoidally around a center frequency). Linear and logarithmic FMs are very similar in spectral makeup to single formant transitions and are thought to underlie speech-sound perception (Holt, 2006). Formant transitions are different, however, in that they reflect changes in the peaks of the spectral energy of vowels, meaning that, unlike FM tones, there is low-level energy present at frequencies other than the modulated spectral peaks.

FM has been shown to be important in communication for humans and animals (e.g., rats, bats, macaque monkeys), with speakers of highly tonal languages (e.g., Mandarin) showing more sensitivity to pure-tone FM. Indeed, people with dyslexia have been shown to have more difficulty with FM detection than normal controls (Stoodley, Hill, Stein, & Bishop, 2006) dyslexia being an impairment in written communication, which relies heavily on oral language abilities (e.g., phonological processing, Hämäläinen, Fosker, Szücs, & Goswami, 2011) and hence speech perception. Furthermore, the neural populations associated with FM processing appear to be sensitive to the "up" and "down" directions and to be central in location rather than peripheral. Shu and colleagues (1993) found that after a sustained, rising FM tone was presented to participants in one ear for several minutes, an unmodulated tone that was presented in the opposite ear was perceived as falling. This behavioural data could suggest neural adaptation effects (Shu, Swindale, & Cynader, 1993).

Neurophysiological studies

Indeed, ERP studies of the MMN and N1 point to several locations of neural populations sensitive to FM direction. Sams and colleagues found that MMF's elicited to directionally deviant pure-tone glides were source-located to the supratemporal auditory cortex, and were proportional to glide magnitude (Pardo & Sams, 1993; Sams & Näätänen, 1991). FMRI studies have shown bilateral activation in auditory brain areas (e.g., HG, STG) to FM. This activation has been shown to be stronger in the RH for slower FMs (e.g., processing spectral info, prosody in speech) with some lateralization effects (Brechmann, Baumgart, & Scheich, 2002; Hall et al., 2000; Schonwiesner, Rubsamen, & von Cramon, 2005). PET studies (Belin et al., 1998; Poeppel et al., 2004) have shown activation to be stronger in the LH for processing faster FMs (e.g., processing temporal information, quick formant transitions in speech-sounds). Furthermore, several studies have suggested that the various RH and LH neural populations responsive to FM directional changes are also responsive to changes in modulation speed (Hsieh, Fillmore, Rong, Hickok, & Saberi, 2012).

Finally, there is debate in the literature as to whether or not central auditory neural populations that encode FM are different than those that encode amplitude modulations

(AMs). FMRI evidence (Hart, Palmer, & Hall, 2003) suggests that the same auditory brain regions code FM and AM. There is also electrophysiological evidence, using the ASSR (Millman, Prendergast, Kitterick, Woods, & Green, 2010; Picton, Skinner, Champagne, Kellett, & Maiste, 1987) that AM and FM are encoded by separate neural populations in the brainstem, but the same populations in the A1. Conversely, other electrophysiological studies suggest that separate populations are sensitive to AM and FM beyond the cochlea (Herrmann et al., 2013; Mäkelä et al., 1987; Moore & Sek, 1995). For example, Mäkelä and colleagues (1987) found large reductions in the amplitude of the N100m to the second stimulus in identical AM or FM pairs. When the second stimulus was modulated in the opposite way (e.g., AM-FM or FM-AM), however, there was less of a reduction in the N100m, suggesting less neural adaptation (i.e., a release from refractoriness) because a new population was being stimulated. Furthermore, Hermann and colleagues (2013b) showed that when frequency selectivity decreases in the cochlea (widening of pass-bands with aging), structures beyond the cochlea (e.g., cochlear nucleus) compensate for this deficiency. Central frequency analysis (as measured by the spread of N1 adaptation in the A1, Herrmann, Henry, & Obleser, 2013) therefore, still occurs normally. They also showed that AM variations become less important the farther up the ascending auditory pathway they travel, especially in fast/spectrally varying stimulus presentations. Also, the degree of N1 adaptation decreased with increasing frequency at low sound pressure levels, but this did not occur at high levels. These results further support a dissociation between FM and AM encoding in the A1 (central) and the cochlea (peripheral).

When all the evidence was taken together, Altmann & Gaese (2013) concluded that there is at least partial overlap of FM and AM neural populations, particularly at lower FM frequencies. They also stated, however, that current MRI technology may not have high enough spatial resolution yet to precisely differentiate between the neural populations based on FM direction or FM-AM differences. With current technology, it seems better to focus on temporally differentiating these two populations by exploring neural adaptation with the ERPs (e.g., the N1, Mäkelä et al., 1987). Altmann & Gaese (2013) also call for more research to link the general auditory processing FM literature to the speech-perception literature, citing the afore-mentioned work of Holt (2006). Such

research would provide invaluable information on the nature of many human communication and related disorders, such as dyslexia and central auditory processing disorders (CAPD). It could also be used to help track the progress of language learning and therapy. Finally, it would also help to lay the groundwork for a clinical tool for assessing receptive language abilities (i.e., speech-perception) in people with expressive communication disorders (e.g., dysarthria, aphasia, cognitive difficulties - Kujala & Näätänen, 2010).

Limitations of FM tone studies

As with speech-sound studies, there are several limitations to the FM tone studies, particularly in terms of generalizing the general auditory perception principles they reveal to speech perception. The issues of stimulus presentation context and parameters also exist with FM tones, but at the opposite end of the "control-naturalness" continuum. For example, there is a need to balance stimulus "naturalness" (more complex, as in vowel studies) and control the number of parameters manipulated (less complex, as in pure-tone studies). Indeed, Poeppel and colleagues state that studies of speech perception need to not only consider auditory neuroscience theory, but also linguistic theory (e.g., phonetics - Poeppel et al., 2008). Thus, more research needs to be done incorporating both perspectives.

Speech-perception: From cochlea to A1

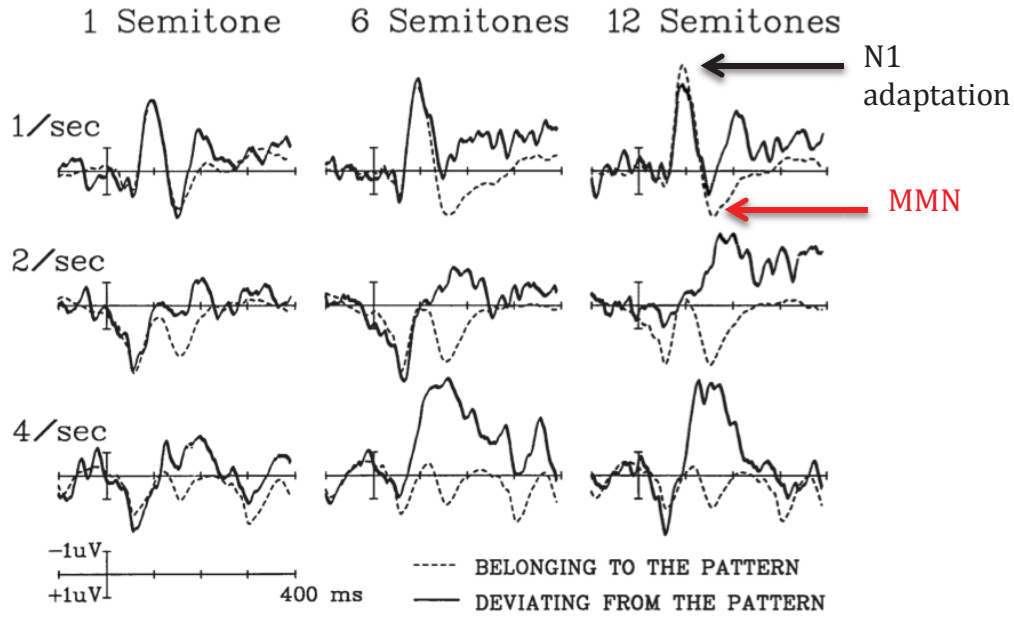
In sum, neural encoding of sound begins at the cochlea, and ends at the A1. The cochlea performs Fourier analysis on incoming sound waves, (e.g., formants, breaking them down into their component sinusoids, and these are further encoded in various steps travelling up the auditory pathway). The general consensus is that the interface between encoding and discrimination in general auditory perception occurs somewhere on the STG (e.g., A1 - Altmann & Gaese, 2013; Poeppel et al., 2008). The interface between auditory perception and higher-level speech perception processes, like phonemic mapping, occurring at the posterior STG (e.g., planum temporale, Griffiths & Warren, 2002; Poeppel et al., 2008). Thus, there should be neural populations in the left STG that are sensitive to formant transition directions, the basic units of speech perception, and activity from these populations can be recorded via EEG. No studies to date have

examined VISC from the perspective that the N1 may be sensitive to individual formant trajectories, as is suggested by the FM literature. Again, this can be examined using N1 adaptation as an index of neural sensitivity, as long as it is reliably separated from the MMN. Finally, along this vein, EEG studies with vowels have generally not considered the implications of the more general pure tone and FM tone literature for careful presentation of stimuli. The present study aims to rectify these issues.

The current study

We designed a unique stimulus presentation sequence to distinguish between changes related to N1 adaptation and true discriminatory MMN responses. We aimed to use the N1 as a measure of which neural populations are activated by different stimuli, and the MMN as a measure of discriminability. Indeed, there is support for doing this with stimuli that consist of pure tones, as long as strict parameters are followed in terms of stimulus pattern presentation so that the N1 and MMN (see above) can be accurately distinguished from one another (Alain et al., 1994; Bendixen & Schröger, 2008; Näätänen & Picton, 1987; Sussman & Gumenyuk, 2005). For instance, Alain and colleagues (1994) presented pure-tones at ISIs of approximately 1 second, and tone separations of 12 semitones (707 – 1414 Hz). They showed a true pattern MMN in response to the repetition deviant along with an apparent reduction in the size of the N1 (see Figure 3). As with many of the vowel studies, they did not analyze this change in the N1, unfortunately, so it is not known if this was a significant difference. If significant it would suggest neural adaptation. We used a similar paradigm to distinguish N1 adaptation from MMNs, so that we could determine whether formant trajectories are encoded by distinct neuronal populations in the A1.

Figure 3 Clear, distinct N1 adaptation and MMN for deviant pure-tones, deviating from standards by 12 semitones, at a 1 second ISI (Alain et al., 1994, p. 141).



In the creation of our paradigm (see Chapter 3) steps were taken to ensure efficient and timely data collection without conflating the N1 and the MMN together. For instance, both Alain and colleagues (1994) and Sussman and Gumenyuk (2005) who used the oddball paradigm, showed that with shorter ISIs (e.g., 500 ms and 600 ms respectively) and smaller frequency separations, the MMN and N1 become conflated together and difficult to analyze, but slightly longer ISIs (800-1200 ms) have been most effective for showing N1 adaptation (Näätänen & Picton, 1987). We used an ISI of 680 ms (i.e., stimulus onset asynchrony of 800 ms). Additionally, Bendixen & Schröger (2008) showed that as few as two repetitions of a short pattern such as ABAB (less than 10s duration, Sams, Hari, Rif, & Knuutila, 1993; Sussman & Gumenyuk, 2005) are necessary to establish a mental rule that, when broken, can elicit a distinct MMN. We used approximately two pattern repetitions to elicit our MMNs, which decreased the amount of time it took to collect data.

By ensuring that our paradigm conformed to the above parameters, we could predict the relative contributions of N1 refractoriness and the MMN to the overall waveforms elicited by deviant stimuli presented in various, pseudo-randomized patterns. We subtracted certain of these waveforms from one another to isolate either component for analysis. Additionally, we used stimuli that were simple enough (e.g., varying in only

one feature) to adequately reflect changing activation across a neuronal population so that we could be certain about our calculations.

To determine whether our speech-sound (vowel) results agreed with the tone literature upon which our stimulus presentation paradigm is based, we also studied FM tone glides that are analogous to the formant transitions in our speech sound stimuli. We then compared the results between these two halves of our study (i.e., FM tone and vowel halves). Given the importance of VISC for speech perception, it is plausible that that changes in formants are also processed preattentively and that the same neural populations coding changes in FM tones also code changes in equivalent formant frequencies. Most importantly, our stimuli varied such that the long-term spectrum of each stimulus was the same but the direction of formant-frequency change (i.e., F_2 trajectory) was varied within the stimulus in a systematic way. Standards and deviants had equivalent spectral content, but their F_2 or glide trajectories were mirror opposites of one another. Furthermore, we statistically analyzed the differences in both the N1 and the MMN elicited by these formant changes. This made it possible to determine whether formant trajectory is explicitly encoded in the A1 in an absolute sense and whether or not these changes give rise to a robust MMN.

Goals

1. Firstly, we wanted to see how the processing of the direction of spectral transitions (formant and pure-tone) in the auditory cortex relates to the N1 and the MMN (i.e., whether formant trajectories are encoded by discrete neural populations and whether they give rise to clear MMNs).
2. Secondly, this study aimed to show electrophysiological responses related to VISC so that future studies can study the relationship between (behavioral) VISC perception and neural encoding.
3. Thirdly, while this study falls under the domain of basic research, we hoped it would add to the growing body of research that supports the clinical utility of the MMN as an index of speech discriminability (especially in the absence of behavioural data collection).

Hypotheses

1. First, we hypothesized that changes in the direction of formant trajectories and FM tone sweeps will lead to a release of N1 adaptation (showing that they are encoded by unique neural populations).
2. Second, we hypothesized that changes in the direction of formant trajectories and FM tone sweeps will give rise to a reliably distinguishable MMN.
3. Third, we hypothesized that changes in the size of the N1 and MMN will be similar for FM tone sweeps and formant trajectories (showing similarities in the way that FM tone sweeps and formant trajectories are encoded in the A1).

CHAPTER 3 METHODOLOGY

Participants

10 adults (faculty and students from the School of Human Communication Disorders at Dalhousie University in Halifax, Nova Scotia, Canada; SHCD) participated in this study (2 male; ages 24 - 40, M = 27.3, SD = 4.81). They were all right-handed, had normal or corrected to normal vision, normal hearing (as screened below) and no known neurological impairments.

Stimuli

In one half of the experiment, stimuli were FM tones (glides), while in the other, stimuli were vowels, /ɪ/ and /e/ (in the subjective impression of the experimenters), synthesized with a MATLAB implementation of a Klatt synthesizer (Klatt, 1980). Vowel stimuli were synthesized to be 120 ms in total duration at a sound level of 74 dB SPL. F_1 and F_3 were held constant, with only F_2 varying throughout the stimulus duration, with the F_2 formant trajectories of each vowel being exact temporal reversals of each other (i.e., / versus \). F_0 was synthesized at 100 Hz, F_1 at 430 Hz, F_3 at 2570 Hz, and F_2 was linearly frequency modulated from 1660-1910 Hz for /e/ (see Figure 4) and 1910-1660 Hz for /ɪ/ (see Figure 5) over the entire vowel duration. All other parameters were left at their defaults according to Klatt (1980). Stimuli were also resampled to 32 kHz (length of stimulus was 3840 samples), and padded with 21760 zeros for a SOA of 800 ms (25600 samples in total). FM tones were linearly frequency modulated in the same manner as F_2 transitions in vowels, but without f_0, F_1 or F_3, and scaled to have the same amplitude as F_2 (68 dB SPL band power). The tone sweeps were ramped to have similar root-mean square (RMS) window shape over time as the synthesized vowels (i.e., fast onset of 1 ms; slower offset of 10 ms similar to the Gaussian smoothing of ends of stimuli in Altmann et al., 2011), after being filtered between 1300 and 2250 Hz.

Figure 4 Synthesized vowel /e/, with f_0, F_1 and F_2 frequency and time values displayed (yellow-orange). FM tone (ascending) frequency and time values corresponding to those of F_2 in /e/ are overlaid (represented by black line).

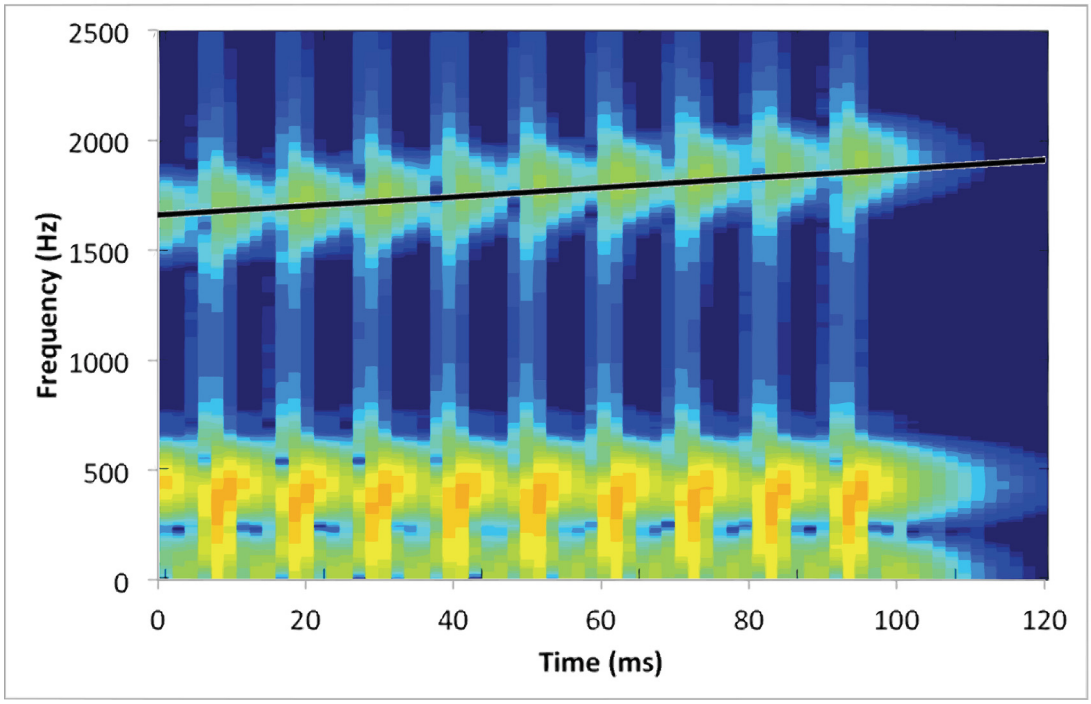
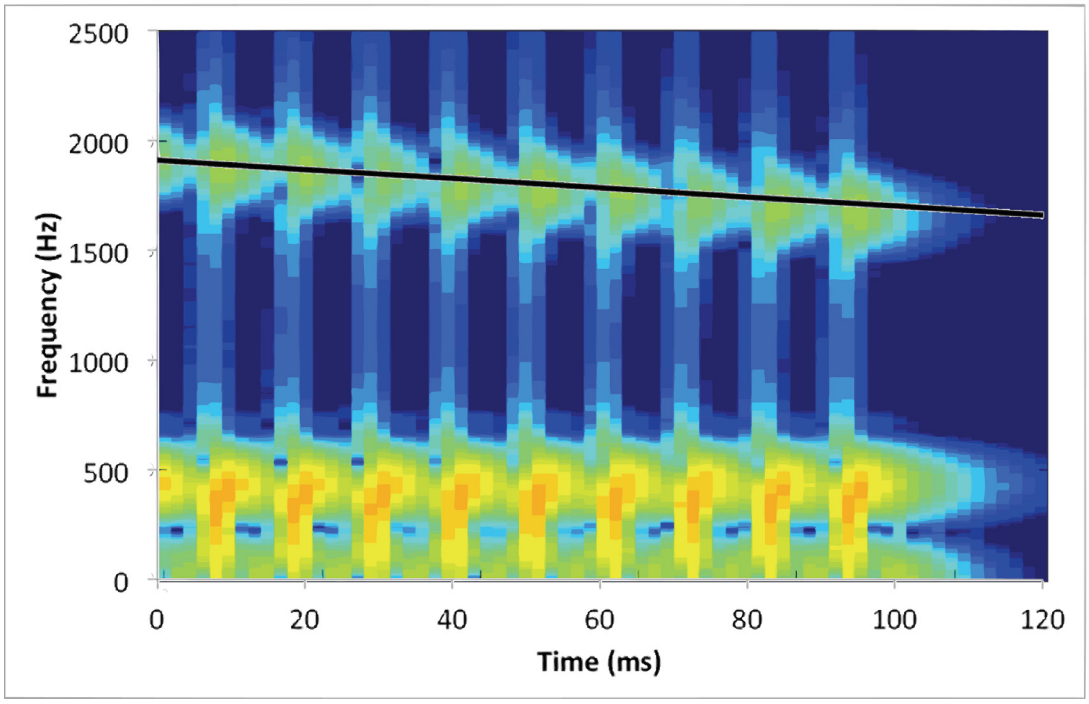


Figure 5 Synthesized vowel /ɪ/, with f_0, F_1 and F_2 frequency and time values displayed (yellow-orange). FM tone (descending) frequency and time values corresponding to those of F_2 in /ɪ/ are overlaid (represented by black line).



Procedure

Hearing screening

After consenting to participate in the study, participants' hearing was screened according to standard audiological procedures in a sound-attenuated room. Specifically, brief tones were presented through insert headphones to right and left ears, respectively, at 25 dB HL, at 250, 500, 1000, 2000, 4000, or 8000 Hz.

Task

Once the hearing screening and EEG preparation were complete, participants watched a subtitled movie of their choice while stimuli were presented (see below). Participants were instructed to ignore the sound stimuli and focus attention on enjoying the movie. Subjects were seated in a comfortable chair in a sound-attenuated and electrically shielded electrophysiology booth for the duration of the experiment.

Stimulus presentation

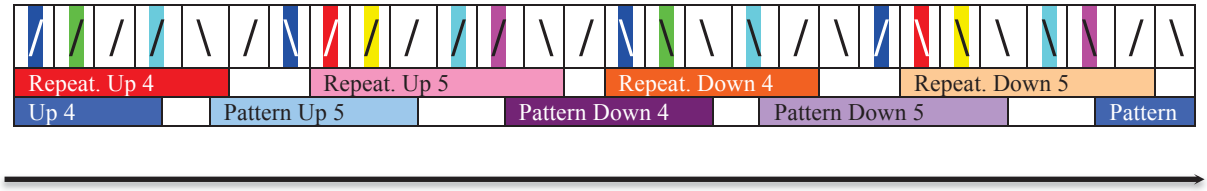
Stimuli were presented binaurally using a custom virtual instrument designed in LabVIEW (National Instruments, Austin TX), and played by a National Instruments PXI 4461 Dynamic Signal Acquisition Card (National Instruments, Austin, TX) routed through a GSI 61 audiometer (Grason-Stadler, Eden Prairie, MN) and ER-3 type insert earphones. Stimuli were presented at a level of 74 dB SPL for vowels and 68 dB SPL for FM tone sweeps (i.e., the level of F2 in the vowels). Stimuli were presented continuously with an 800 ms SOA (onset-onset), with a 680 ms ISI (offset-onset). This stimulus presentation speed is at the slow-end (e.g., 900 ms, and 750-650 ms) of the range successfully used in the literature to elicit a pattern MMN (Alain et al., 1994; Sussman & Gumenyuk, 2005). Half of the participants heard the FM tone sequence before the vowel sequence, while the other half completed vowels before tones, to account for potential fatigue effects in the data. Participants were assigned to sequence order in a counterbalanced manner.

Presentation paradigm design

Stimuli were presented in a pattern designed to distinguish changes in adaptation (that give rise to N1s) from pattern violations (that give rise to MMNs). For both formant transitions and FM tones, there were two general sequences of stimulus presentation – "repeated" and "pattern" (see Figure 6). The repeated sequences consisted of the same stimulus repeated successively several times and were used for the purposes of analyzing N1 adaptation. (While the repeated sequence resembles an "oddball" sequence, the "oddball MMN" was not analyzed due to potential contamination by the N1 effect). The pattern sequences required at least four standard stimuli to create a pattern (one pattern example is established by two standard stimuli, e.g., $A_1A_2A_1A_2B$ or /\/\\) and were used for eliciting a pure MMN with no contaminating N1 effects. There were two repeated and two pattern micro-sequences created for both types of transition (/ = up, \ = down), each of which were presented 100 times. The first of each micro-sequence type contained four standards followed by a deviant (e.g., ////\), while the second of each micro-sequence type contained five standards followed by a deviant (e.g., /////\). This created a total of 8 different micro-sequences for presentation (see Figure 6).

These micro-sequences overlapped so that the last two stimuli of the preceding micro-sequence began the pattern of the following micro-sequence. This helped form a pseudorandom order that was imperceptible to participants, and was recycled until enough data was collected. The long-term pattern of the paradigm was believed to be imperceptible in the subjective impression of the experimenters. Also, it contained 28 stimuli and a full duration was 22.4 s, which is longer than the 10 s human auditory sensory memory trace (Sams et al., 1993). Furthermore, the overlapping nature of the micro-sequences had the added bonus of speeding up the data collection process, which made the task less onerous for participants. (This is similar to Altmann and colleagues' (2011) "roving oddball" paradigm, but with pattern-MMN microsequences inserted in between direction changes). Finally, as noted above, we presented a subtitled movie for participants to watch while the auditory stimuli were presented to them in order to preclude conscious attention being devoted to long-term paradigm pattern (Bendixen & Schröger, 2008; Sussman & Gumenyuk, 2005).

Figure 6 One full run of the pseudorandom order, with a total duration of 22.4 s. Forward slashes represent rising F_2 transitions and up glides, while backslashes represent falling F_2 transitions and down glides. Each microsequence has a unique colour (see descriptions in text). Specific stimuli used in calculations are highlighted in unique colours (see descriptions in text).



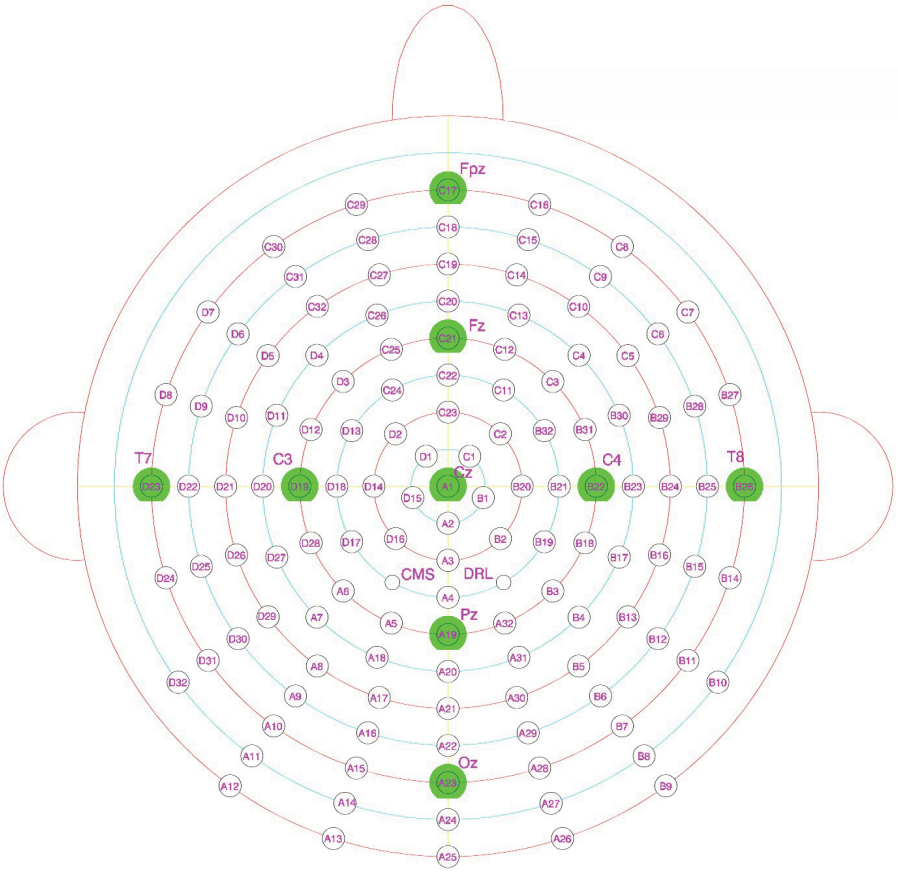
The breakdown of total time for each half of study (FM tones and vowels) is as follows: SOA (800ms) x number of stimuli per run (28) x number of runs (100) = 37.33 minutes. Thus, each half of the study took approximately 37 minutes to complete for a total recording duration of approximately 80 minutes. For both vowels and FM tones, this yielded 200 "down" and 200 "up" "repeated adaptee" with three preceding identical stimuli (e.g., \\\\, light blue highlighted slashes in Figure 6). This also yielded 100 "up and 100 "down" repeated adaptees with four preceding identical stimuli (e.g., \\\\\, pink highlighted slashes). Equivalent numbers of "alternating adaptees" were elicited for comparison with the repeated adaptees, either 3 or 4 alternating stimuli preceding (e.g., /\/\ and \/\/\, dark blue and red highlighted slashes, respectively). 100 "up" and 100 "down" "repetition deviants" were obtained with both four preceding standards (e.g., /\/\\, green-highlighted slashes) and five preceding standards (e.g., \/\/\\, yellow-highlighted slashes). The alternating adaptees with three preceding alternations (see above) doubled

in use as the comparators for the 4-standard MMNs, while the alternating adaptees with four preceding alternations doubled as the comparators for the 5-standard pattern MMNs.

Data recording and storage

ERP data were collected via 140-channels using a BioSemi Active-Two biopotential system (BioSemi Instrumentation, 2006). 128 channels were connected via a cap (in addition to 2 other channels for grounding, CMS and DLR; see Figure 7), and there were 10 off-cap electrodes, half of which were placed on the right side of the face, while the other half were placed on the left. Placement locations for the "off-caps" were on the mastoid (TP9/10), in front of the tragus (FT9/10), on the cheekbone (F9/10), beside the outer corner of the eye (horizontal EOG), and under the eye (vertical EOG).

Figure 7 Layout for the BioSemi Active-Two biopotential system 128-channel electrode cap with key 10/20 mapping system electrode locations overlaid (in green; BioSemi Instrumentation, 2006).



Data were recorded at a sample rate of 2048 Hz and stored automatically during the experiment via Actiview 7.0 Software (Biosemi, Amsterdam, Netherlands), running within LabVIEW (National Instruments, Austin, TX). The data consisted of continuous raw waveforms that were passively elicited in the brain in response to change in the direction of FM tone sweeps/formant transitions. Prior to analysis, waveforms were downsampled to 512 Hz, and segmented into epochs for each stimulus condition (e.g., repetition deviant and repeated adaptee) consisting of -200 ms prestimulus (baseline) to +800 ms post-stimulus time windows. Data were also re-referenced to the average raw waveform of all electrodes. BESA (Brain Electrical Source Analysis, Megis AG, Gräfelfing, Germany) software was used to process the data (e.g., remove artefacts, remove activity from eye-blinks, average data) prior to statistical analysis. Statistical analyses were then conducted in BESA Statistics (Megis AG, Gräfelfing, Germany) and MATLAB (The Mathworks, Natick, MA).

Statistical calculations

To determine statistically significant differences between the waveforms for specific conditions, non-parametric bootstrap analyses were performed on the data over multiple electrodes (128) and time-points (0-300 ms post-stimulus presentation) (Maris & Oostenveld, 2007). This is a specific method of multiple-comparisons correction that controls for the family-wise error (FWE) rate associated with performing a large number of statistical comparisons (i.e., data from several spatial locations over hundreds of time points). It has the advantage of being less conservative, and more sensitive to statistical differences than the typically used Bonferroni-correction. Here, electrode-time pairs (i.e., data samples) that differ between the two conditions being compared, as measured by a *t*-value above a threshold set at 0.05, are clustered together based on both temporal and spatial adjacency. Cluster-level statistics are then calculated by summing the *t*-values within each identified cluster. The largest of the cluster-level statistics is the means by which the difference between conditions is evaluated. The *p*-value is calculated by combining the data samples from all conditions into one large pool, from which the same number of trials are drawn as there were in the first condition. The test statistic (i.e., the *t*-value) is then calculated for this random partition of the data. This procedure is

repeated 1000 times to create a histogram of the *t*-values. The number of random partitions that resulted in a *t*-value that is higher than the one actually observed is then calculated, resulting in a *p*-value. A *p*-value smaller than 0.05 represents a significant difference (Maris & Oostenveld, 2007).

The first hypothesis—that FM tones and formant trajectories are encoded by unique neural populations and should therefore exhibit a release from adaptation—was assessed by comparing the amplitude of the N1 from repeated adaptees with the N1 from the alternating adaptees (see Figure 6). The N1 following 3- or 4-repeated adaptees should be smaller in size because of adaptation due to repeated stimulation of the same neural population. If unique neural populations are involved in coding each trajectory, the N1 following a stimulus embedded in an alternating trajectory sequence should be less adapted and therefore larger. We could be sure that this N1 change was not contaminated by an MMN either because these stimuli were not different from the preceding stimuli (e.g., repeated adaptee, from repeated microsequence) and no pattern was broken (e.g., alternating adaptee, from pattern microsequence). Bootstrap analyses (see above) were used to compare N1 amplitudes, with the N1 being defined as the largest negative peak occurring 70-150 ms post-stimulus. The difference between the two is a measure of the release from adaptation that occurs due to the change in the direction of the FM tone or formant trajectory, and thus a measure of the extent to which each is encoded by a unique population of neurons. We hypothesized that the N1 would be significantly larger to the "alternating adaptees" for both FM tones and formant trajectories than for "repeated adaptees."

The second hypothesis—that FM tones and formant trajectories would give rise to an MMN—was assessed by comparing the response to a "repetition deviant" to the response to the immediately preceding standard stimulus (see Figure 6). We could be sure that any increased negativity associated with a repetition deviant could not be attributed to a release from N1 adaptation because the repetition would increase N1 adaptation. Furthermore, the preceding standard stimuli were part of an alternating, equiprobable pattern, meaning that relative amounts of adaptation should be similar between standards and repetition deviants (although slightly increased for the latter). Bootstrap analyses were used to compare waveforms for significant differences in the

time-window of the MMN (100-250 ms post-stimulus). We hypothesized that there would be a significantly increased negativity for the repetition deviant for both FM tones and formant trajectories.

The third hypothesis— that changes in the size of the N1 and MMN will be similar for FM tone sweeps and formant trajectories —was assessed by observing patterns in statistically significant results (for N1 adaptation and the MMN) within each condition and comparing them across conditions (i.e., absolute amplitudes of the N1 and MMN and degree of adaptation).

Summary of MMN elicitation and calculation effectiveness

According to parameters in the above presentation protocol, stimulus duration, loudness, ISI, SOA were all within the typical ranges used in the literature to elicit N1s and MMNs (see Chapter 2). Additionally, several steps were taken to ensure that the N1 and MMN would be easily separable using this protocol in conjunction with the MMN calculation method: 1) Stimuli were presented at a fast enough rate to elicit a pattern MMN (which has no N1 effect), but slow enough to show a distinct N1 (Alain et al., 1994; Sussman & Gumenyuk, 2005). 2) Stimuli were presented fast enough to show N1 adaptation effects (Näätänen & Picton, 1987). 3) The pattern MMN was derived from subtracting the preceding (alternating) stimulus—which should also have a relatively equally adapted N1—from the repetition deviant. 4) Attention was diverted from auditory stimuli to reduce N1 attention effects, while maintaining the amplitude of the MMN. 5) The difference between standard and deviant stimuli (for both FM tones and vowels) was kept small to increase the latency between the N1 and MMN.

CHAPTER 4 RESULTS

Data preprocessing

Data was preprocessed with eye correction (based on principle components analysis; PCA) turned on and bandpass filtered (1 – 30Hz). This was done to ensure that the following artefact scan would not reject trials due to the mere presence of eye activity. Specifically, the artefact scan involved the removal of any bad electrode channels on a subject-by-subject basis, as well as automatic artefact rejection based on amplitudes exceeding 120 μV (in the passband). At least 91-100% of the data was accepted for each participant for each type of waveform (i.e., 4-standard repetition deviant) used in N1 and MMN calculation (see Table 1).

Table 1 Percent data accepted for all conditions by participant.

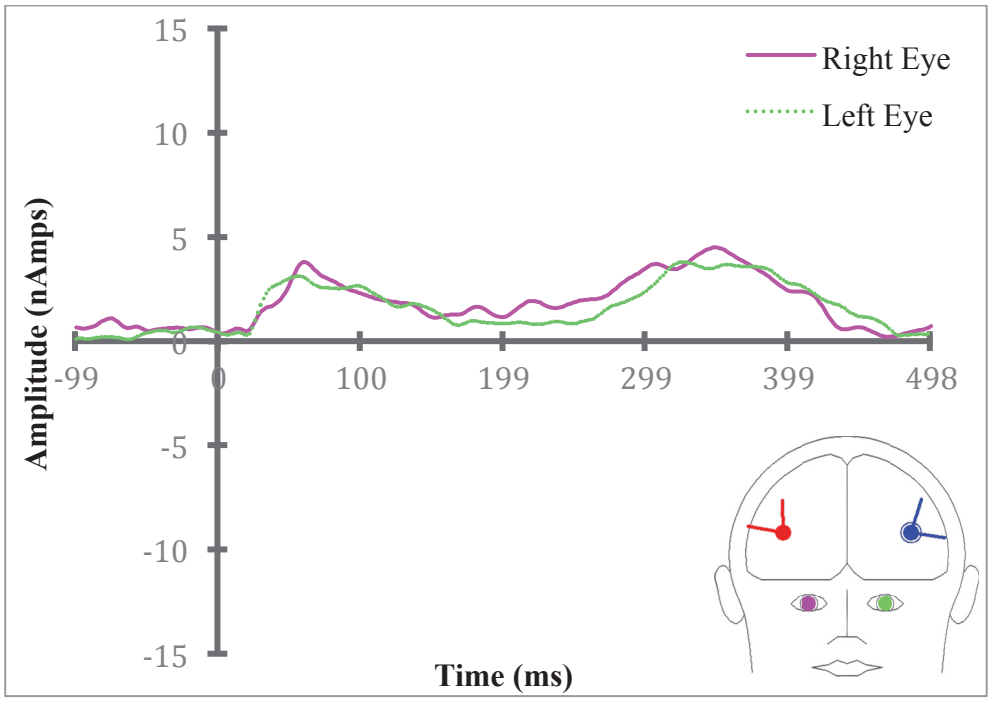
Participant Number	Percent data accepted (all conditions)
100	98-100%
101	98-100%
102	94-97%
103	93-99%
104	98-100%
105	91-96%
106	97-100%
107	97-100%
108	98-100%
109	99-100%

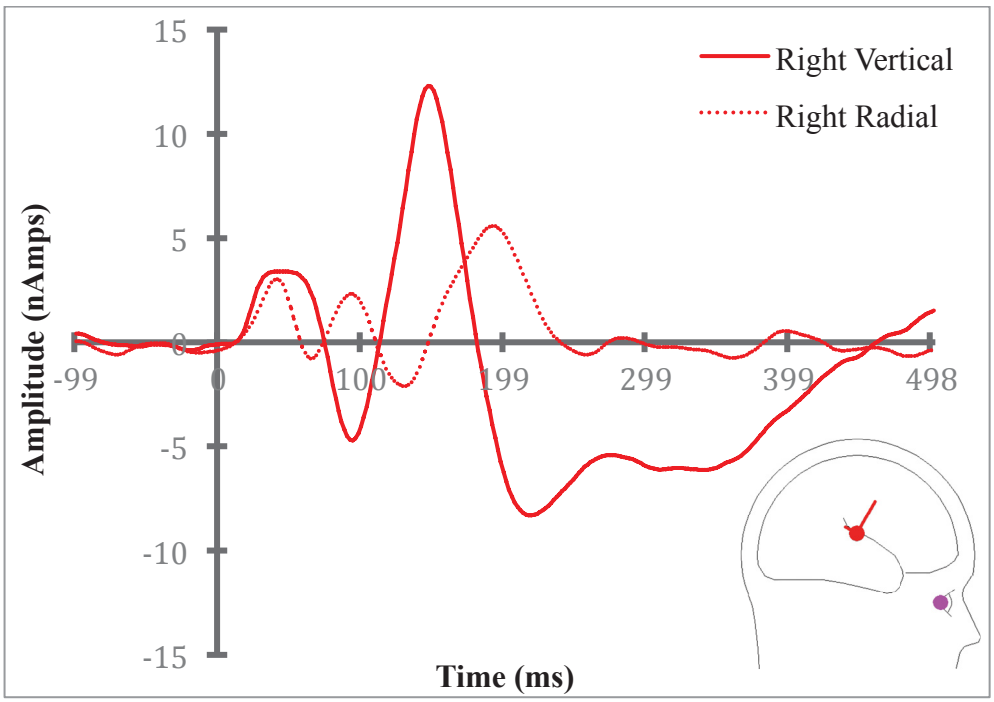
Source analysis

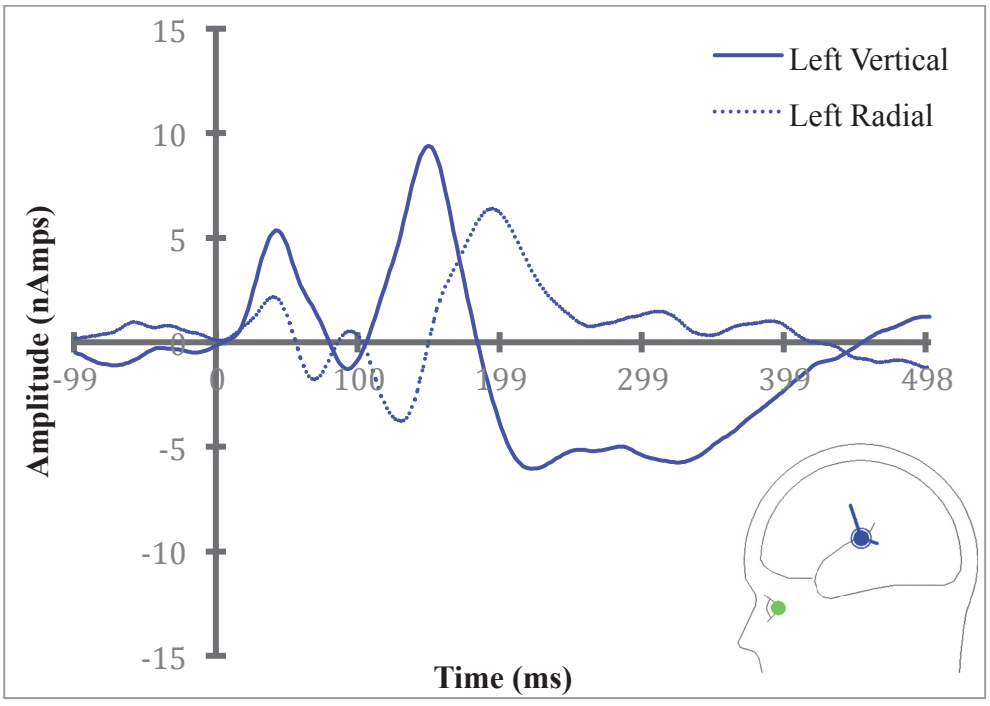
Eye-correction was disabled for the accepted epochs of the above data in order to perform source analysis. This is because eye correction can distort the topographical scalp data used to estimate cortical sources. Source analysis was conducted on the grand-averaged waveforms to isolate any activity associated with eye movements and guide the fitting of a discrete source model using BESA software (i.e., to find current sources for the electrical activity measured at the scalp). Specifically, two symmetrical regional sources were placed in each hemisphere—one to capture eye-blinks and another to capture activity in response to stimuli, hypothesized to emanate from a location near HG. The two regional sources were first manually placed over the eyes. A spatial PCA was then computed and the remaining sources were fit over a time period in which most of the

variance (> 90%) was associated with the first principal component. This period was from 75 to 200 ms. The two eye-related sources were then refit based on the entire epoch, under the assumption that most of the remaining activity would be eye-related. The resulting model had a residual variance of 2.671% for vowels (see Figure 8), and 2.987% for FM tones. The regional sources in both right and left HG were broken out into three dipoles with orthogonal orientations (based on activity from 75 to 200 ms). The vertical dipoles showed waveforms with clear P1-N1-P2 complexes—the right one being slightly larger than the left—while the radial dipoles produced waveforms that fit the criteria to be T-complexes (see Chapter 2). The third and smallest dipoles were not analyzed further. No significant differences were found in the source locations for adapted N1s and MMNs, controlling for FWE.

Figure 8 Sources for activity scalp activity in response to vowel stimuli (2.671 residual variance). Sources for eye activity are pink and green, with matching-coloured waveforms. Right-hemisphere dipoles (vertical for N1 and radial for MMN) are in red, while equivalent left-hemisphere dipoles are in blue.







Due to the noisiness of the data, source-analysis beyond the level of the grand-averaged data for all participants and conditions (e.g., subject-by-subject, condition-by-condition) proved inconclusive (i.e., similar sources were not reliably found across

similar conditions or subjects). While some subjects had clear vertical and radial sources, with similar waveforms for the grand-averaged data (Figure 9), many did not.

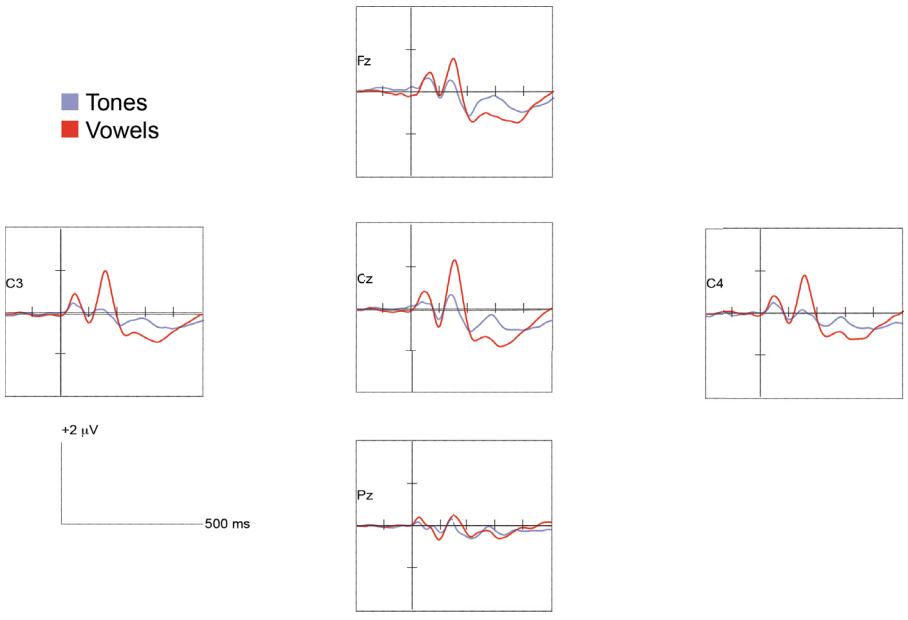
Scalp data analysis

Since source waveforms were unreliable at the individual subject levels, and statistical analysis requires individual subject data, the initial raw scalp data (i.e., with eye correction turned on) was used for all statistical analyses (see Table 1).

Onset responses (scalp waveforms)

When all waveforms were averaged together for tones and subsequently for vowels, a P1-N1-P2 complex emerged for both types of stimuli at electrode Cz (see Figure 9). Upon visual inspection of the data the N1 appeared to be slightly larger and the P2 considerably larger for the vowels than for the FM tones. Bootstrap analyses on the difference between the vowel and FM tone data revealed that the increased positivity for vowels at around the P2 time (i.e., 162 ms) approached significance at electrode C3 (p = 0.084). When the time-window for analysis was narrowed to 100-200 ms post-stimulus onset, however, this difference became significant (p = 0.046). No other differences between the vowel and tone waveforms were found to be significant (or approaching significance).

Figure 9 Eye-corrected, grandaveraged onset responses for all tones (blue/light) and vowels (red/dark) and 5 different scalp electrode sites (Fz, C3, Cz, C4, Pz).



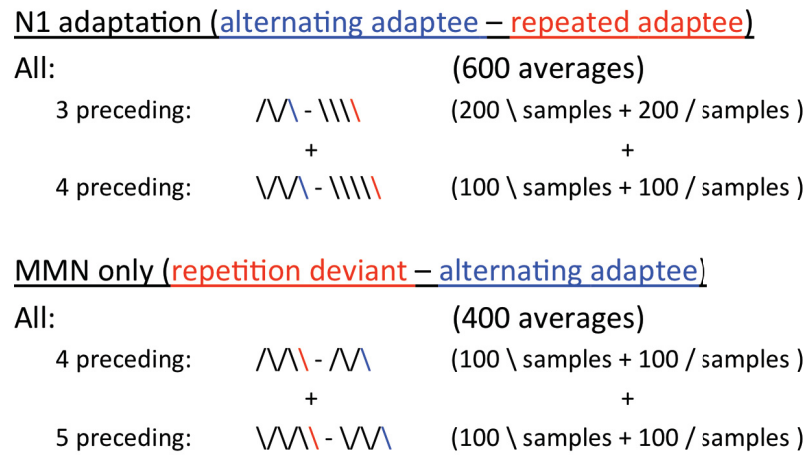
"Up" versus "down" comparison

The difference between "up" stimuli and "down" stimuli for all conditions averaged together at the grand-average level was analyzed, but did not yield any significant results.

Vowels: N1 adaptation and pattern MMN

For scalp data, significant variability in cortical responses continued to prove to be a complicating factor at the individual subject level. Data were, therefore, collapsed across conditions to gain more power in the analysis (see Figure 10). For all comparisons, the "up" and the "down" waveforms were collapsed together because there were no significant differences found in the "up" and "down" scalp patterns and because the difference between standards and deviants should be equivalent for "ups" and "downs." All waveforms elicited by repeated adaptees, regardless of the number of preceding stimuli, were collapsed together as well. This was done for waveforms elicited by each of alternating adaptees, repetition deviants, and the stimuli immediately preceding repetition deviants. This yielded a grand total of 600 repeated adaptees and alternating adaptees for calculating N1 adaptation and 400 repetition deviants and immediately preceding stimuli for calculating the pattern MMN (for each subject).

Figure 10 Final calculations used in scalp data analysis for determining the presence of N1 adaptation and the pattern MMN.



N1 adaptation: Vowels

To examine the first hypothesis, that FM tones and formant trajectories are encoded by unique neural populations, bootstrap analysis was used to look for significant differences between the repeated and alternating adaptee waveforms around the N1 time. Three significant differences between the repeated and alternating adaptee waveforms were revealed (where a release from adaptation would be expected in the case of the response to the stimulus embedded in an alternating pattern). The first consisted of a decreased positivity for the repeated adaptee below electrode F4 at about 50 ms post-stimulus onset (p = 0.002; see Figure 11—red vertical line marks time-point of significant difference). The second consisted of an increased negativity (or decreased positivity) for the repeated adaptee posterior to and below electrode C4 at about 100 ms post-stimulus onset (p = 0.005; see Figure 12). The third consisted of a decreased positivity for the repeated adaptee at electrode C3 at about 140 ms post-stimulus onset (p = 0.005; see Figure 13).

Figure 11 Significant (p = 0.002) decreased positivity for the repeated adaptee below F4 (at Biosemi electrode C10; black arrow) at about 50 ms post-stimulus onset (red vertical line marks time-point of significant difference). Top graph displays compared waveforms (repeated adaptee in black and alternating adaptee in grey), while bottom graph displays the difference waveform.

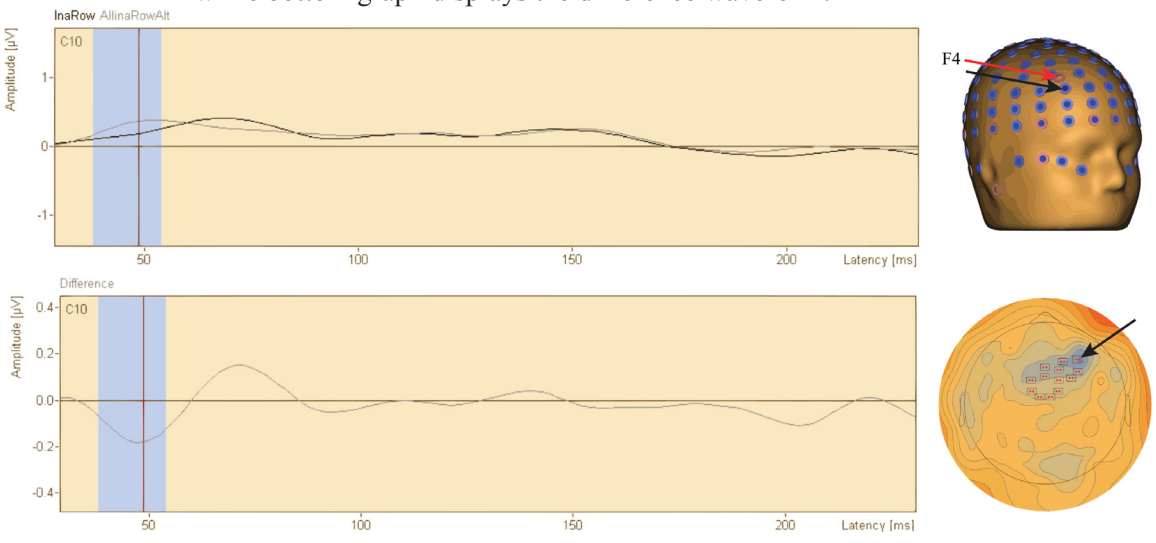


Figure 12 Significant (p = 0.005) increased negativity (or decreased positivity) for the repeated adaptee posterior to and below C4 (at Biosemi electrode B17; black arrow) at about 100 ms post-stimulus onset (red vertical line). Top graph displays compared waveforms (repeated adaptee in black and alternating adaptee in grey), while bottom graph displays the difference waveform.

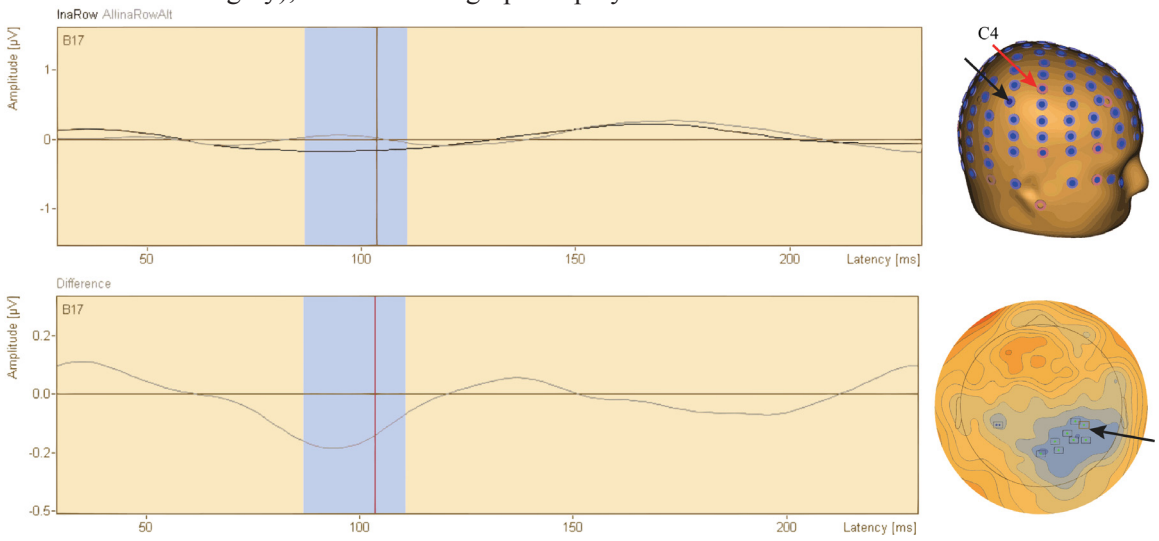
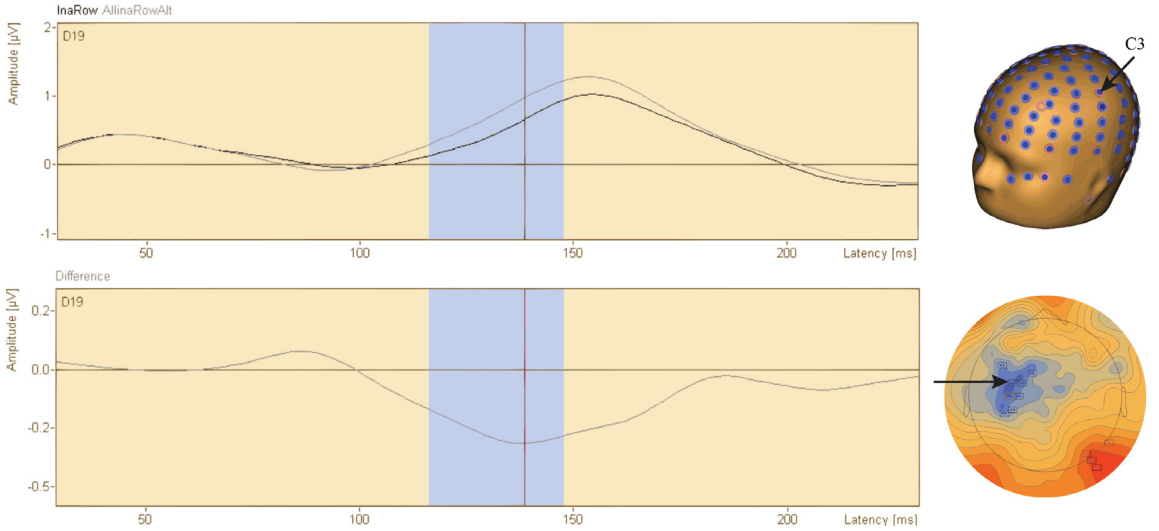


Figure 13 Significant (p = 0.005) decreased positivity for the repeated adaptee at C3 (at Biosemi electrode D19; black arrow) at about 140 ms post-stimulus onset (red vertical line). Top graph displays compared waveforms (repeated adaptee in black and alternating adaptee in grey), while bottom graph displays the difference waveform.



Pattern MMN: Vowels

To examine the second hypothesis, that FM tones and formant trajectories would give rise to an MMN, bootstrap analysis was used to look for significant differences between the repetition deviant and immediately preceding standard stimulus waveforms

around the MMN time. Four significant differences between the repetition deviant and its immediately preceding (standard) stimulus waveforms were revealed. The first consisted of a decreased negativity for the repetition deviant posterior to and above P10 at 43 ms post-stimulus onset ($p = 0.002$; see Figure 14—red vertical line marks time-point of significant difference). The second consisted decreased positivity for the repetition deviant above P10 at 92 ms post stimulus onset ($p = 0.002$; see Figure 15). The third consisted of a decreased positivity for the repetition deviant below C3 at 140 ms post stimulus onset ($p < 0.0001$; see Figure 16). The fourth consisted of an increased negativity for the repetition deviant below Fz at 178 ms post-stimulus onset ($p = 0.001$; see Figure 17).

Figure 14 Significant ($p = 0.002$) decreased negativity for the repetition deviant posterior to and above P10 (at Biosemi electrode B7; black arrow) at 43 ms post-stimulus onset (red vertical line marks time-point of significant difference). Top graph displays compared waveforms (repetition deviant in black and the preceding stimulus in grey), while bottom graph displays the difference waveform.

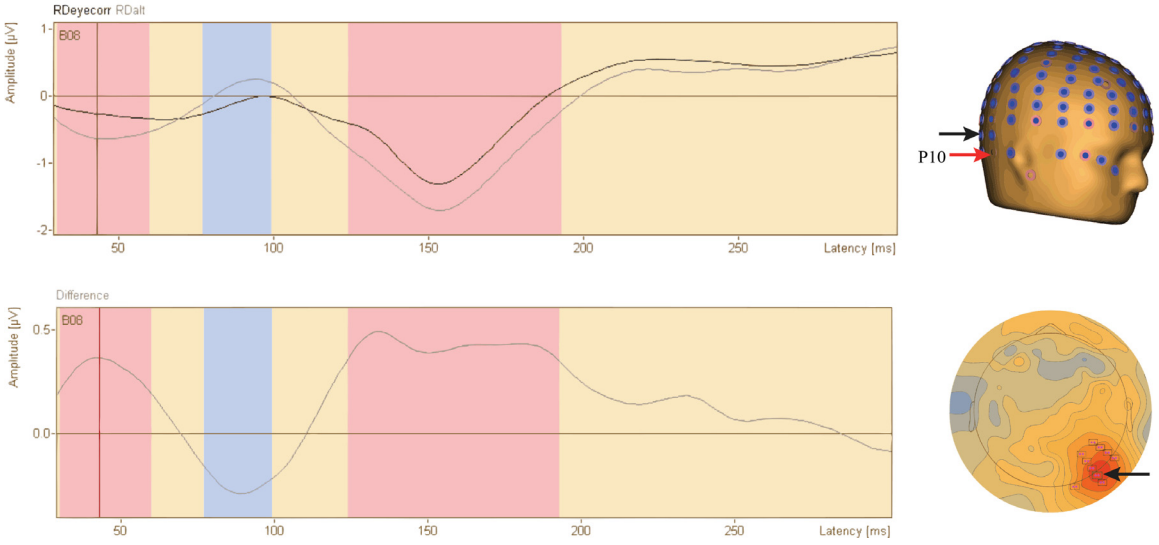


Figure 15 Significant ($p = 0.002$) decreased positivity for the repetition deviant above P10 (at Biosemi electrode B11; black arrow) at 92 ms post stimulus onset (red vertical line). Top graph displays compared waveforms (repetition deviant in black and the preceding stimulus in grey), while bottom graph displays the difference waveform.

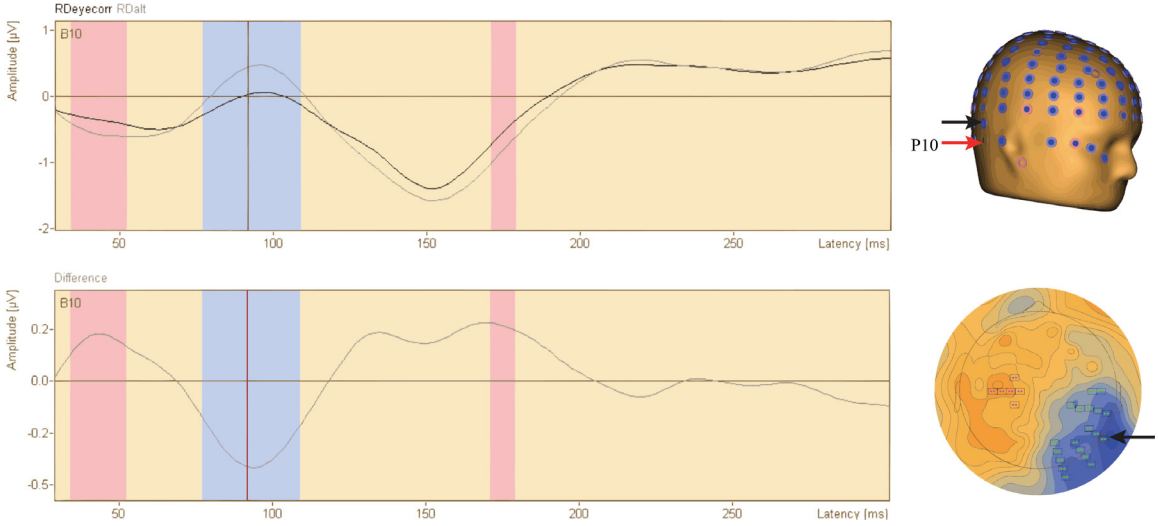


Figure 16 Significant ($p < 0.0001$) decreased positivity for the repetition deviant below C3 (at Biosemi electrode D20; black arrow) at 140 ms post stimulus onset (red vertical line). Top graph displays compared waveforms (repetition deviant in black and the preceding stimulus in grey), while bottom graph displays the difference waveform.

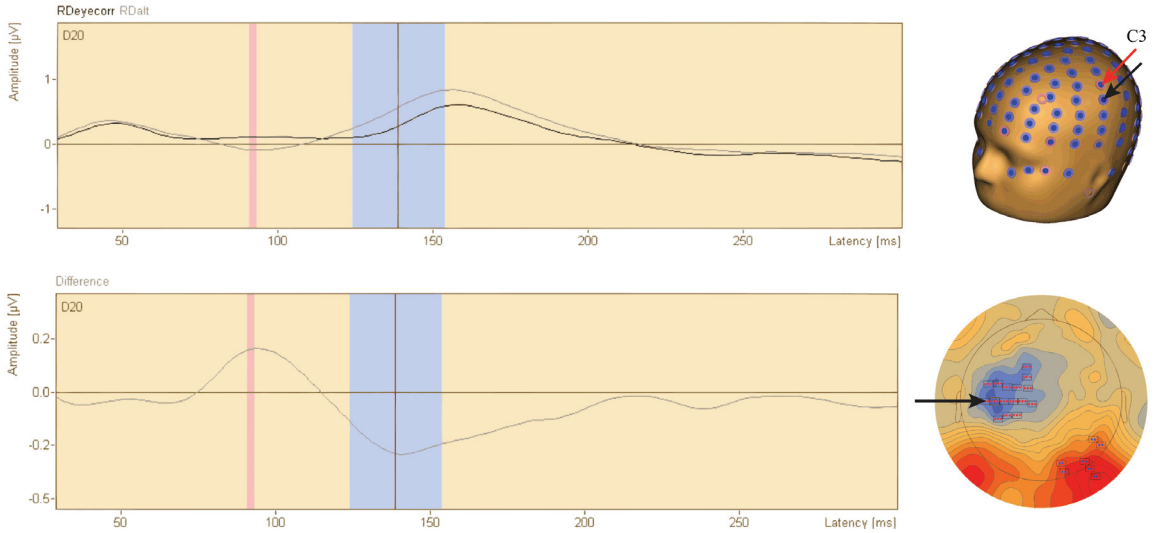
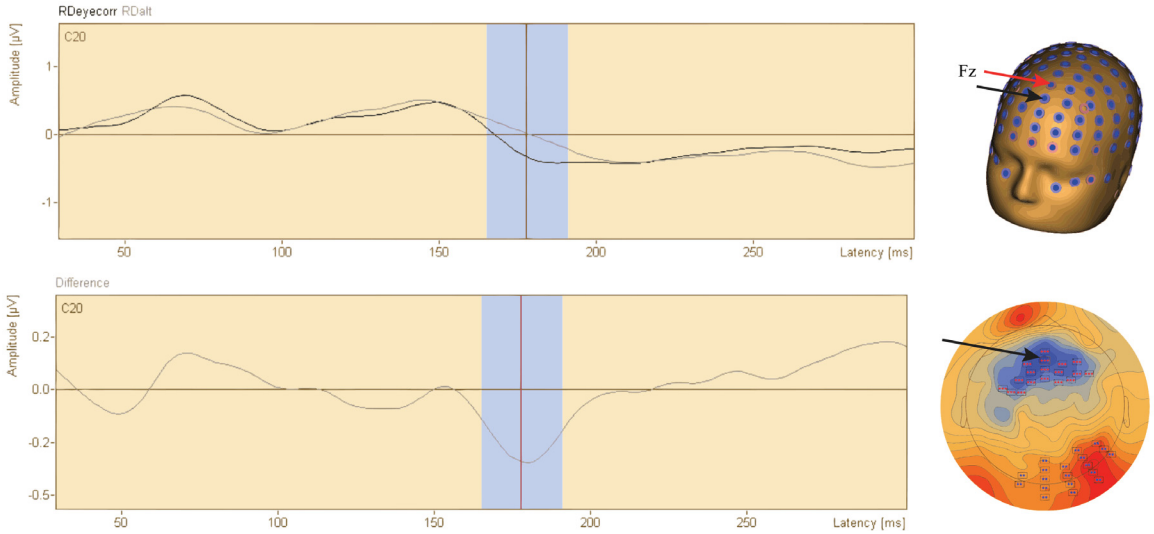


Figure 17 Significant (p = 0.001) increased negativity for the repetition deviant below Fz (at Biosemi electrode C20; black arrow) at 178 ms post-stimulus onset (red vertical line). Top graph displays compared waveforms (repetition deviant in black and the preceding stimulus in grey), while bottom graph displays the difference waveform.



FM Tones: All conditions

None of the statistical analyses performed on the FM tone data yielded any significant differences between conditions, specifically complicating evaluation of the third hypothesis, that FM tone and vowel data would be similar in all conditions.

CHAPTER 5 DISCUSSION

The aim of this thesis was to determine how the processing of the direction of formant transitions and equivalent FM tones in the auditory cortex relates to N1 and MMN production. Other aims included bringing together the behavioural research on VISC and the research on AEPs and adding to the groundwork for a clinically viable tool to reliably assess speech-sound discrimination without the need for behavioural feedback. The current data provide valuable insight into these areas of inquiry.

Source analysis

PCA was useful to do at the grand-average level across conditions because it gave us two reliable dipoles in each hemisphere that explained the majority of the variance (as well as dipoles for eye activity; see Figure 8). The first was a vertical dipole in HG (left and right), which was oriented toward Cz/Fz, where the N1 and MMN are typically measured. This produced a waveform that looks very similar to that which occurs at Cz at the level of the scalp (in uV; see below). This suggests that this is indeed the source for the N1 and MMN measured at the scalp. The second source was a radial dipole in HG (left and right), which may have been the source of some unexpected activity, such as T-complexes, in various waveforms. The T-complex typically is measured at central-temporal scalp sites (Tonnquist-Uhlen et al., 2003); however, we did not find any visual or statistically significant differences within our conditions at these sites. Furthermore, these sites are far enough away from Fz and Cz that any T-complexes should make minimal contributions to the components measured at these locations (e.g., N1 and MMN).

The considerable between subject variability in source waveform may be explained by between-subject differences in the sizes and orientations of cortical folds—and thus dipole orientations—and insufficient amounts of data collected to correctly estimate dipoles in individual subjects. These complicate analysis of waveforms at the source level because the source waveforms may not correspond to any true sources. They do not preclude analyses at the level of the scalp because this data does not require additional assumptions about underlying generators. The activity produced by different source locations may still summate at the scalp (i.e., the inverse problem; Luck 2005),

and differences between subject sources and orientations simply add error-related variance to subsequent statistical analyses.

Scalp data analysis

Onset responses

The finding that the difference in onset responses between vowel and FM tone data around the P2 time was approaching significance may be explained in a number of ways. It could be that this failure to achieve significance accurately reflects the lack of a real difference in responses between tone and vowel stimuli (failure to reject the null hypothesis). However, given that the difference was significant using a slightly narrower analysis time-window, and some of the individual subject data was noisy, it is more likely that the original analysis simply failed to have enough power. Additionally, the fact that there was a significant difference between conditions within the vowel data, but not with FM tone data (see below), could be taken to mean that there is, in fact, a difference between how vowels and tones are processed in the A1, as reflected in the P2.

There is a precedent for P2 enhancement in the discrimination training literature for studies of both tone-pattern distinction (Atienza, Cantero, & Stickgold, 2004)and vowel distinction learning (Tremblay et al., 2009). In the case of vowel-distinction learning in particular, the P2 increase occurred along with, but independent of the N1, resulting in shortened latencies for both components (Reinke, He, Wang, & Alain, 2003). In later studies, both decreases and increases in the amplitude of the P2 were found to indicate successful voice-onset discrimination training of CV pairs (respectively, Alain, Campeanu, & Tremblay, 2010; Tremblay et al., 2009). Indeed, it could be that the enhanced P2 for vowels reflects subjects "picking up" on the subtle distinction in formant trajectory throughout the duration of our study and that this component can be used instead of the N1 to assess the neural encoding of sound, specifically for vowels. That said, there is some debate as to whether the P2 represents precisely the same neural mechanism as the N1, with some researchers showing changes in the P2 that are not necessarily reflected in the N1 (Martin et al., 2008; Tremblay, Billings, & Rohila, 2004). Nevertheless, the P2 does share a close relationship with the N1 and, like the N1, may still be a valid indicator of adapted neural populations that encode specific stimulus

parameters (see below). Indeed, it is present at or near Cz for all conditions for vowels (e.g., N1 and MMN; see below).

One issue in the P2 elicited both within the grand-averaged data and the specific conditions below is that the timing is quite early—as early as 140 ms within the specific conditions. This is considered early because the latency of the P2 typically found in the literature falls around 200 ms post-stimulus onset (Martin et al., 2008). It could be that the P2 is extremely sensitive to the spectral complexity of vowels in particular presented at a fast rate, occurring earlier in time and higher in amplitude than normal. This extreme latency might suggest, however, that the increased positivity is not a genuine P2. It is not unreasonable, however, for a P2 to occur as early as 140 ms (or 162 ms). One study found that the latencies of the P2 elicited by various stimuli (30 ms duration 800 Hz, 80 dB tones, with ISI's varying between 100 – 1000 ms) ranged from 140 – 180 ms post-stimulus onset and were negatively correlated with increased ISI (Wang, Mouraux, Liang, & Iannetti, 2008).

"Up" versus "down" comparison

The finding that there were no statistically significant differences between "up" and "down" data was to be expected because the relative differences between standards and deviants for both sets of data were the same. It could be, however, that had we collected more data, that we would have found a slight difference. Indeed, Maiste & Picton (1989) found that, when they modulated a continuously presented tone with "up" and "down" ramps, the N1 to "up" ramps was slightly larger (Maiste & Picton, 1989).

Vowels: N1 adaptation and pattern MMN

Turning to the specific differences between conditions, there were three specific hypotheses tested in this study. First, we hypothesized that reliably distinguishable changes in the size of the N1 would be caused by formant trajectory differences, with the N1 being larger for formant trajectories following a sequence of alternating stimuli rather than repeated stimuli. Second, we hypothesized that differences in both FM tone glide trajectories and F_2 trajectories would give rise to a reliably distinguishable MMN. Third, we hypothesized that changes in the size of the N1 and MMN would be similar for both FM tones and vowels.

N1 adaptation

With regards to the first hypothesis, there was evidence of generalized adaptation throughout the entire duration of the repeated adaptee waveform, when compared to that of the alternating adaptee (which should be less adapted) for vowels. Specific instances of this adaptation included decreased positivities at 50 ms and 140 ms (being F4 and C3, respectively) for responses to stimuli following 3 or 4 identical trajectories (see Figures 16 and 18). There was, however, no reduction in negativity for the repeated adaptee waveform at (or near) Cz around 100 ms (where one would expect to see N1 adaptation) when compared to the alternating adaptee waveform. In fact, at 100 ms, the opposite pattern expected was observed behind electrode C4 for the repeated adaptee waveform, which had an increased negativity (see Figure 12). It is difficult to compare the repeated and alternating adaptees at this point, however, because the two waveforms appear to be traveling in opposite directions, which suggest that adaptation per se is not occurring here. It could be, however, that our stimulus presentation protocol was "too good" at minimizing N1 effects to isolate the MMN (see Chapter 3), so much so that it was only visible at the grand-averaged, all conditions level of analysis.

The most interesting result from this analysis is the reduced positivity at C3 at 140 ms post-stimulus onset. As with the grand-averaged data from all conditions, this may represent an early P2, which appears to be sensitive to quickly presented vowels. Indeed, the waveforms for the repeated adaptee and alternating adaptee only differ (visually and statistically) in the amplitude of the positive component at 140 ms. The reduction in the response of the early P2 of the repeated adaptee could be directly related to the neural encoding of sound. Indeed, while the significance of the P2 and related adaptation are poorly understood in the literature, recent studies have shown P2 adaptation to be associated with proficiency in the processing of a variety of sounds. For example, P2 adaptation has been correlated with young age in response to repeated meaningful nonspeech sounds (Leung, He, Grady, & Alain, 2013) and successful cochlear-implant use with repeated pure-tones and speech sounds (Zhang et al., 2011). Furthermore, if the N1 and P2 do indeed operate independently of one another, then perhaps the early P2 observed here is still reflecting encoding (like the N1), but is less resistant to the extreme adaptation that our presentation paradigm may have caused in the N1 (see above).

The results from the statistical analysis of the difference between the repetition deviant and its immediately preceding standard for vowels also address the first hypothesis. The early decreased negativity above P10 at 43 ms could be expected in that, as with the repeated adaptee, the repetition deviant is repeated. The location and timing of the decreased negativity is very dissociated from the N1 time and location, however. The same explanation applies for the decreased positivity occurring in the same vicinity at 92 ms, although a decreased negativity would be expected at Cz. P10 is far enough away, however, that we may be seeing an inversion of the N1. Additionally, we did see expected patterns of adaptation around the locations for the N1 (Cz) and MMN (Fz). At C3, similar to the waveform for the repeated adaptee, there was a decreased positivity (early P2) again at 140 ms for the repetition deviant. Also, at around the N1 time, the repetition deviant was slightly less negative (adapted), as would be expected, although this difference did not achieve significance.

Pattern MMN

As with the other conditions, the early P2 was clearly seen near Fz for both the repetition deviant and its preceding standard, but there was a very focal negative deviation of the repetition deviant beginning around 150 ms, peaking at 178 ms, and re-joining the waveform for the preceding standard at 200 ms. This appears to be an isolated component, due to its focal nature, which is very likely to be a legitimate pattern MMN. It is not surprising that we found a significant MMN, but no significant difference in the N1, even though we had less waveforms to average for the MMN than the N1. This is because, although MMNs typically have lower amplitudes and hence lower SNRs than N1s at longer ISIs (Martin et al., 2008), this does not hold true at high presentation rates. This further suggests that the N1 is extremely minimized by our paradigm, and that the P2 may be the component of interest to examine to assess neural encoding. Furthermore, this provides support for the Memory Trace account of MMN production (Näätänen et al., 2005) because the N1 peaks in both the repetition deviant and the preceding standard are of essentially the same amplitude and latency, and thus equally adapted, but an MMN is nonetheless present. Caution is warranted in accepting these conclusions, however, because our stimulus presentation paradigm may not be

optimal for assessing the Adaptation Model because the N1 is extremely minimized in all conditions. It may be that using the same paradigm with longer ISIs may elicit differential N1 adaptation in addition to P2 adaptation for more conclusive analysis. This manipulation would have the effect of minimizing the size of the MMN, however, meaning that N1 adaptation and the MMN may need to be assessed using separate paradigms, which are optimized for analyzing a specific component of interest.

From the discussion of the data thus far, there are a number of conclusions that can be drawn: First, it shows that there are neural populations that are sensitive to formant transition direction in vowels, although the evidence for this sensitivity was obtained in relation to an apparent P2, rather than the N1. Second, it shows that, as long as data are collapsed appropriately across similar conditions (see Figure 10), our stimulus presentation paradigm is capable of eliciting a pattern MMN for analysis. This may have future clinical applications if it can be optimized. Third, it points to differences in neural populations involved in the discrimination of vowels and FM tones in the A1, specifically with regards to the early P2 adaptation seen for the vowels, and more generally with the larger P2 responses for vowels than tones. These populations may be separate from those encoding raw acoustic energy of sounds (e.g., vowel space account; see Chapter 2). Fourth, it provides support for the Memory Trace account of MMN production (Näätänen et al., 2005) although it is unclear if memory traces are created by our vowels due to differences in formant-trajectory space per se, top-down categorical processes, or a combination of both. Given the fact that FM tone with equivalent spectral change did not elicit P2 adaptation or MMNs, the latter two options seem most probable.

FM Tones: All conditions (tones v. vowels)

The third hypothesis—that changes in the size of the N1 and the MMN would be similar for FM glides and vowels—was only partially supported by the present data. The size of the N1 was the same in both conditions, as predicted, but there was no evidence of neural adaptation at this level for either FM tones or vowels for comparison purposes. Evidence of neural adaptation was present, however, at the level of the P2 for vowels only. Thus, if the third hypothesis were to be rephrased to consider neural adaptation through the lens of the P2 (rather than the N1), then this hypothesis is not supported the data. This is due to the fact that there was also no MMN present for the FM tones.

Again, while this data suggests that the A1 encodes and distinguishes FM tones and vowels differently (e.g., different neural populations), there are several possible explanations for the observed differences between conditions that warrant caution in accepting this conclusion:

Stimulus amplitude, rate and duration

First, the overall amplitude of the FM tone stimuli may have been too low to elicit neural adaptation effects and pattern MMNs. This is because our tones were synthesized to match only the amplitude of F_2 in the vowel data, which did not account for the contributions of the other formants (i.e., f_0, F_1, and F_3) to the overall amplitudes of the vowel stimuli. This means that the overall amplitude of the vowel stimuli was larger than that of the tones, which may have contributed to the P2 increase, evidence of adaptation, and the MMN for vowels only. Indeed, many studies of synthesized vowels have emphasized the importance of synthesizing all stimuli to have the same overall amplitude to ensure that any significant differences between conditions are not due to differences in amplitude (Aaltonen et al., 1994; Altmann & Gaese, 2013; Korczak & Stapells, 2010). Thus, perhaps synthesizing our FM tones to match the full amplitude of the vowels would have caused them to elicit a stronger P2, with associated neural adaptation, and MMN, although this would make the comparison to the F_2 transition itself less direct. To circumvent this, perhaps the overall dB gap between the tones and vowels in our study could be filled in with low-level energy at all frequency bands (so as not to mimic specific formants) in the tones. This type of stimulus might tap into a "later auditory" or "earlier phonemic" process.

Second, the stimulus duration (120 ms) may have been too quick for the tone stimuli. Given that the sound amplitude was lower for the tones, and that they were less spectrally complex, presenting them at a slower rate may have been necessary to make them more distinguishable by the auditory system. This is unlikely, however, because tone pips as short as 30 ms with ISIs as short as 100 ms have been used to elicit reliable N1s (Wang et al., 2008).

Third, it could be that we simply required more data to find statistically significant neural adaptation, and MMNs for FM tones, especially if their distinctiveness

was less salient than that of the vowels (i.e., due to differences in spectral complexity or enhanced processing due to categorical perception; see below).

Spectral complexity

In addition to the above-mentioned factors, the effect that the spectral complexity of the vowels would have on the size of the P2 (and associated adaptation) and MMN may have been underestimated. A study by Tervaniemi and colleagues (2000) may provide some insight here. They presented pure-tones to participants (standard = 500 Hz band) and complex pure-tones (standard = 500 Hz + 1000 Hz + 1500 Hz bands) which lasted 75 ms in duration, with an ISI of 300 ms. The complex tones mimicked steady-state (no VISC) vowels in that the second and third harmonic partials were either 3 and 6 dB lower in intensity than the first component (i.e., f_0). Oddball deviants varied in their degree of frequency deviance from the standard (at all frequency bands). The authors found that the MMNs elicited by deviants were larger in amplitude and shorter in latency for the complex pure-tones than for the simple pure tones (Tervaniemi et al., 2000). Furthermore, these passive electrophysiological findings correlated well with behavioural data. Interestingly, upon further analysis, the authors found that, while the amount of frequency deviance between standards and deviants was related to the amplitude of the MMN, so too was the spectral complexity of the stimuli (these two factors did not interact). While no data was presented regarding the overall sound amplitudes of the stimuli, the authors concluded that the "spectral richness" of sounds facilitates their (pre)attentive discrimination.

Caution is called for in comparing Tervaniemi and colleagues' (2000) findings to our own, however, because a study that is more similar to ours found conflicting results. Specifically, Aaltonen and colleagues (1994) found the opposite pattern of results, showing increased MMN amplitudes for tones. This could be due to the fact that Tervaniemi and colleagues' (2000) vowel-like stimuli simply included more deviance in them, deviating at all frequency bands, whereas Aaltonen's (1994) vowels deviated only at f_0 or F_2 (never both at once). Both studies only examined "steady state" tones and vowel-equivalents, however, without examining frequency modulations at formant bands (i.e., VISC). It could be that frequency modulated bands at the formant level add an extra

layer of spectral complexity to our stimuli, which would explain why we obtained results more similar to those of Tervaniemi and colleagues (2000) despite the fact that our deviations were more similar to those of Aaltonen (1994). Our stimuli have the added bonus, however, of (likely) being natural-sounding due to having low-level energy at frequencies other than formant peaks and due to the presence of VISC.

It is unclear as to why our vowels elicited significant P2 adaptation and MMNs, whereas the FM tones did not. It could be that the spectral complexity of the vowels prevented the auditory system from habituating to our stimuli in general. As is often seen with the N1, the P2 to the less spectrally complex tones in particular may have been attenuated due to the long duration of one run of the task (e.g., over 10 minutes), especially due to the fact that participants were not actively attending to the stimuli (Näätänen & Picton, 1987). This is unlikely to be the case, however, in light of Tervaniemi and colleagues' (2000) findings, that their data for behavioural discrimination of tones and vowel-like stimuli, matched their passively elicited MMN results. Alternatively, it could be that the differences between standards and deviants for FM tones may have been too small to elicit significant results, but the spectral complexity of the vowels made them salient enough to yield significant results.

Explaining the salience of vowels

There are three key possibilities as to why vowels may be more salient than tones. First, from an "absolute" perspective, it could be that either completely different or simply extra neural populations are engaged by our spectrally rich vowels compared to less rich FM tones. More precisely, it could be the trajectories of formant transitions engage neural populations that are entirely separate from those that encode the trajectories of FM tones (i.e., unique formant-trajectory space populations; see Chapter 2). Otherwise, it could be that formant transitions and FM tones engaged similar (or the same) neural populations, while the extra spectral components of the vowels activated other neural populations. If more neurons are engaged in the processing of spectrally complex sounds, such as vowels, this of itself may make them more salient, and therefore give rise to more robust ERPs—an entirely "bottom-up" effect. (To be complete, the increased amplitude of the vowel stimuli could also account for more neurons being

activated). Indeed, our vowel encoding and discrimination appeared to be affected by a bottom-up spectral complexity given the pattern of significant results, despite the fact that the absolute spectral change between standards and deviants were equivalent for FM tones and vowels.

Second, it could be that the human auditory system is better at distinguishing spectrally complex stimuli than spectrally simple stimuli—in a more "top-down" manner, based on peoples' previous experience with spectrally complex stimuli. Indeed, Tervaniemi and colleagues concluded, "Since our auditory system has extensive experience with spectrally rich stimulation like speech and music, it appears plausible that it reacts more vigorously to changes in spectrally complex than simple (ecologically unrepresentative) information" (Tervaniemi et al., 2000, p. 32). This explanation is sensible because it is efficient for the central nervous system to become faster at processing the types of stimuli it encounters on a daily basis (such as speech, music, and other complex sensory experiences). If it were true that this "spectral complexity effect" occurs, this would go against a purely bottom-up explanation to speech perception (i.e., a bottom-up approach could be enhanced by top-down training effects).

Third, it could be that humans have ingrained, well-practiced vowel-recognition systems that allow them to process vowels more quickly than tones (even if they were matched for spectral complexity and not just absolute spectral change). This would be an "even more top-down" process than the mere spectral complexity effect, and would be especially active if participants perceived our stimuli as good exemplars of our target vowels /ɪ/ and /e/ (Iverson & Kuhl, 2000). Indeed, speech sounds are a specific type of spectrally complex acoustic stimulus that we spend every day listening to. We need to communicate well for our survival, so it makes sense that our auditory and language systems have strong neural connections that facilitate vowel identification (more so than other spectrally complex stimuli). These would involve long-term memory networks such as those involved in categorical perception (Kujala & Näätänen, 2010) and short-term memory networks such as those involved in the creation of memory traces (Tiitinen et al., 1994). It may be that categorical perception facilitates the creation of memory traces. Indeed, if this were an animal study with birds as subjects, we may have seen stronger ERPs for our FM tones instead of our vowels (Doupe & Kuhl, 1999).

The early P2 and MMN: Indicators of formant encoding and discrimination

In sum, the early P2 for vowels, along with its demonstrable refractory effects and MMN, as revealed in our study might be a window into encoding and discrimination of vowels at the single-formant level. Specifically, changes in the P2 reflect the learning of the basic units of language (vowels) at the boundary of "late auditory" and "early phonetic" processing in the auditory cortex. Measuring differences in P2 refractory effects may give us insight into differences in the specific neural populations involved in encoding variations in formant trajectories for specific vowels. Similarly, measuring changes in the pattern MMN may give us insight into vowel discrimination based on formant trajectory differences alone (as seen in VISC). Since our stimulus presentation paradigm allows us to thoroughly examine the discrimination and encoding of vowels separately, it is a potentially very useful experimental and clinical tool for further studies of speech perception.

CHAPTER 6 CONCLUSION

Key findings and contributions

In conclusion, there are several contributions the results of our study make to the literature on electrophysiology and speech perception:

4) Our stimulus presentation paradigm isolated the MMN from the N1 so that the MMN is a pure measure of participants' ability to discriminate formant transitions
5) Our paradigm also distinguished P2 adaptation as a measure of the neural encoding of formant transitions in synthetic vowels.
6) Our paradigm is a tool that may be used in further electrophysiological studies of speech perception. Such research may optimize it for specific research questions (e.g., which properties of vowels are neural populations sensitive to) and for use in the clinical setting to test speech perception in "hard-to-test" clinical populations.

Limitations and future work

There are a number of limitations to our study that may be rectified in future work. First, given the amount of between-subject variability in our data, it would have been useful to include more participants in our study for more statistical power. Indeed, similar studies typically include around 15-20 participants (e.g., Aaltonen et al., 1994; Deguchi et al., 2010).

Second, we may be able to elicit differential N1 adaptation effects using the same stimulus presentation paradigm, but with longer ISIs. Future work could involve using this modified paradigm to examine graded N1 adaptation effects only, while using our original paradigm in a separate condition to examine the MMN only. While this would increase data collection time, only half data could be collected within each paradigm if we continue to collapse data across conditions as done in the above analysis (i.e., we only needed 100 N1s and 200 MMNs, but obtained at least double that amount).

Third, we did not compare pattern MMN results to those of the oddball MMN, which might be useful for evaluating attention and N1 (and P2) adaptation effects. If there were any attention or N1 effects in either MMN, we would expect them to be more prominent in the oddball MMN, which would likely be larger in amplitude.

Fourth, we could try to elicit significant P2 adaptation and MMNs to FM tones (i.e., for comparison with vowels) by matching the overall loudness of the tones to that of the vowels (i.e., not just the F_2 loudness). This could be done either by using the same tone stimuli and increasing the overall loudness of the frequency-modulated component itself or by adding equal amounts of low-level energy at all harmonics of f_0. Another way to try to elicit significant result with tones might be to simply give them a longer ISI (which would result in a larger P1-N1-P2 complex). Alternatively, the energy in F_2 frequency range could be isolated from the vowel stimuli and used as a less spectrally complex control (instead of the FM tones).

Fifth it would be interesting to use our stimulus presentation paradigm to examine P2 adaptation and the pattern MMN to formant transitions with trajectories that are not mirror images of one another (as in /ɪ/ and /e/). To be able to assess P2 adaptation effectively, it would be useful to examine more than just two different stimuli (i.e., the cardinal vowels /i/, /u, and /a/). These could be used to create various pattern MMN sequences to examine standards and deviants with varying amounts of deviance (e.g., more and less extreme F_2 transitions, transitions of varying trajectories etc.). The greater the deviance, the larger the release from refractoriness for the P2 would be expected to be. The same would hold true for the pattern MMN if it reflects a more "bottom-up" type of discrimination.

Sixth, related to the above, our study confounds the onset-offset, onset-trajectory and onset-slope approaches (Morrison & Nearey, 2007) to explaining the mechanism of discrimination in VISC. Future studies could use our paradigm to with synthetic vowels with formants that are modified in such a way that allows us to distinguish between these three approaches. For example, "elbowed" stimuli, with elbows at various temporal locations throughout the duration of the stimulus could be used to assess the onset-offset v. onset-direction approaches (Chiddenton & Kiefte, 2013; Morrison & Nearey, 2007).

Seventh, future studies could experiment with the "naturalness continuum" of vowels using our stimulus presentation paradigm to examine how this affects the size of the P2 and pattern MMN. For example, the low-level energy in our synthetic vowels could be removed so that our stimuli were more tone-like (as in Tervaniemi and colleagues' (2000) study). Conversely, naturally produced vowels, which would be more

spectrally complex, could be recorded and used. In terms of clinical applications, this latter manipulation would be very important to examine if our paradigm is to be used as a tool to examine real-world speech perception.

Eighth, it would be interesting to include more active task requirements in a future study using our paradigm. Indeed, we did not include a behavioural discrimination component to our study to compare with our electrophysiological recordings. While this reduced the amount of time required of the participants, a behavioural discrimination task with similar results would confirm the conclusions of our study for "real-world" speech perception (Kujala, Tervaniemi, & Schröger, 2007). Also, some participants reported becoming sleepy over time, perhaps because this was a passive-listening experiment in a dark room with only a captioned movie that lasted a long time (80 minutes). Future studies may include a more engaging task, such as a Multiple Object Tracking (MOT) game, which does not specifically assess sound discrimination, but has been shown to be effective in pattern MMN studies (Sculthorpe et al., 2008).

Finally, once the above issues have been examined, including optimization of our paradigm for reliable single-subject recording, future studies could examine between-subject variation in results across a range of ages and clinical populations (Sussman, Steinschneider, Gumenyuk, Grushko, & Lawson, 2008). This would help to establish the clinical utility of the stimulus presentation paradigm we have created for determining the speech-perception abilities of "hard-to-test" populations.

References

Aaltonen, O., Eerola, O., Lang, A. H., Uusipaikka, E., & Tuomainen, J. (1994). Automatic discrimination of phonetically relevant and irrelevant vowel parameters as reflected by mismatch negativity. *The Journal of the Acoustical Society of America, 96*, 1489.

Aiken, S. J., & Picton, T. W. (2006). Envelope following responses to natural vowels. *Audiology & Neurotology, 11*(4), 213.

Aiken, S. J., & Picton, T. W. (2008). Envelope and spectral frequency-following responses to vowel sounds. *Hearing Research, 245*(1), 35. doi:10.1016/j.heares.2008.08.004

Alain, C., Campeanu, S., & Tremblay, K. (2010). Changes in sensory evoked responses coincide with rapid improvement in speech identification performance. *Journal of Cognitive Neuroscience, 22*(2), 392-403.

Alain, C., & Woods, D. L. (1997). Attention modulates auditory pattern memory as indexed by event-related brain potentials. *Psychophysiology, 34*(5), 534-546.

Alain, C., Woods, D. L., & Ogawa, K. H. (1994). Brain indices of automatic pattern processing. *Neuroreport, 6*(1), 140-144.

Altmann, C. F., & Gaese, B. H. (2013). Representation of frequency-modulated sounds in the human brain. *Hearing Research, 307,* 74-85.

Altmann, C. F., Klein, C., Heinemann, L. V., Wibral, M., Gaese, B. H., & Kaiser, J. (2011). Repetition of complex frequency-modulated sweeps enhances neuromagnetic responses in the human auditory cortex. *Hearing Research, 282*(1), 216-224.

Atienza, M., Cantero, J. L., & Stickgold, R. (2004). Posttraining sleep enhances automaticity in perceptual discrimination. *Journal of Cognitive Neuroscience, 16*(1), 53-64.

Barreda, S., & Nearey, T. M. (2012). The direct and indirect roles of fundamental frequency in vowel perception. *The Journal of the Acoustical Society of America, 131*(1), 466-477. doi:10.1121/1.3662068

Belin, P., Zilbovicius, M., Crozier, S., Thivard, L., Fontaine, A., Masure, M. C., & Samson, Y. (1998). Lateralization of speech and auditory temporal processing. *Journal of Cognitive Neuroscience, 10*(4), 536-540.

Bendixen, A., & Schröger, E. (2008). Memory trace formation for abstract auditory features and its consequences in different attentional contexts. *Biological Psychology, 78*(3), 231-241. doi:10.1016/j.biopsycho.2008.03.005

Binder, J., Frost, J., Hammeke, T., Bellgowan, P., Springer, J., Kaufman, J., & Possing, E. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cerebral Cortex, 10*(5), 512-528.

BioSemi Instrumentation. (2006). Cap 128 layout - large. Retrieved 01/27, 2014, from http://www.biosemi.com/pics/

Bouchard, K. E., Mesgarani, N., Johnson, K., & Chang, E. F. (2013). Functional organization of human sensorimotor cortex for speech articulation. *Nature, 495*(7441), 327-332.

Brechmann, A., Baumgart, F., & Scheich, H. (2002). Sound-level-dependent representation of frequency modulations in human auditory cortex: A low-noise fMRI study. *Journal of Neurophysiology, 87*(1), 423-433.

Butler, R. A. (1968). Effect of changes in stimulus frequency and intensity on habituation of the human vertex potential. *The Journal of the Acoustical Society of America, 44*, 945.

Cacace, A. T., & McFarland, D. J. (2005). The importance of modality specificity in diagnosing central auditory processing disorder. *American Journal of Audiology, 14*(2), 112-123. doi:10.1044/1059-0889(2005/012)

Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., & Näätänen, R. (1998). Development of language-specific phoneme representations in the infant brain. *Nature Neuroscience, 1*(5), 351-353.

Chiddenton, K., & Kiefte, M. (2013). Perception of vowel-inherent spectral change. Paper presented at the *Proceedings of Meetings on Acoustics, 19* 060279.

Conboy, B. T., & Kuhl, P. K. (2011). Impact of second-language experience in infancy: Brain measures of first-and second-language speech perception. *Developmental Science, 14*(2), 242-248.

Crivello, F., & Mazoyer, B. (1999). Positron emission tomography of the human brain. *Modern techniques in neuroscience research* (pp. 1083-1097) Springer.

D'Arcy, R. C., Bardouille, T., Newman, A. J., McWhinney, S. R., DeBay, D., Sadler, R. M., . . . Esser, M. J. (2012). Spatial MEG laterality maps for language: Clinical applications in epilepsy. *Human Brain Mapping,*

Davis, H., & Zerlin, S. (1966). Acoustic relations of the human vertex potential. *The Journal of the Acoustical Society of America, 39*, 109.

Deguchi, C., Chobert, J., Brunelliere, A., Nguyen, N., Colombo, L., & Besson, M. (2010). Pre-attentive and attentive processing of french vowels. *Brain Research, 1366*, 149-161. doi:10.1016/j.brainres.2010.09.104

Diekmann, V., Erné, S. N., & Becker, W. (1999). Magnetoencephalography. *Modern techniques in neuroscience research* (pp. 1025-1054) Springer.

Doupe, A. J., & Kuhl, P. K. (1999). Birdsong and human speech: Common themes and mechanisms. *Annual Review of Neuroscience, 22*(1), 567-631.

Feng, Y., Yin, S., Kiefte, M., & Wang, J. (2010). Temporal resolution in regions of normal hearing and speech perception in noise for adults with sloping high-frequency hearing loss. *Ear and Hearing, 31*(1), 115-125.

Fox, R. A., & Jacewicz, E. (2009). Cross-dialectal variation in formant dynamics of american english vowels. *The Journal of the Acoustical Society of America, 126*(5), 2603-2618. doi:10.1121/1.3212921

Frahm, J., Fransson, P., & Krüger, G. (1999). Magnetic resonance imaging of human brain function. *Modern techniques in neuroscience research* (pp. 1055-1082) Springer.

Garrido, M. I., Kilner, J. M., Stephan, K. E., & Friston, K. J. (2009). The mismatch negativity: A review of underlying mechanisms. *Clinical Neurophysiology : Official Journal of the International Federation of Clinical Neurophysiology, 120*(3), 453-463. doi:10.1016/j.clinph.2008.11.029

Griffiths, T. D., & Warren, J. D. (2002). The planum temporale as a computational hub. *Trends in Neurosciences, 25*(7), 348-353. doi:10.1016/S0166-2236(02)02191-4

Hall, D. A., Haggard, M. P., Akeroyd, M. A., Summerfield, A. Q., Palmer, A. R., Elliott, M. R., & Bowtell, R. W. (2000). Modulation and task effects in auditory processing measured using fMRI. *Human Brain Mapping, 10*(3), 107-119.

Hämäläinen, J. A., Fosker, T., Szücs, D., & Goswami, U. (2011). N1, P2 and T-complex of the auditory brain event-related potentials to tones with varying rise times in adults with and without dyslexia. *International Journal of Psychophysiology, 81*(1), 51-59. doi:http://dx.doi.org/10.1016/j.ijpsycho.2011.04.005

Hart, H. C., Palmer, A. R., & Hall, D. A. (2003). Amplitude and frequency-modulated stimuli activate common regions of human auditory cortex. *Cerebral Cortex (New York, N.Y.: 1991), 13*(7), 773-781.

Herrmann, B., Henry, M. J., & Obleser, J. (2013). Frequency-specific adaptation in human auditory cortex depends on the spectral variance in the acoustic stimulation. *Journal of Neurophysiology, 109*(8), 2086-2096.

Herrmann, B., Henry, M. J., Scharinger, M., & Obleser, J. (2013). Auditory filter width affects response magnitude but not frequency specificity in auditory cortex. *Hearing Research, 304,* 128-136.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of american english vowels. *The Journal of the Acoustical Society of America, 97*(5 Pt 1), 3099-3111.

Hillenbrand, J. M., & Nearey, T. M. (1999). Identification of resynthesized /hVd/ utterances: Effects of formant contour. *The Journal of the Acoustical Society of America, 105*(6), 3509-3523.

Holt, L. L. (2006). Speech categorization in context: Joint effects of nonspeech and speech precursors. *The Journal of the Acoustical Society of America, 119*, 4016.

Hsieh, I. H., Fillmore, P., Rong, F., Hickok, G., & Saberi, K. (2012). FM-selective networks in human auditory cortex revealed using fMRI and multivariate pattern classification. *Journal of Cognitive Neuroscience, 24*(9), 1896-1907. doi:10.1162/jocn_a_00254; 10.1162/jocn_a_00254

Humes, L. E., Dubno, J. R., Gordon-Salant, S., Lister, J. J., Cacace, A. T., Cruickshanks, K. J., . . . Wingfield, A. (2012). Central presbycusis: A review and evaluation of the evidence. *Journal of the American Academy of Audiology, 23*(8), 635-666.

Iverson, P., & Kuhl, P. K. (2000). Perceptual magnet and phoneme boundary effects in speech perception: Do they arise from a common mechanism? *Perception & Psychophysics, 62*(4), 874-886.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition, 87*(1), B47-B57.

Jacewicz, E., & Fox, R. A. (2012). The effects of cross-generational and cross-dialectal variation on vowel identification and classification. *The Journal of the Acoustical Society of America, 131*(2), 1413-1433. doi:10.1121/1.3676603

Jacobsen, T., & Schröger, E. (2001). Is there pre-attentive memory-based comparison of pitch? *Psychophysiology, 38*(4), 723-727.

Jin, S., & Liu, C. (2013). The vowel inherent spectral change of english vowels spoken by native and non-native speakers. *The Journal of the Acoustical Society of America, 133*(5), EL363-EL369.

Kewley-Port, D. (1982). Measurement of formant transitions in naturally produced stop consonant–vowel syllables. *The Journal of the Acoustical Society of America, 72*, 379.

Kiefte, M., & Kluender, K. R. (2005). The relative importance of spectral tilt in monophthongs and diphthongs. *The Journal of the Acoustical Society of America, 117*(3 Pt 1), 1395-1404.

Kiefte, M., & Kluender, K. R. (2008). Absorption of reliable spectral characteristics in auditory perception. *The Journal of the Acoustical Society of America, 123*(1), 366-376. doi:10.1121/1.2804951

Kim, H., Hasegawa-Johnson, M., & Perlman, A. (2011). Vowel contrast and speech intelligibility in dysarthria. *Folia Phoniatrica Et Logopaedica : Official Organ of the International Association of Logopedics and Phoniatrics (IALP), 63*(4), 187-194. doi:10.1159/000318881

Klatt, D. H. (1980). Software for a cascade/parallel formant synthesizer. *The Journal of the Acoustical Society of America, 67*(3), 971-995.

Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *The Journal of the Acoustical Society of America, 87*(2), 820-857.

Korczak, P. A., & Stapells, D. R. (2010). Effects of various articulatory features of speech on cortical event-related potentials and behavioral measures of speech-sound processing. *Ear and Hearing, 31*(4), 491-504. doi:10.1097/AUD.0b013e3181d8683d

Kraus, N., McGee, T. J., Carrell, T. D., Zecker, S. G., Nicol, T. G., & Koch, D. B. (1996). Auditory neurophysiologic responses and discrimination deficits in children with learning problems. *Science, 273*(5277), 971-973.

Kraus, N., McGee, T., Sharma, A., Carrell, T., & Nicol, T. (1992). Mismatch negativity event-related potential elicited by speech stimuli. *Ear and Hearing, 13*(3), 158-164.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science, 255*(5044), 606-608.

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience, 5*(11), 831-841. doi:10.1038/nrn1533

Kujala, T., Tervaniemi, M., & Schröger, E. (2007). The mismatch negativity in cognitive and clinical neuroscience: Theoretical and methodological considerations. *Biological Psychology, 74*(1), 1-19.

Kujala, T., & Näätänen, R. (2010). The adaptive brain: A neurophysiological perspective. *Progress in Neurobiology, 91*(1), 55-67. doi:10.1016/j.pneurobio.2010.01.006

Leung, A. W., He, Y., Grady, C. L., & Alain, C. (2013). Age differences in the neuroelectric adaptation to meaningful sounds. *PloS One, 8*(7), e68892.

Lidji, P., Jolicœur, P., Moreau, P., Kolinsky, R., & Peretz, I. (2009). Integrated preattentive processing of vowel and pitch. *Annals of the New York Academy of Sciences, 1169*(1), 481-484.

Liégeois-Chauvel, C., de Graaf, J. B., Laguitton, V., & Chauvel, P. (1999). Specialization of left auditory cortex for speech perception in man depends on temporal coding. *Cerebral Cortex, 9*(5), 484-496.

Luck, S. J. (2005). *An introduction to the event-related potential technique*. Cambridge, MA: MIT press.

Maiste, A. C., Wiens, A. S., Hunt, M. J., Scherg, M., & Picton, T. W. (1995). Event-related potentials and the categorical perception of speech sounds. *Ear and Hearing, 16*(1), 68-89.

Maiste, A., & Picton, T. (1989). Human auditory evoked potentials to frequency-modulated tones. *Ear and Hearing, 10*(3), 153-160.

Mäkelä, J., Hari, R., & Linnankivi, A. (1987). Different analysis of frequency and amplitude modulations of a continuous tone in the human auditory cortex: A neuromagnetic study. *Hearing Research, 27*(3), 257-264.

Mann, V. A. (1980). Influence of preceding liquid on stop-consonant perception. *Perception & Psychophysics, 28*(5), 407-412.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. *Journal of Neuroscience Methods, 164*(1), 177-190.

Martin, B. A., Kurtzberg, D., & Stapells, D. R. (1999). The effects of decreased audibility produced by high-pass noise masking on N1 and the mismatch negativity to speech sounds/ba/and/da. *Journal of Speech, Language and Hearing Research, 42*(2), 271.

Martin, B. A., Tremblay, K. L., & Korczak, P. (2008). Speech evoked potentials: From the laboratory to the clinic. *Ear and Hearing, 29*(3), 285-313. doi:10.1097/AUD.0b013e3181662c0e

Martinez-Montes, E., Valdes-Sosa, P. A., Miwakeichi, F., Goldman, R. I., & Cohen, M. S. (2004). Concurrent EEG/fMRI analysis by multiway partial least squares. *NeuroImage, 22*(3), 1023-1034.

May, P. J., & Tiitinen, H. (2010). Mismatch negativity (MMN), the deviance-elicited auditory deflection, explained. *Psychophysiology, 47*(1), 66-122. doi:10.1111/j.1469-8986.2009.00856.x

Medwetsky, L. (2011). Spoken language processing model: Bridging auditory and language processing to guide assessment and intervention. *Language, Speech, and Hearing Services in Schools, 42*(3), 286-296. doi:10.1044/0161-1461(2011/09-0090)

Millman, R. E., Prendergast, G., Kitterick, P. T., Woods, W. P., & Green, G. G. (2010). Spatiotemporal reconstruction of the auditory steady-state response to frequency modulation using magnetoencephalography. *NeuroImage, 49*(1), 745-758. doi:10.1016/j.neuroimage.2009.08.029; 10.1016/j.neuroimage.2009.08.029

Moore, B. C., & Sek, A. (1995). Effects of carrier frequency, modulation rate, and modulation waveform on the detection of modulation and the discrimination of modulation type (amplitude modulation versus frequency modulation). *The Journal of the Acoustical Society of America, 97*(4), 2468-2478.

Morrison, G. S., & Nearey, T. M. (2007). Testing theories of vowel inherent spectral change. *The Journal of the Acoustical Society of America, 122*(1), EL15-22. doi:10.1121/1.2739111

Näätänen, R., Sams, M., Alho, K., Paavilainen, P., Reinikainen, K., & Sokolov, E. (1988). Frequency and location specificify of the human vertex N1 wave. *Electroencephalography and Clinical Neurophysiology, 69*(6), 523-531.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., . . . Luuk, A. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature, 385*, 432-434.

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology, 118*(12), 2544-2590.

Näätänen, R., Pakarinen, S., Rinne, T., & Takegata, R. (2004). The mismatch negativity (MMN): Towards the optimal paradigm. *Clinical Neurophysiology, 115*(1), 140-144.

Näätänen, R., Jacobsen, T., & Winkler, I. (2005). Memory-based or afferent processes in mismatch negativity (MMN): A review of the evidence. *Psychophysiology, 42*(1), 25-32. doi:10.1111/j.1469-8986.2005.00256.x

Näätänen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology, 24*(4), 375-425.

Neel, A. T. (2008). Vowel space characteristics and vowel identification accuracy. *Journal of Speech, Language and Hearing Research, 51*(3), 574.

Obleser, J., Elbert, T., Lahiri, A., & Eulitz, C. (2003). Cortical representation of vowels reflects acoustic dissimilarity determined by formant frequencies. *Cognitive Brain Research, 15*(3), 207-213.

Pardo, P. J., & Sams, M. (1993). Human auditory cortex responses to rising versus falling glides. *Neuroscience Letters, 159*(1-2), 43-45.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America, 24*, 175.

Picton, T. W., Campbell, K. B., Baribeau-Braun, J., & Proulx, G. B. (1978). The neurophysiology of human attention: A tutorial review. *Attention and Performance VII, 6*, 429-467.

Picton, T. W., Skinner, C. R., Champagne, S. C., Kellett, A. J., & Maiste, A. C. (1987). Potentials evoked by the sinusoidal modulation of the amplitude or frequency of a tone. *The Journal of the Acoustical Society of America, 82*(1), 165-178.

Poeppel, D., Guillemin, A., Thompson, J., Fritz, J., Bavelier, D., & Braun, A. R. (2004). Auditory lexical decision, categorical perception, and FM direction discrimination differentially engage left and right auditory cortex. *Neuropsychologia, 42*(2), 183-200.

Poeppel, D., Idsardi, W. J., & van Wassenhove, V. (2008). Speech perception at the interface of neurobiology and linguistics. *Philosophical Transactions of the Royal Society B: Biological Sciences, 363*(1493), 1071-1086.

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

Sams, M., Hari, R., Rif, J., & Knuutila, J. (1993). The human auditory sensory memory trace persists about 10 sec: Neuromagnetic evidence. *Journal of Cognitive Neuroscience, 5*(3), 363-370. doi:10.1162/jocn.1993.5.3.363; 10.1162/jocn.1993.5.3.363

Sams, M., & Näätänen, R. (1991). Neuromagnetic responses of the human auditory cortex to short frequency glides. *Neuroscience Letters, 121*(1-2), 43-46.

Schonwiesner, M., Rubsamen, R., & von Cramon, D. Y. (2005). Hemispheric asymmetry for spectral and temporal processing in the human antero-lateral auditory belt cortex. *The European Journal of Neuroscience, 22*(6), 1521-1528. doi:10.1111/j.1460-9568.2005.04315.x

Sculthorpe, L. D., Collin, C. A., & Campbell, K. B. (2008). The influence of strongly focused visual attention on the detection of change in an auditory pattern. *Brain Research, 1234*, 78-86.

Seubert, C. N., & Herman, M. (2012). Auditory evoked potentials. *Monitoring the nervous system for anesthesiologists and other health care professionals* (pp. 47-68) Springer.

Shamma, S. (2001). On the role of space and time in auditory processing. *Trends in Cognitive Sciences, 5*(8), 340-348.

Shu, Z. J., Swindale, N. V., & Cynader, M. S. (1993). Spectral motion produces an auditory after-effect. *Nature, 364*(6439), 721-723. doi:10.1038/364721a0

Skoe, E., & Kraus, N. (2010). Auditory brainstem response to complex sounds: A tutorial. *Ear and Hearing, 31*(3), 302.

Stilp, C., Alexander, J., Kiefte, M., & Kluender, K. (2010). Auditory color constancy: Calibration to reliable spectral properties across nonspeech context and targets. *Attention, Perception and Psychophysics, 72*(2), 470. doi:10.3758/APP.72.2.470

Stoodley, C. J., Hill, P. R., Stein, J. F., & Bishop, D. V. M. (2006). Auditory event-related potentials differ in dyslexics even when auditory psychophysical performance is normal. *Brain Research, 1121*(1), 190-199. doi:http://dx.doi.org/10.1016/j.brainres.2006.08.095

Sussman, E., Steinschneider, M., Gumenyuk, V., Grushko, J., & Lawson, K. (2008). The maturation of human evoked brain potentials to sounds presented at different stimulus rates. *Hearing Research, 236*(1–2), 61-79. doi:http://dx.doi.org/10.1016/j.heares.2007.12.001

Sussman, E. S., & Gumenyuk, V. (2005). Organization of sequential sounds in auditory memory. *Neuroreport, 16*(13), 1519-1523.

Tervaniemi, M., Ilvonen, T., Sinkkonen, J., Kujala, A., Alho, K., Huotilainen, M., & Näätänen, R. (2000). Harmonic partials facilitate pitch discrimination in humans: Electrophysiological and behavioral evidence. *Neuroscience Letters, 279*(1), 29-32.

Tiitinen, H., May, P., Reinikainen, K., & Näätänen, R. (1994). Attentive novelty detection in humans is governed by pre-attentive sensory memory. *Nature, 372*(6501), 90-92. doi:10.1038/372090a0

Tonnquist-Uhlen, I., Ponton, C. W., Eggermont, J. J., Kwong, B., & Don, M. (2003). Maturation of human central auditory system activity: The T-complex. *Clinical Neurophysiology, 114*(4), 685-701. doi:http://dx.doi.org/10.1016/S1388-2457(03)00005-1

Tremblay, K. L., Billings, C., & Rohila, N. (2004). Speech evoked cortical potentials: Effects of age and stimulus presentation rate. *Journal of the American Academy of Audiology, 15*(3), 226-237.

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

Official Journal of the International Federation of Clinical Neurophysiology, 120(1), 128-135. doi:10.1016/j.clinph.2008.10.005

Wallach, G. P. (2011). Peeling the onion of auditory processing disorder: A language/curricular-based perspective. *Language, Speech, and Hearing Services in Schools, 42*(3), 273-285. doi:10.1044/0161-1461(2011/09-0090)

Wang, A. L., Mouraux, A., Liang, M., & Iannetti, G. D. (2008). The enhancement of the N1 wave elicited by sensory stimuli presented at very short inter-stimulus intervals is a general feature across sensory systems. *PLoS One, 3*(12), e3929.

Werker, J. F., & Yeung, H. H. (2005). Infant speech perception bootstraps word learning. *Trends in Cognitive Sciences, 9*(11), 519-527.

Wilson, W. J., Arnott, W., & Henning, C. (2013). A systematic review of electrophysiological outcomes following auditory training in school-age children with auditory processing deficits. *International Journal of Audiology, 52*(11), 721-730.

Winkler, I., Kujala, T., Tiitinen, H., Sivonen, P., Alku, P., Lehtokoski, A., . . . Näätänen, R. (1999). Brain responses reveal the learning of foreign language phonemes. *Psychophysiology, 36*(5), 638-642.

Zhang, F., Hammer, T., Banks, H., Benson, C., Xiang, J., & Fu, Q. (2011). Mismatch negativity and adaptation measures of the late auditory evoked potential in cochlear implant users. *Hearing Research, 275*(1), 17-29.

APPENDIX A: Consent Form

School of Human Communication Disorders Consent Form

Neuronal adaptation and formant transition direction in vowels: An MMN study.

Primary Contact
Nathanael Crawford
M.Sc. Human Communication Disorders (Candidate)
Nathanael.crawford@dal.ca

Secondary Contact
Dr. Steven Aiken
Assistant Professor
School of Human Communication Disorders, Faculty of Health Professions
Department of Surgery, Faculty of Medicine
Department of Psychology, Faculty of Science Dalhousie University
6th Floor • 1256 Barrington Street • Halifax • NS • B3J 1Y6 • Canada
P.(902) 494-1057
F.(902) 494-5151
steve.aiken@dal.ca

Introduction

We invite you to take part in a research study being conducted by Nathanael Crawford and Dr. Steve Aiken. Your participation in this study is voluntary and you may withdraw from the study at any time. Your academic status will not be affected by whether or not you participate. The study is described below. This description tells you about the risks, inconvenience, or discomfort that you might experience. Participating in the study might not benefit you, but we might learn things that will benefit others. You should discuss any questions you have about this study with Nathanael or Dr. Aiken.

Purpose of the Study

There are several reasons why a person may misunderstand the speech of someone else. One of the many possible reasons is a person's difficulty in perceiving speech. Many clinicians working with people who have communication disorders test speech perception by collecting "yes-no" responses from them. However, speech perception is complex and may break down for a number of reasons, which cannot be easily assessed using 'yes-no" tests. The distinction between speech perception and speech understanding *can* be identified using an electroencephalogram (EEG), to record brainwaves specifically related to speech perception (Martin, Tremblay & Korczak, 2008). One such brainwave, known as the Mismatch Negativity (MMN), can tell us about the brain's response to sound and how well the brain is healing after being damaged (Kujala & Näätänen, 2010). More research is needed, however, before clinicians can use this technique to create therapy plans for their clients. In our study, we will examine whether we can measure predictable changes in the MMN when different vowels and tones are presented (e.g., /ɪ/ as in 'wish' and /e/ as in 'day'). The results of this study may be very useful for clinical assessment of speech encoding in hard-to-test populations (e.g., infants wearing hearing aids).

Study Design

This study is composed of two almost identical experiments, which each take approximately 37 minutes and are completed in succession. Each experiment has the same two components. The first involves watching a silent movie of your choice with subtitles throughout both experiments. The second involves listening to a series of sounds that either ascend or descend in pitch while watching the movie. In the first experiment, the sounds will be pure tones, while in the second experiment the sounds will be vowels. We will record changes in you brainwaves elicited by these sounds using an electroencephalogram.

Who can Participate

You may participate in this study if you are between 18-50 years of age, have no neurological impairments, and you have normal vision, motor abilities and hearing. Your hearing will be measured before we begin the experiment to ensure that it is at an appropriate level for you to participate.

Who will be Conducting the Research

Nathanael Crawford and Dr. Steve Aiken will be conducting the study. Nathanael will be present at all times during data acquisition.

What You will be Asked to Do

The study requires one visit to the School of Human Communication Disorders at Dalhousie University. The entire session will require approximately 140 minutes of your time. Upon arrival, your hearing will be tested prior to beginning the study. During the study, we will record your brain waves in response to various sound stimuli presented at a comfortable level. You will be asked to wear earphones that sit in the ear canal (similar to music player earphones), and a number of electrodes will be attached to your face with stickers. The electrodes will be placed on you temple, forehead and nose. A 140-electrode cap will then be placed on your head and water-soluble gel will be squirted gently underneath these electrodes, on top of the skin, using a blunt syringe. This preparation period will take between 15 and 30 minutes. (This preparation time can be minimized if you come to the experiment with clean hair—preferably no hair gel or products). During the study, you will be asked to sit upright in a comfortable chair, while listening to the sounds and watching a silent, subtitled movie of your choice. You will be given an opportunity to rest between sessions if you wish. After the experiment, you will be given the opportunity to wash your hair if you choose, or you may wish to wipe off the (water soluble) adhesive gel with a towel.

Possible Risks and Discomfort

There is minimal risk involved with this study. There may be temporary red marks on the skin remaining after the electrodes and the adhesive are removed, and the insert earphones may be slightly itchy or uncomfortable. Also, there may be some gel left in your hair, which you can easily wash out at our sink after the experiment has been completed. Also, there is a minimal chance that boredom will be experienced.

Possible Benefits

A possible benefit to you of participating in this study is knowing the results of your hearing screening. In general, the main benefit to your participation is contributing to the increase in understanding of how the brain encodes and discriminates speech sounds.

Compensation/Reimbursement

Your participation is strictly on a volunteer basis, and we thank you for your time. There will, therefore, be no monetary compensation.

Confidentiality and Anonymity

Though it is not possible for you to remain entirely anonymous, since you will be known to the researcher, you will not be identified in any reports and publications. The data that you are providing will be treated and stored in a confidential manner. The data will be coded using numbers and not names, and it will be stored on a password-protected computer. The only individuals with access to the computer will be the two investigators. The data will be securely stored for at least five years in the password-protected computer. You will not be provided with your individual results, but are invited to read the group averaged results of this study when published in an academic journal or to attend Nathanael Crawford's thesis defense to learn more about them. You may also provide your email address on the following page if you would prefer a summary of the results to be sent to you instead.

Questions

If there are any questions with regards to your participation in the study, please do not hesitate to contact Nathanael Crawford by email at nathanael.crawford@dal.ca. You will be provided with any new information that may affect your decision to participate in the study as soon as it becomes available.

Problems or Concerns

If you have any difficulties with, or wish to voice concern about, any aspect of your participation in this study, you may contact Catherine Connors, Director of Dalhousie University's Office of Human Research Ethics Administration, for assistance (902) 494-1462, catherine.connors@dal.ca.

School of Human Communication Disorders Consent Form

Neuronal adaptation and formant transition direction in vowels: An MMN study.

I have read the explanation about this study. I have been given the opportunity to discuss it and my questions have been answered to my satisfaction.

I hereby consent to take part in this study. However, I realize that my participation is voluntary and that I am free to withdraw from the study at any time.

_____	_____	_____
Participant Signature	Participant Name	Date

*Optional: Participant email address: _____

_____	_____	_____
Researcher Signature	Researcher Name	Date

APPENDIX B: Recruitment Dialogue and Inclusion Criteria Checklist

Neuronal adaptation and formant transition direction in vowels: An MMN study.

Study description
You are invited to participate in a speech perception study conducted by Nathanael Crawford at Dalhousie University's School of Human Communication Disorders. This study is composed of two almost identical experiments, which each take approximately 37 minutes and are completed in the same session. Each experiment has the same two components. The first involves watching a silent, subtitled movie of your choice. The second involves listening to a series of sounds while you watch the movie. We will record changes in your brainwaves using an electroencephalogram while you watch the movie. We want to see how your brainwaves are affected by the sounds you are hearing. The results of this study may help in the creation of a tool for measuring speech perception in "hard-to-test" populations (e.g., infants). Participation is on a volunteer basis and the total time commitment is approximately 140 minutes.

Who can participate?
You may participate in this study if you are between 18-50 years of age, have no known neurological impairments and you have normal (or corrected-to-normal) vision and hearing. Your hearing will be screened before we begin the experiment to ensure that it is at an appropriate level for you to participate.

Interested?
If you have any questions or would like to participate, please complete the checklist below and return it to Nathanael Crawford in person or at nathanael.crawford@dal.ca. You will be contacted as soon as possible, and if you meet all the criteria, a date can be agreed upon for you to participate in the study.

Study Participation Checklist

Name: _____ Email: _____

Check the following if it applies to you:

__ I am between 18-50 years of age
__ I have no known neurological impairments
__ I have normal vision or corrected-to-normal vision
__ I have normal hearing (no hearing aids)

Comments: _____

_____