THE NATURE OF SPEECH REPRESENTATION IN VARYING-STANDARD MMN PARADIGM

by

Chao Han

A dissertation submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Linguistics

2023

THE NATURE OF SPEECH REPRESENTATION IN VARYING-STANDARD MMN PARADIGM

by

Chao Han

Approved: _____
Benjamin Bruening, Ph.D.
Chair of the Department of Linguistics and Cognitive Science

Approved: _____
John A. Pelesko, Ph.D.
Dean of the College of Arts and Sciences

Approved: _____
Louis F. Rossi, Ph.D.
Vice Provost for Graduate and Professional Education and
Dean of the Graduate College

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Arild Hestvik, Ph.D
Professor in charge of dissertation

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Zhenghan Qi, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Irene Vogel, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

William Idsardi, Ph.D.
Member of dissertation committee

ACKNOWLEDGMENTS

I want to express my deepest gratitude to my advisor, Arild Hestvik, for his unwavering guidance and support throughout my Ph.D. life and my dissertation. I would like to particularly thank him for his tolerance and understanding regarding the delay I faced during the completion of this dissertation.

I also want to give a special thank you to my committee member, William Idsardi, for his generosity in giving his time and invaluable input throughout the process. His knowledge and expertise have been crucial in shaping the outcome of this dissertation.

I also want to extend my gratitude to my committee members Irene Vogel and Zhenghan Qi for their advice and mental support during my Ph.D. journey and their feedback on my dissertation.

I want to thank my lab buddies Ryan Rhodes and Enes Avcu. I had so much fun when we three worked together. The first experiment of my dissertation is developed on Ryan's idea. Enes gave me much information and support during my job haunting.

In addition, I want to thank my friends Chen Zhou and Sree Sarkar. I thank Chen for her companionship and encouragement, which made my days brighter. I thank Sree for having lunch together, a great way to take a break from the research.

Finally, I would like to thank my family for taking care of my son while I was away. My dissertation would not have been possible without their sacrifice and dedication.

TABLE OF CONTENTS

Appendix

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Mismatch Negativity (MMN) studies have utilized the varying-standard oddball paradigm to investigate the effect of linguistic category on speech perception. But the evidence is missing for an MMN driven by a within-category acoustic contrast between standards and deviants. With three experiments, the current dissertation asks whether the memory trace in the varying-standard paradigms retains gradient information that could lead to an MMN response.

Experiment 1 looked for an MMN response using the varying-standard paradigm without a categorical contrast between standards and deviants. The standards were realized by [tæ] with voice onset times (VOTs) of 42, 48, and 55ms VOT [tæ]s, and the deviants were realized by [tæ] with a 119ms VOT. The MMN was computed as the difference in brain response to deviants minus the same stimulus in a roving-standard control condition. A within-category MMN was observed, suggesting that the memory trace in the varying-standard paradigm must contain gradient information.

Experiment 2 asked what drove the within-category MMN in Experiment 1. The observed within-category MMN could be due to the deviant VOT contrasting to the phonetic knowledge retrieved from long-term memory. Alternatively, it could be driven by a contrast between the deviant VOT and a statistical summary based on the presented standard VOTs. In that case, the MMN magnitude should be modulated by the statistical structure of the presented stimuli (Garrido, Sahani, & Dolan, 2013). Experiment 2 presented one group with a 128ms VOT [tæ] deviant embedded in a normal distribution of varying standards with a mean VOT of 64ms and a “wide”

standard deviation of 15ms. It presented a second group with the same deviant and mean standard VOTs but with a “narrow” standard deviation of 5ms. The result replicated Experiment 1 with a within-category MMN, but no difference between the two standard deviation groups was observed. The lack of difference in the MMN magnitude suggests that a statistical summary of the presented stimuli could not explain the within-category MMN obtained in Experiment 1.

As the conclusion of Experiment 2 was based on a null result, Experiment 3 was conducted to provide positive evidence for a statistical summary of the presented stimuli. Experiment 3 swapped the standards and deviants used in Experiment 2, such that the standard VOTs formed a distribution with a mean VOT of 128ms while the deviant VOT was 64ms. If the varying standards activate phonetic knowledge as a memory trace, the deviant VOT should not lead to an MMN, as the deviant VOT corresponds to the most typical phonetic realization of /tæ/. On the other hand, if the brain computes a statistical summary of standard VOTs, the acoustic difference between the standard and deviant VOT values should lead to an MMN response. The result was that the deviant VOT elicited a robust MMN, indexing the brain’s sensitivity to the statistical structure of the presented stimuli.

The results of the three experiments confirmed that a speech MMN could be driven by a within-category difference other than a categorical contrast. This finding suggests that the memory representation in the varying-standard paradigm retains gradient information along with a category representation. Furthermore, the gradient information comes from the statistical summary based on the acoustic properties of the presented stimuli.

Chapter 1

RESEARCH QUESTION OVERVIEW

Humans can effortlessly and efficiently extract regularities from the ever-changing environment. This ability, also referred to as statistical learning (Conway, 2020), is critical for understanding speech. For example, the same word could be uttered with a great deal of inter-speaker variability in acoustic cues due to genders, ages, accents, etc. However, we can still extract invariant perceptual categories regardless of the variability. A perceptual category could be a phoneme category stored in long-term memory. For example, the phoneme category /t/ in American English, if produced at the onset position of a stressed syllable, becomes an aspirated [t^h] with a VOT typically ranging from about 15ms to about 160ms (Chodroff & Wilson, 2018). The knowledge about the detailed acoustic and articulatory realizations of a phoneme category in specific environments (e.g., at the onset position of a stressed syllable) must be stored in the long-term memory, such that people can judge whether a given phonetic token is a good or typical realization of a phoneme category surfacing in the given environment. The current dissertation will refer to that knowledge as "phonetic knowledge".

Phonetic knowledge is qualitatively different from knowing whether a phonetic token belongs to some category. Categorical information is binary, as an individual exemplar either belongs to or does not belong to a category. In contrast, the content of phonetic knowledge must contain detailed information that is gradient. A speech perception process involves mapping from gradient acoustic signals to binary category

representations, mediated by phonetic knowledge. Therefore, both categorical and gradient information is utilized during speech perception. Previous neuroimaging studies have identified the neuronal populations respectively responding to the categorical and gradient information during speech perception. For example, Blumstein, Myers, and Rissman (2005) conducted an fMRI study and found that the left inferior frontal sulcus showed a sensitivity only to an across-category VOT difference but not to a within-category VOT difference. In contrast, the left superior temporal regions showed a more graded sensitivity to both the across-category and within-category VOT differences. Going for a higher spatial resolution, the human intracranial recordings from the auditory cortex have shown that speech perception involves a processing path starting with a gradient acoustic discrimination and gradually becoming more categorical in distinct neuronal populations (Chang et al., 2010; Fox, Leonard, Sjerps, & Chang, 2020).

However, there is a lack of evidence that the brain pre-attentively retains the gradient information of speech along with the category representation. The imaging techniques used in the above studies have limited temporal resolution and thus cannot answer when gradient information is retained relative to categorical representation. With a higher temporal resolution, electroencephalogram (EEG) techniques have been used to investigate speech perception. Researchers typically examined the Mismatch Negativity (MMN) component for information encoded and retained during speech perception. The MMN is an event-related potential (ERP) component reflecting the brain's pre-attentive novelty detection process, which is typically driven by a clash between infrequent acoustic stimuli (i.e., deviants) and the memory trace of frequent acoustic stimuli (i.e., standards) (R. Näätänen et al., 2012). An MMN response

indicates the brain's sensitivity to the difference between standards and deviants and thus reveals what information about standards and deviants is encoded and retained in the memory trace. The standards can be realized by a fixed stimulus, corresponding to a single-standard paradigm, or by stimuli that differ in acoustic properties, corresponding to a varying-standard paradigm. According to Phillips et al. (2000), when the stimuli are different phonetic realizations of the same phoneme categories, the brain "has access to representations of discrete phonological categories (p. 1051)". However, as was mentioned earlier, speech perception involves not only the retrieval of a category representation but also the sensitivity to gradient acoustic signals and the retrieval of gradient information about phonetic knowledge. Therefore, the memory trace is generated in the varying-standard paradigm should contain both categorical information and gradient information, and we should be able to observe an MMN driven by an across-category contrast (i.e., difference in categorical information) as well an MMN driven by a within-category contrast (i.e., difference in gradient information). Surprisingly, MMN studies using the varying-standard paradigm (see 2.2 for details) always observed the former type of MMN but not the latter. Therefore, evidence is still missing that the memory trace generated in the varying-standard MMN paradigm retains gradient information along with the category (but see Toscano et al. (2010) for the perceptual encoding of gradient information in a different paradigm).

Note that a within-category MMN is not equal to whether listeners are sensitive to a within-category difference, which has been well documented. Early studies by Rosch and colleagues have found that not all members of a category were treated equally (Mervis & Rosch, 1981; Rosch, 1975). For a given category (e.g.,

bird), some members are judged as good or typical examples (e.g., *robin*) while others as poor or atypical examples (e.g., *penguin*). Relative to the atypical members, the typical members can be more easily judged as belonging to a category, a phenomenon called the typicality effect (McCloskey & Glucksberg, 1978; Posner & Keele, 1968; Rosch, 1975; Rosch, 1973). In speech perception, we also see a trajectory of research focus going from a coarse discrimination of categorical difference to a more fine-grained sensitivity to within-category gradience. Early studies on speech perception demonstrated the categorical perception where listeners showed better discrimination for an across-category sound pair than for a within-category sound pair (Liberman, Harris, Hoffman, & Griffith, 1957). The categorical perception does not necessarily mean listeners are not sensitive to fine-grained acoustic differences. Discrimination tasks have shown that listeners could better discriminate a sound pair with a larger acoustic difference than a smaller one, although both pairs involved an across-category difference (/ba – pa/) (Pisoni & Tash, 1974). Later studies found that the sensitivity to acoustic details also affects word recognition. The evidence comes from the acoustic variation's influence on the priming effect's magnitude in a lexical decision task. The magnitude of semantic priming (e.g., prime: *king*; target: *queen*) decreased as the prime-initial [k] changed from carrying a typically long VOT (e.g., *[k]ing*) to carrying an atypically short VOT (e.g., *[g]ing*) (Andruski, Blumstein, & Burton, 1994). Besides the behavioral measures, studies using the eye-tracking technique also found that when participants heard a target sound (e.g., [b]), the probability of their visual fixation to the visual target (e.g., a letter *b*) increased as the target sound became more acoustically different from the sound (e.g., [p]) corresponding to a visual competitor (e.g., a letter *p*) (McMurray, Aslin, Tanenhaus,

Spivey, & Subik, 2008; McMurray, Tanenhaus, Aslin, & Spivey, 2003). Those results suggest that speech recognition involves a probabilistic mapping from gradient acoustic signals to category representations. Moreover, a probabilistic mapping entails a comparison between the acoustic properties of a proximal speech stimulus[1] and the phonetic knowledge stored in long-term memory, both the acoustic properties and the content of phonetic knowledge being gradient by nature.

The within-category findings do not necessarily mean, however, that gradient information about speech – whether it is about proximal stimuli or phonetic knowledge – is retained as part of the memory trace in a varying-standard MMN paradigm, which taps into pre-attentive memory encoding in the absence of an explicit behavioral task. That is, the task demands of such studies either explicitly or implicitly encouraged participants to access gradient information. Specifically, the priming task (Andruski et al., 1994) required participants to recognize prime and target words; the eye-tracking tasks (McMurray et al., 2008, 2003) required participants to match speech sounds to visual stimuli. It is possible that, without the task demand on a speech recognition process, once a category representation is elicited, the relevant gradient information is discarded from the memory trace (but can be accessed when needed). A corollary of this is that we would not observe an MMN whenever standards and deviants belong to the same category, as the memory trace contains identical categorical information for all stimuli. Indeed, few studies found a within-category MMN; those that did, used a single-standard paradigm and thus can be

---

[1] By "proximal speech stimulus", we mean the presented speech stimulus a listener just heard. We chose this term to convey a contrast between a representation based on the presented stimulus and a representation retrieved from long-term memory.

interpreted as primarily tapping into nonspeech auditory perception rather than speech perception (see 2.1.2 for details). The goal of the current study aims to examine whether the memory trace in an MMN paradigm contains gradient information along with categorical information.

The current study used the MMN (see Section 2.1 for a detailed introduction) to measure the pre-attentive deviance detection response for within-category contrasts, combined with the varying-standards paradigm (see 2.2 for experiment details). The rest of the dissertation is organized as follows: Chapter 2 introduces the MMN and the different models to explain the MMN, followed by a review of the varying-standard paradigm. Chapter 3 is about an experiment examining whether the brain is sensitive to a within-category contrast with the varying-standard paradigm. Chapter 4 is about an experiment aiming to identify the source of the gradient information, testing the effect of the statistical distribution of standards. Chapter 5 is about an experiment offering decisive evidence for the source of the gradient information. Chapter 6 discusses the implications of the findings and concludes the dissertation.

Chapter 2

BACKGROUND

The current dissertation tests whether a within-category MMN is observable as the evidence for gradient information of speech being pre-attentively retained along with a category representation. To achieve this goal, I conducted electroencephalogram (EEG) experiments and focused on the event-related potential (ERP) component of MMN. The current chapter reviews the mechanisms of the MMN and shows why the MMN component is suitable for investigating this question.

2.1 MMN

2.1.1 What is MMN

The MMN is an event-related potential (ERP) component of novelty detection (Fitzgerald & Todd, 2020; R. Näätänen et al., 2012). One example of novelty detection happens when we encounter a rare sound among a sequence of repeated sounds. The unexpected, rare sound produces an enhanced negative-going ERP response compared to the ERP response produced by frequent sounds. That negative deflection of the rare sound ERP relative to the frequent sound ERP is the MMN. In an experimental setting, an MMN can be elicited using an oddball paradigm in which a sequence of frequent stimuli (i.e., standards) are interspersed with infrequency stimuli (i.e., deviants), as shown in Figure 1. There is also a visual MMN (Stefanics, Kremláček, & Czigler, 2014), but the current dissertation only discusses auditory MMN.

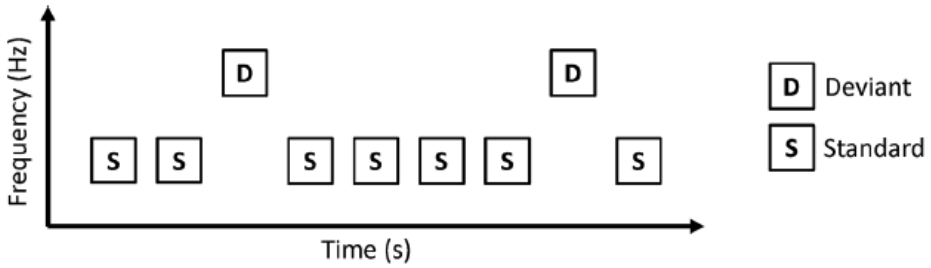


Figure 1: A classic oddball paradigm for a frequency MMN. Adapted from Kirihara et al. (2020).

The MMN reflects the brain's pre-attentive and automatic response to a detected change violating a regularity the brain extracts based on a sequence of auditory stimuli preceding the change (Näätänen, 1990). An MMN typically peaks within the 100 – 300ms time window after the onset of the detected change and exhibits a scalp topography of frontocentral distribution (Näätänen, Paavilainen, Alho, Reinikainen, & Sams, 1989). Figure 2 shows waveforms and topographical plots of an MMN effect.

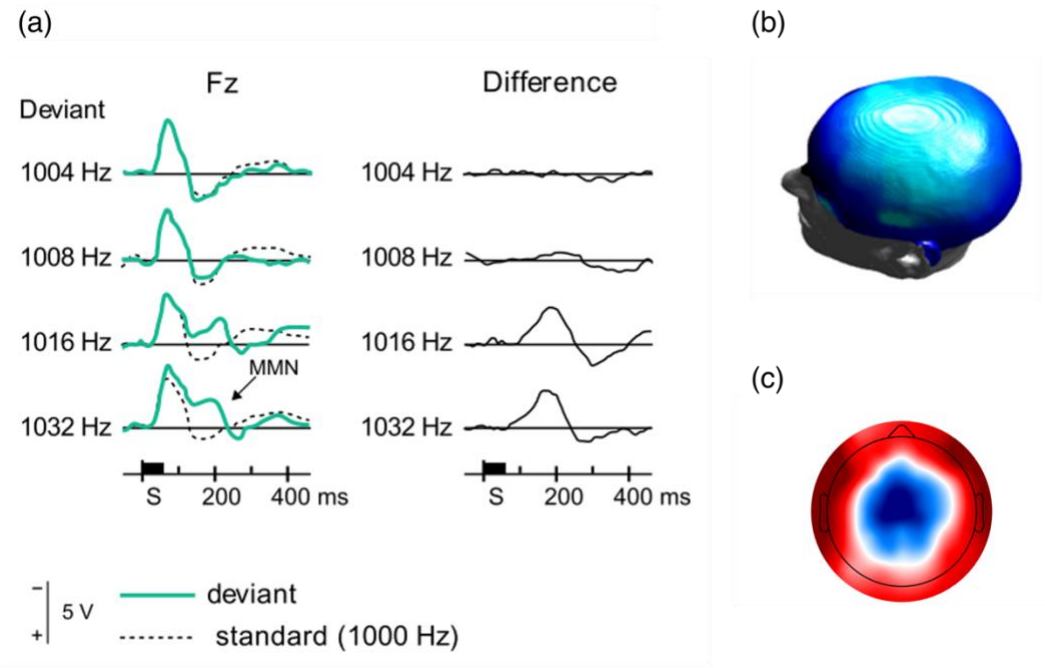


Figure 2: (a) Waveforms (negative up) of frequency MMNs. A larger frequency change yields a larger (more negative) MMN magnitude (indexed by the difference waveform). (b) A 3D topographical plot of an MMN effect. Negativity is indexed by brightness. (c) A 2D topographical plot of an MMN effect. The depth of blue indexes negativity. Adapted from Näätänen et al. (2007) and Garrido et al. (2007).

The MMN was first identified in Näätänen, Gaillard, and Mäntysalo (1978), where they found a post-N1 deflection – an N1 response is an auditorily evoked ERP component with a negative deflection peaking about 100ms after the auditory stimulus onset – responding to a change in the location of a sound source, even when participants were not paying attention to the sound source. Subsequent studies have found an MMN responding to the change in the stimulus presentation location (Schröger & Wolff, 1996; Winkler & Schröger, 1995), tone frequency (Alain, Achim, & Woods, 1999; Haenschel, Vernon, Dwivedi, Gruzelier, & Baldeweg, 2005; Horváth

et al., 2008; Schröger, 1996), tone intensity (Rinne, Särkkä, Degerman, Schröger, & Alho, 2006; Schröger, 1996), tone duration (Ylinen, Shestakova, Huotilainen, Alku, & Näätänen, 2006). The MMN was also observed in complex linguistic stimuli involving changes in consonants (Allen, Kraus, & Bradlow, 2000; Dehaene-Lambertz, 1997; Joanisse, Robertson, & Newman, 2007; Sharma, Kraus, McGee, Carrell, & Nicol, 1993) and vowels (Miglietta, Grimaldi, & Calabrese, 2013; Näätänen et al., 1997; Shestakova et al., 2002; Sittiprapaporn, Tervaniemi, Chindaduangratn, & Kotchabhakdi, 2005; Winkler et al., 1999). Since the MMN reflects the detection of a regularity violation, an MMN entails that the brain has extracted some regularity out of the presented stimuli. Regularity can be a simple repetition of some acoustic property or an abstracted pattern based on various stimuli. Studies have observed an MMN in violating complex spectrotemporal patterns (Näätänen, Schröger, Karakas, Tervaniemi, & Paavilainen, 1993; Schröger, Paavilainen, & Näätänen, 1994). For example, in Schröger, Paavilainen, and Näätänen (1994), the standard stimulus was composed of consecutive pure tone segments of different frequencies. The deviant stimulus was composed of identical tone segments but differed from the standard stimulus in ordering the tone segments. The observed MMN indicated that the brain extracted a pattern from the repetitive structure inherent in the presented acoustic signals.

There are two approaches to measuring the magnitude of an MMN. A non-identity approach involves subtracting the ERP response to the stimuli serving as standard from the ERP response to different stimuli serving as deviants. The other approach is to subtract the ERP response to the stimuli serving as standard from the ERP response to the same stimuli serving as deviants. Since the comparison is

between identical stimuli, the resulting MMN is referred to as an *identity MMN*. The non-identity MMN and the identity MMN could lead to different results, hence different conclusions (Peter, McArthur, & Thompson, 2010). To determine the right approach to use in the current study, we need to understand what part of information each approach of computing an MMN conveys. It is thus important to understand what information an MMN reveals. Below I introduce two major models of the neurophysiological and cognitive mechanisms underlying an MMN response, as well as the predictive coding framework that attempts to reconcile the two competing models.

2.1.1.1 Adaptation-based model

The adaptation-based model explained the MMN as an enhanced N1 response (May & Tiitinen, 2010). An N1 response is an auditorily evoked ERP component with a negative deflection peaking about 100ms after the auditory stimulus onset. It is generated in the auditory cortex (Näätänen & Picton, 1987) and can become attenuated as the stimulus repeats – an effect called N1 suppression. The neurophysiological basis for N1 suppression is thought to be neural adaptation – the attenuation of neural responses when the same pattern of neuronal activity is repeatedly initiated (Butler, Spreng, & Keidel, 1969). It is not hard to imagine that in a typical oddball paradigm where all standards are of the same stimulus, the same set of neuronal populations responding to the acoustic properties of that stimulus is repeatedly activated and thus becomes attenuated. When a deviant is presented, a different set of neuronal populations become activated. As the new set of neural populations has not gone through the adaptation process, the neural activity is less attenuated compared to that of standards, leading to an enhanced (more negative) N1. In an oddball paradigm

where not all the standards are of the same stimulus, the source of neural adaptation becomes less clear. It is possible that some neuronal receptive fields are broad enough to respond to all the stimuli serving as standards, and it is the repeated activation of those neurons that leads to suppression. Another possibility is that there are neurons responding to the abstracted regularities in the auditory cortex (Nelken, 2004) and these neurons undergo the adaptation process, as some neuronal populations have been found to respond to abstract features (Chang et al., 2010; Fox et al., 2020). The enhanced effect on N1 continues to the time window of what is thought to be an MMN. Thus, according to the adaptation-based model, the MMN is merely an enhanced N1 due to the release from the neural adaptation.

Note that I have used *adaptation* to describe the attenuation of neuronal responses. People have used different terms, including *habituation* (Callaway, 1973; May & Tiitinen, 2010; Rosburg et al., 2006), *refractoriness* (Jacobsen, Horenkamp, & Schröger, 2003; Jacobsen & Schröger, 2001; Jacobsen, Schröger, Horenkamp, & Winkler, 2003) and *adaptation* (O'Shea, 2015). These terms are not strictly synonyms and have different implications. *Habituation* has a clear definition, and we can use the criteria proposed by Thompson and Spencer (1966) to determine whether the N1 suppression is a result of a habituation process. According to Thompson and Spencer (1966), a habituation process must meet the following (adapted) criteria:

1. The response to standards should attenuate gradually instead of rapidly.
2. The response to the standard immediately following a deviant should be enhanced relative to the standard immediately preceding the deviant (a dishabituation process).
3. The extent of suppression is sensitive to the probability of the standard stimulus but insensitive to the interstimulus interval (ISI).

Previous studies have found that the N1 suppression occurred rapidly instead of gradually (Budd, Barry, Gordon, Rennie, & Michie, 1998; Ritter, Vaughan, & Costa, 1968), violating the first criterion. Barry et al. (1992) compared the N1 response of the standard immediately preceding a deviant to that of the standard immediately following the deviant. They found no difference between the two N1 responses, which suggests a lack of a dishabituation process, violating the second criterion. The same finding was obtained in Budd et al. (1998). Furthermore, Budd and his colleagues found that the degree of the N1 suppression was affected by the ISI, with the shortest ISI (1s) resulting in more suppression than a moderate ISI (3s), which in turn led to more suppression than the longest ISI (10s). The fact that the N1 suppression is an increasing function of the ISI also suggests that the N1 suppression is not a habituation process (Hari, Kaila, Katila, Tuomisto, & Varpula, 1982; May & Tiitinen, 2004). So far, we can conclude that a habituation process cannot explain the N1 suppression. O'Shea (2015) pointed out that the term *refractoriness* is also problematic, as the physiological meaning of *refractoriness* refers to a state where neurons hardly generate more electrical activities due to fatigue. However, to obtain a state of neuronal fatigue, we need a much faster stimulation rate (over 10 Hz, Fernández-Alfonso & Ryan, 2004) than the rate at which standards are presented in a typical MMN experiment. Also, a refractoriness state lasts for only a few milliseconds, while the MMN does not emerge until hundreds of milliseconds. Therefore, O'Shea (2015) suggested replacing *refractoriness* with *adaptation* when referring to the neurophysiological mechanism of an N1 suppression. Interpreting the N1 suppression as an *adaptation* process implies that the brain actively changes its neuronal response patterns to repeated stimulation. This is in line with the research suggesting that

MMN, as an enhanced N1, is modulated by top-down predictions (Fitzgerald & Todd, 2020) (see 2.1.1.3 for details).

One advantage of interpreting the MMN as an enhanced N1 is that it fits into the attempt to integrate human and animal research on the auditory novelty system (Escera & Malmierca, 2014). Single-unit recordings first discovered the suppression of neuronal responses to repeated stimulus in animal studies (Gross, Bender, & Rocha-Miranda, 1969; Gross, Schiller, Wells, & Gerstein, 1967; Miller & Desimone, 1994; Miller, Li, & Desimone, 1991; Miller, Li, & Desimone, 1993). Later animal studies using intracortical measurements and single-unit recordings have attempted to link this neuronal response suppression to MMN (Javit, Steinschneider, Schroeder, Vaughan, & Arezzo, 1994; Ulanovsky, Las, & Nelken, 2003). For example, Ulanovsky, Las, and Nelken (2003) found that some neurons in a cat's primary auditory cortex reduced firing with repeated stimuli. However, the responses recovered when a deviant was encountered, a phenomenon they considered to be the single-neuron correlate of the MMN. With human subjects, Elangovan et al. (2005) compared the MMN to a "difference wave". Their MMN was obtained by subtracting the frequent standard ERP from the infrequent deviant ERP; the "difference wave" was obtained by subtracting the ERP of the standard stimulus presented alone from the ERP of the deviant stimulus presented alone. They found that the MMN showed the same pattern of same scalp topography as the "difference wave" and that the N1 and P2 of the difference wave were significant predictors of the MMN amplitude.

A few challenges come up for the adaptation-based model. One challenge is the timing of the N1 suppression. Stimulus-specific repetition suppression is typically observed at 20-30ms after the stimulus onset (Ulanovsky et al., 2003), while an N1 or

MMN is typically observed after 100ms. Yet, Grimm and Escera (2012) suggested that a deviance detection process could show ERP modulations within the first 50ms after sound onset, way before the time window where an N1 or MMN is observed. Thus, a deviance detection process might start with a stimulus-specific repetition and manifests an N1 suppression in a later time window. Another challenge comes from the observations of the MMN responding to unexpected stimulus omission (e.g., Yabe, Tervaniemi, Reinikainen, & Näätänen, 1997). That is, when stimulation is absent in the position where a standard is otherwise expected, the MMN is elicited to the stimulus omission. Given the absence of stimulation, there should be no ground for an enhanced N1. Thus, the adaptation-based model would predict no MMN. In contrast, the MMN to the stimulus omission can be well explained by the memory-based model, which is introduced below.

2.1.1.2 Memory-based model

The memory-based model of MMN treated N1 and MMN as two functionally distinct components. The model suggests that an MMN emerges from the conflict between the incoming deviant and the memory trace of a standards-based representation. After being exposed to standards, the brain extracts the regularity, generates, and maintains a memory trace of that regularity (Cowan, 1984). If the new stimulus is inconsistent with the representation in the memory trace, the MMN is elicited (Alain, Woods, & Knight, 1998; Kujala, Tervaniemi, & Schröger, 2007; Näätänen, Paavilainen, Rinne, & Alho, 2007; Näätänen, 1990, 1992; Näätänen et al., 1989; Winkler, 2007). Researchers supporting the memory-based model pointed out that the adaptation-based model ignores the function of memory in the MMN formation while the MMN indeed depends on memory traces (Winkler, Reinikainen,

& Näätänen, 1993). Studies have found a functional distinction between an MMN at the mastoids and an MMN over the frontal region: while the former could be generated from the auditory cortex only, the latter could involve the frontal cortex, which is responsible for conscious expectation and detection of the change (Giard, Perrin, Pernier, & Bouchet, 1990).

The memory-based model implies a "genuine" MMN component independent of an N1 (Schröger, 1996). Indeed, a few studies have provided experimental support to dissociate the MMN and the N1 (O. Korzyukov et al., 1999; Rentzsch, Shen, Jockers-Scherübl, Gallinat, & Neuhaus, 2015; Rinne et al., 2006). For example, Rentzsch et al. (2015) found that the MMM magnitude and the N1 suppression were correlated in a healthy control group but not in a group of schizophrenia patients, suggesting that MMN cannot be explained by an N1 suppression alone. Rinne et al. (2006) found that an intensity increment elicited both an MMN and an N1 enhancement. In contrast, an intensity decrement elicited an MMN only, suggesting two separate mechanisms underlying an MMN and an N1. Korzyukov et al. (1999) conducted a source modeling of an MMN and an N1 elicited by the deviants presented alone, assuming the equivalent current dipoles. They found that although the sources of the MMN and the N1 are both located bilaterally in the superior temporal cortex, the MMN source is anterior to the N1 source, providing evidence for a neurophysiological dissociation between an MMN and an N1[2]. Another advantage of

---

[2] However, Jääskeläinen et al. (2004) suggested that the adaptation of the neuronal activity could change the center of gravity of the underlying source configuration when the source is modeled with equivalent current dipoles, which explains the source loci difference between the N1 and the MMN.

the memory-based model is that it naturally explains the MMN response to a stimulus omission. A lack of stimulus also contrasts with a representation of the regularity retained in the memory trace. An MMN is elicited whenever the regularity is violated, whether or not an actual stimulus is present.

One major criticism of the memory-based model is that it implies the existence of a separate generator exclusive to the MMN. That is, there should be some neuronal cells responding to a stimulus when it serves as a deviant but not when it serves as a standard, or some neuronal cells responding to a stimulus when it serves as a standard but not when it serves as deviant. However, with all the single-neuron level studies, such exclusive neuron cells are never found (May & Tiitinen, 2010). However, the criticism might not be legitimate because the memory-based model is a psychological explanation but not a physiological one. In contrast, the adaptation-based model is a physiological explanation. In that sense, the adaptation model and the memory model are, by nature, not mutually exclusive. This is especially true when we consider the studies that found the N1 suppression itself can be modulated by a top-down prediction (hence must involve memory encoding and retrieval) – a critical property for a "genuine MMN" (see 2.1.1.3 for the evidence). Those studies placed a novelty detection process in the Predicative Coding framework, which is introduced below.

2.1.1.3 From memory to prediction

The memory-based model implies a prediction about the upcoming stimulus. That is, the brain generates a prediction about the upcoming stimulus based on the memory trace. If the new stimulus violates the prediction, an MMN is elicited. Therefore, the memory-based model naturally fits in the Predictive coding framework. The predictive coding framework explains how a biological system performs

perceptual learning from the environment and makes an inference (Auksztulewicz & Friston, 2016). In a seminal paper, Rao and Ballard (1999) proposed a model for visual perception. They proposed that different brain regions for visual processing constitute a hierarchically organized pathway, starting from the retina to the lateral geniculate nucleus and further to the visual cortex. In this model, every brain region is a layer in the hierarchical organization. An upper layer predicts the neuronal activity of a lower layer, and a lower layer sends the prediction error to the upper layer. A prediction error occurs when, at a given layer, the prediction sent from the higher layer fails to match the incoming signal. Heilbron and Chait (2018) evaluated the key assumptions of the Predicative Coding Theory in the auditory modality. They found that the existing animal, human, and modeling studies well support the neuronal responses shaped by the hierarchically organized predictions with increasing levels of abstraction.

Following the predictive coding view, an MMN is driven by a prediction error during the hierarchical inference of the auditory system (Escera & Malmierca, 2014; Garrido, Kilner, Stephan, & Friston, 2009). The information flow of an auditory event in the brain contains a bottom-up route where the lower-layer auditory information and the prediction error are fed into the upper layer. The upper level, in turn, sends down the prediction about the upcoming auditory stimulation via a top-down route. As shown in Figure 3. An MMN reflects a process of updating the representation of the previously extracted regularity based on the standards (István Winkler, 2007),

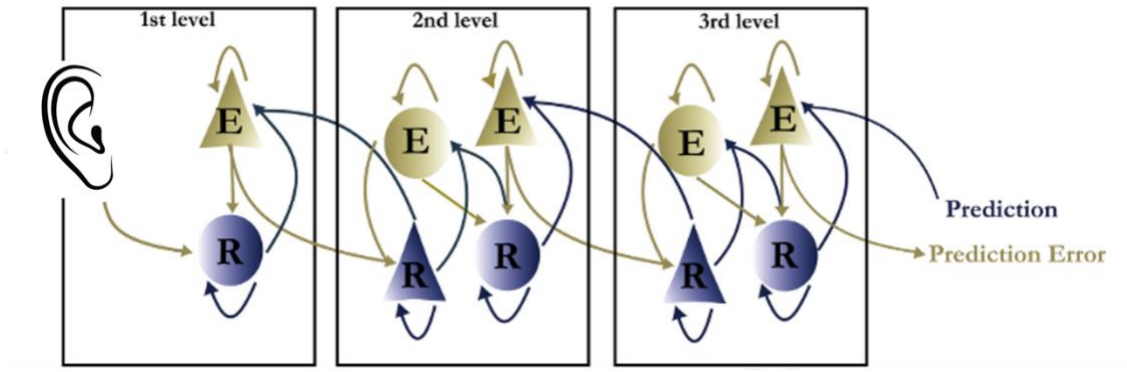


Figure 3: A predictive coding model for auditory perception. The auditory cortex is considered a hierarchical structure (here consisting of three levels) containing top-down and bottom-up routes. In a top-down route, representation units (R) send out (via blue arrows) predictions about the upcoming auditory stimulation. In a bottom-up route, prediction errors are returned (via gold arrows) to representation units. Adapted from Grotheer and Kovács (2016).

Explaining the MMN with the Predictive coding framework is supported by the findings that predictions can modulate the MMN. Lecaignard et al. (2015) found that MMN amplitude decreased as the predictability of deviants increased, even when the magnitude of the deviance remained the same. The predictive coding framework is also supported by computational modeling. Wacongne, Changeux, and Dehaene (2012) built a neuronal model for the MMN based on the Predictive coding framework. The model successfully predicts future stimuli based on the transition statistics of the past input and updates the prediction according to the prediction errors, accounting for the major properties of an MMN.

At first glance, the Predictive coding framework is inconsistent with the adaptation-based model – however, recent studies on the repetition suppression point to a connection between neuronal adaptation and prediction. The current theories state

that neuronal adaptation is not merely a bottom-up process. Instead, it is also a result of minimizing the prediction error through adaptive changes in predictions about the sensory input, resulting from inference and learning implemented under the Predictive coding framework (Auksztulewicz & Friston, 2016; Grotheer & Kovács, 2016). In line with this view, Summerfield et al. (2011) found that the expectation of repetition modulates repetition suppression. They presented subjects with pairs of faces that were either identical or not. In one block, there were more pairs of identical faces than those of non-identical faces (repetition expected). The pattern was reversed in the other block (repetition unexpected). The repetition suppression was measured by comparing the response to the second face in identical pairs to the second face in non-identical pairs. Both the ERP and the theta-band (4-8Hz) spectral power showed that the repetition suppression was more robust when the repetition was unexpected. Grotheer and Kovács (2016) further pointed out that conventional *repetition suppression* might be composed of a passive repetition suppression which occurs in an early time window, and an active expectation suppression which can extend to a much later time window. The experimental support for dissociating expectation suppression and passive suppression comes from Todorovic and de Lange (2012). They presented tone pairs where the second tone is either a repetition or non-repetition of the first tone. They also manipulated the first tone such that it either could predict or could not predict the second tone. They found reduced neural activity (passive repetition suppression) for the repeated tones compared to the non-repeated tones in the early time window (40 – 60ms). In addition, they found another reduced neural activity (expectation suppression) for the expected tones compared to the unexpected tones in the intermediate time window (100 – 200ms) and a later time window extending to

500ms, suggesting the repetition suppression and the expectation suppression have distinct time course.

Back to MMN, if we want MMN to reflect a violation of an expectation based on a memory trace, the MMN better include the modulation motivated by the expectation suppression, which is associated with the waveform of standards. Now consider the two ways of computing MMN. As a reminder, the non-identity approach involves subtracting the ERP response of standards from the ERP response of different stimuli serving as deviants. In contrast, the identity approach subtracts the ERP response of standards from the ERP response of the same stimuli serving as deviants. The identity approach removes the repetition suppression of standards at all. However, as we have seen above, the repetition suppression could consist of a bottom-up passive suppression component and a top-down active prediction component. Therefore, the identity MMN approach will underestimate an MMN magnitude by removing the active prediction component. In contrast, the non-identity MMN approach will inflate an MMN magnitude by including the bottom-up passive suppression component. That being said, the current study adopted the identity MMN approach in the current study as it puts us on the safe side to draw a conclusion based on a positive result.

2.1.2 Across-category MMN and within-category MMN

The MMN reflects not only low-level acoustic changes contradicting the sensory memory but also higher-level linguistic changes contradicting a long-term memory representation. Numerous studies have conducted MMN experiments using linguistic stimuli (Allen, Kraus, & Bradlow, 2000; Dehaene-Lambertz, 1997; Joanisse, Robertson, & Newman, 2007; Miglietta, Grimaldi, & Calabrese, 2013; Näätänen & Alho, 1997; Sharma, Kraus, McGee, Carrell, & Nicol, 1993; Sharma & Dorman,

1999; Shestakova et al., 2002; Sittiprapaporn, Tervaniemi, Chindaduangratn, & Kotchabhakdi, 2005; Winkler et al., 1999). Studies have observed a robust MMN to an across-category contrast but minimal or no MMN to the within-category contrast. For example, Silva, Melges, and Rothe-Neves (2017) observed a robust MMN to the across-category vowel contrast (i.e., /i – e/) but little MMN to the within-category contrast (i.e., two phonetic realizations of /i/ or two phonetic realizations of /e/) with native Brazilian Portuguese speakers, despite their attempt to control for the magnitude of the acoustic difference. With the same sound pair, a larger MMN was observed when the contrast is across-category compared to when the contrast is purely within-category in a listener's language. This pattern has been observed with segmental level contrast (e.g., vowel length) (Dehaene-Lambertz, 1997; Winkler et al., 1999; Ylinen et al., 2006) as well as suprasegmental level contrast (e.g., Chinese tones) (Yu, Shafer, & Sussman, 2017). For example, Winkler et al. (1999) presented the same /e – æ/ contrast to Finnish speakers and Hungarian speakers. The /e – æ/ contrast is across-category for Finnish speakers but within-category for Hungarian speakers. A robust MMN was observed for the Finnish speakers but not for the monolingual Hungarian speakers. These results show that the language-specific categorical knowledge stored in long-term memory significantly affects MMN generation, echoing the behavioral findings that listeners can discriminate across-category contrast much better than the within-category contrast. Nonetheless, an across-category difference alone is not sufficient for an MMN. The stimulus complexity matters. Pettigrew et al. (2004) presented subjects with English real words contrast /del – gel/. They reported a lack of a robust MMN even though the consonant onsets of the standards and deviants (/d/ vs. /g/) are separate phonemes. In contrast, Allen, Kraus,

and Bradlow (2000) used the stimuli extracted from the /da – ga/ continuum and reported a robust MMN. More interestingly, the standard-deviant pairs in their study were either above or below the just-noticeable-difference level in two conditions. However, the MMNs elicited in both conditions did not differ in latency and amplitude, suggesting the brain is sensitive to subtle acoustic differences (as small as 10Hz) beyond the attention level.

A few studies reported a robust within-category MMN. Sharma et al. (1993) presented participants with either a within-category contrast or an across-category contrast falling in the /da – ga/ continuum. They found that both contrasts elicited robust MMNs, which did not differ in amplitude and latency. They concluded that an MMN reflects acoustic processing and is insensitive to language-specific categorical knowledge. But note that they measured the MMN by subtracting the standard-stimulus ERP from the deviant-stimulus ERP in the same block, which introduced the confounds of the stimulus-specific response (see the last paragraph of 2.1.1.3 for details). In another experiment, Sharma and Droman (1999) focused on the VOT difference and presented subjects with stimuli drawn from a /da – ta/ continuum. The standard-deviant pair formed an across-category contrast (30 vs. 50ms VOT) in one block and a within-category contrast (60 vs. 80ms VOT) in another block. This time, they measured the identity MMN (subtracting the ERP to the stimulus serving as standards from the ERP to the identical stimulus serving as deviants) to control for the stimulus-specific response. The across-category contrast elicited a robust MMN, while a within-category contrast elicited “a minimal MMN (p. 1081)”. Nonetheless, even with the stimulus-specific response controlled for, there are still studies showing a within-category MMN with a comparable size to an across-category MMN. Miglietta,

Grimaldi, and Calabrese (2013) tested a within-category contrast /ɛ – e/ and an across-category contrast /i – e/ in a Southern-Italian variety, finding that the MMN to the two contrasts did not differ in amplitude, although the onset of the across-category contrast is earlier. They suggested that the earlier MMN was due to an easier parsing and encoding of language-specific category knowledge. Another example comes from Sittiprapaporn et al. (2005), which tested the native speakers of Thai with a standard /pɔ/ sequence interspersed with a cross-category deviant /pi/ and a within-category deviant /po/. They found that both contrasts elicited the MMN but with an enhanced MMN for the across-category vowel contrast. Furthermore, they found that while the across-category contrast elicited a greater activation in the left temporal cortex, the within-category contrast elicited a greater activation in the right temporal cortex. They explained the topographical difference as the difference between nonlinguistic versus linguistic processing. That is, within-category discrimination involves detecting changes in the low-level physical acoustic features of the stimuli. In contrast, across-category discrimination involves a higher-level language-specific category knowledge. Also, note that the sensitivity to the across-category contrast does not wipe out the sensitivity to the within-category acoustic difference. Joanisse, Robertson, and Newman (2007) examined the MMN elicited by the standard /da/ (1600Hz F2 onset) versus the acoustically closer deviant /ba/ (1100Hz F2 onset) as well as the acoustically further deviant /ba/ (900Hz F2 onset). They found that the acoustically further deviant elicited greater MMN, suggesting the acoustic information beyond the across-category contrast can influence the MMN.

So far, we have seen that the MMN studies involving within-category contrasts show inconsistent findings. It is possible that whether a within-category contrast elicits

the MMN depends on the stimulus type, the stimulus complexity, and the fine-grained acoustic difference of the contrast. As explained below, the current study's interpretation relies on an MMN to a within-category contrast. I thus include a baseline condition to make sure that the current stimuli and the experiment design can elicit a within-category MMN.

2.2 Varying-standard oddball paradigm

The current dissertation examines the nature of memory trace in a varying-standard paradigm. Varying the standards precludes a representation based on a single standard token. Instead, listeners must abstract a higher-level pattern based on the various standard stimuli. Numerous studies have observed an MMN to the violation of an abstract regularity by varying the standards (Paavilainen, Simola, Jaramillo, Näätänen, & Winkler, 2001; Saarinen, Paavilainen, Schöger, Tervaniemi, & Näätänen, 1992; Tervaniemi, Maury, & Näätänen, 1994). For example, Paavilainen et al. (2001) presented a sequence of pure tones varying in frequency and intensity. The tones followed the rule that a tone with a higher frequency also carried a higher intensity. An MMN was elicited to a tone if it has a high frequency with a low intensity or a low frequency with a high intensity, although both the frequency and the intensity level were frequently presented. The MMN must have come from violating the high-frequency-high-intensity pairing – an abstract regularity.

When the varying standards belong to the same category, the extracted regularity is a discrete category representation retrieved from long-term memory. An MMN would indicate a mismatch between the incoming sound and the phonological category representation. This is well documented in Phillips et al. (2000). They varied the VOT of standards along the /da – ta/ continuum. In one condition, an MMN was

observed when the various standards all fell into the /d/ category. In the other condition, no MMN was observed when the standards straddled the /d – t/ VOT boundary, although the acoustic difference between standards and deviants was the same across the two conditions.

The idea of varying standards to elicit a phonological category is further supported in the experimental research on Phonological Underspecification Theories. In those studies, enhanced and earlier MMNs were observed when the standard stimuli are not phonologically underspecified in the memory trace (e.g., standard /t/ vs. deviant /d/), compared to when the phoneme of the standard stimuli has unspecified features (e.g., standard /d/ vs. deviant /t/). Given the identical acoustic distance, the legitimate explanation for the MMN asymmetry is the difference in the phoneme of the standard stimuli. Studies have confirmed the underspecification in both the segmental and suprasegmental features, including the [coronal] feature in German vowels (Cornell, Lahiri, & Eulitz, 2011), the [coronal] and [plosive] feature in German consonants (Cornell, Lahiri, & Eulitz, 2013; Friedrich, Lahiri, & Eulitz, 2008; Scharinger, Bendixen, Trujillo-Barreto, & Obleser, 2012), the [coronal] and the [spread glottis] feature in English consonant (Cummings, Madden, & Hefta, 2017; Friedrich, Eulitz, & Lahiri, 2006; Hestvik & Durvasula, 2016; Scharinger, Merickel, Riley, & Idsardi, 2011; Schluter, Politzer-Ahles, & Almeida, 2016), the English mid-vowel (Eulitz & Lahiri, 2004; Scharinger, Monahan, & Idsardi, 2016, 2012), the [spread glottis] feature in Arabic and Russian (Schluter, Politzer-Ahles, Al-Kaabi, & Almeida, 2017), the dipping tone in Mandarin Chinese (Politzer-Ahles, Schluter, Wu, & Almeida, 2016), and the positional neutralization in Brazilian Portuguese (Silva, Rothe-Neves, & Melges, 2020).

Studies not focusing on Underspecification Theory have also used the varying-standard paradigm to tap into the phonological category (Barrios, Namyst, Lau, Feldman, & Idsardi, 2016; Dehaene-Lambertz & Pena, 2001; Kazanina, Phillips, & Idsardi, 2006; Shafer et al., 2021; Shestakova et al., 2002). For example, Barrios et al. (2016) found that advanced L2 learners showed an MMN to non-native phoneme distinctions, and concluded that L2 learners could form new phonemes beyond the native sound inventory. Kazanina et al. (2006) tested the influence of native language on Russian and Korean speakers' VOT perception, with stimuli drawn from a /d – t/ continuum. Their [d] and [t] stimuli are phonemically distinct in Russian but are allophones of /t/ in Korean. They found that only Russian speakers showed the MMN, confirming the effect of linguistic knowledge.

In sum, using the varying-standard paradigm to elicit a speech MMN is an effective way to understand how speech sounds are represented in the brain. However, the relevant studies have relied on the assumption that varying standards elicit a category representation. It is still unclear whether the memory trace in a varying-standard paradigm retains gradient information along with a category representation. In the next chapter, I introduce the first of the three experiments to investigate the nature of memory trace in a varying-standard paradigm.

Chapter 3

EXPERIMENT 1: WITHIN-CATEGORY MMN IN VARYING-STANDARD PARADIGMA

The current dissertation addresses whether the memory trace contains gradient information of speech along with a category representation elicited in a varying-standard paradigm. As introduced in the previous chapter, using an oddball paradigm with various stimuli serving as standards within a category enforces a representation of that category. For example, if the stimuli for standards are various phonetic realizations of /tæ/, those different [tæ]s will elicit a categorical representation associated with /tæ/. Note that the categorical representation is not necessarily a phoneme category /t/. With the different phonetic realizations of /tæ/, the perceptual category could be an allophonic category of an aspirated /t/ occurring in the onset position of a stressed syllable. Nevertheless, no matter whether the perceptual category is a phoneme category or an allophonic category, the phonetic knowledge must include the gradient information about how the phonetic realizations of /t/ as in the onset position of a stressed syllable.

The current experiment used CV syllables made of an alveolar stop plus the vowel [æ] as the stimuli and manipulated the VOT of the stop to elicit a category representation. To examine whether the memory trace contains gradient information along with a category representation, the VOTs of both the standard and the deviant stimuli fall in a range that can be perceived as [t]. In particular, the VOT of each standard [tæ] corresponds to a typical realization of /tæ/ (e.g., 48ms VOT). In contrast,

the VOT of the deviant [tæ] corresponds to an atypical realization of /tæ/ (e.g., 119ms VOT). If the memory trace contains gradient information besides the elicited category representation of [tæ], we should observe an MMN to the deviants. Note that the gradient information could come from two sources: the acoustic properties of the proximal stimuli and the phonetic knowledge retrieved from long-term memory. If the gradient information is about the acoustic properties of the proximal stimuli, then the acoustic difference between the 48ms VOT and the 119ms VOT could lead to an MMN (and Experiment 1 included a single-standard condition to confirm that). On the other hand, gradient information could also come from the phonetic knowledge retrieved from long-term memory.

The phonetic knowledge could take the form of a prototype corresponding to the most typical phonetic realization of /tæ/. It could also take the form of a probability distribution of the empirical VOT realizations of a /t/ produced at the onset position of a stressed syllable. The prototype-like phonetic knowledge is consistent with Prototype Theory of categorization in cognitive science (E. Rosch, 1988). The support for phonetic knowledge taking the form of a probability distribution of empirical phonetic realizations can be found in Kronrod, Coppess, and Feldman (2016). They proposed a model where listeners use their knowledge of the probability distribution of the phonetic realizations to infer speakers' intended productions. In that model, listeners' knowledge of a category is a normal distribution with a mean μ and a standard deviation (SD) σ. Their model closely1 captures different degrees of categorical effects. According to Chodroff and Wilson (2018), the probability distribution of the empirical VOTs of a word-initial /t/ resembles a normal distribution

with a mean of about 60ms and a certain SD (Figure 4)[3]. Therefore, the phonetic knowledge of a /t/ occurring at the onset position of a stressed syllable could take the form of a distribution with a mean of 60ms. An atypical deviant VOT would be viewed as an outlier of that distribution, assuming that the brain adopted a criterion of, say, 3SD, to declare a member to be an outlier. It would thus elicit an MMN response, as the brain is sensitive to the outlier of a represented statistical structure (Garrido et al., 2013). Alternatively, if the phonetic knowledge is about the acoustic properties of a prototype, the prototype would approximate the mean of the distribution of the empirical VOT, which is estimated to be 60ms. In that case, an atypical deviant VOT would also contrast the prototype and thus elicit an MMN response. To summarize, a memory trace containing gradient information would lead to an MMN response, whether the gradient information is from the statistical summary of the proximal stimuli or the phonetic knowledge retrieved from long-term memory.

[3] The empirical VOT distribution in fact does not pass the normality or log-normality test. However, the normality test is practically useless with the current large sample size of 7827.

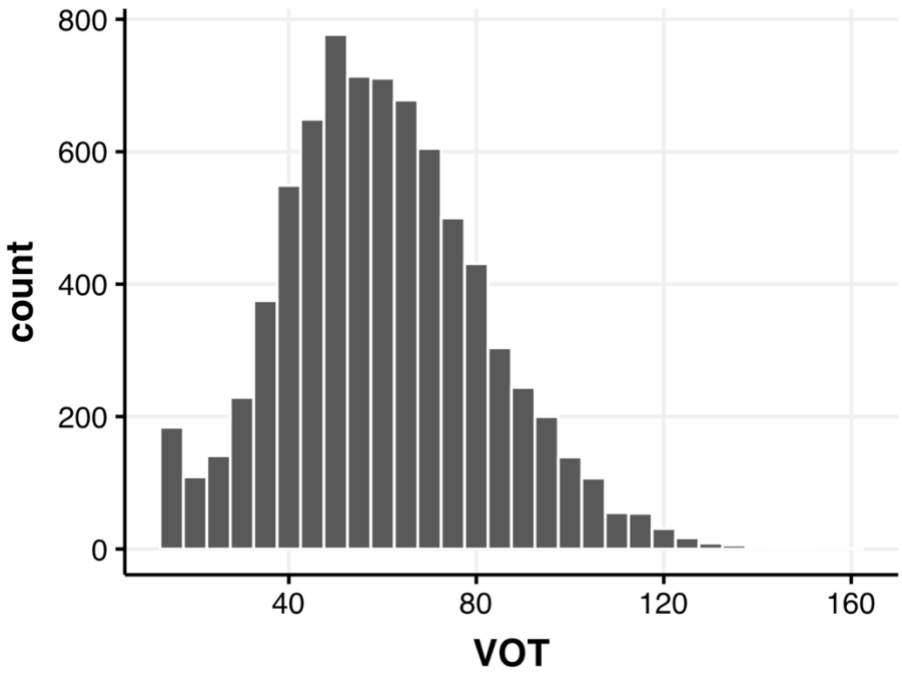


Figure 4: Frequency distribution of empirical VOTs of /t/. The distribution has a bell shape with a mean of 60. The VOTs are from the 7827 [t] tokens produced at the onset of stressed word-initial syllables. The syllables were extracted from a corpus of sentences produced by 180 native American English speakers. The data come from Chodroff and Wilson (2018).

3.1 Previous efforts

Our lab has done experiments tapping into the content of memory trace during a varying-standard MMN paradigm. (Rhodes, Avcu, Han, & Hestvik, 2022; Rhodes, Han, & Hestvik, 2019). Following the same logic to elicit a phonological category, Rhodes, Han, and Hestvik (2019) had subjects passively listen to synthesized speech sounds drawn from a /dæ-tæ/ continuum. They constructed two conditions with various standards: In the low-T condition, the standard syllable onset had a VOT of

60, 65, and 70ms; in the high-T condition, the VOT became 75, 80, and 85ms. The deviant stimulus for both conditions was a [dæ] with a 15ms VOT. The two conditions are illustrated in Figure 5.

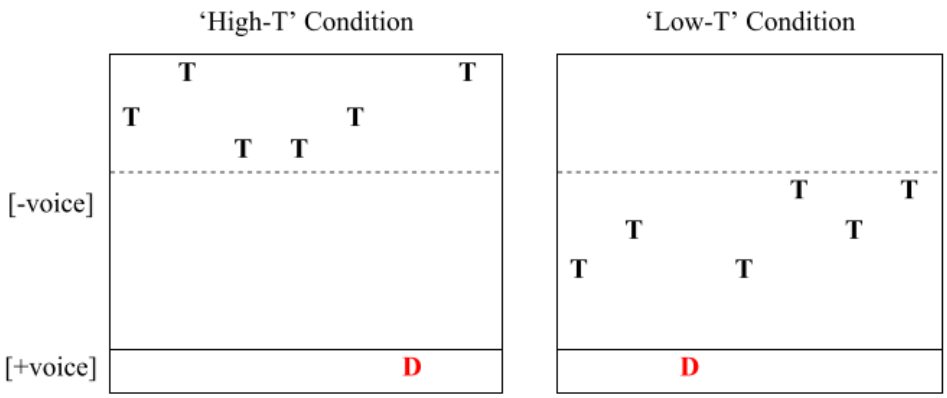


Figure 5: Two experimental conditions in Rhodes, Han, and Hestvik (2019). In both conditions, standards belong to a voiceless category (T), and deviants belong to a voiced category (D). The solid black line represents the boundary between [+voice] and [-voice]. The dotted line separates the High-T standards from the Low-T standards. Adapted from Rhodes, Han, and Hestvik (2019).

If the memory trace retains gradient information of the proximal stimuli, we expect the two conditions to elicit MMN with different magnitudes, as the acoustic distances between the standards and the deviants are different in the two conditions. It turned out that no MMN magnitude difference was found, so we concluded that the memory trace in the varying-standard paradigm is solely a discrete category representation (and another interpretation would be that the two conditions elicited the

identical phonetic knowledge of a category[4]. However, since the design used an across-category contrast (standard /t/ vs. deviant /d/), we suspected that the MMN response was saturated due to an across-categorical difference. Therefore, in Rhodes et al. (2022), we designed another experiment focusing on a within-category contrast. In the low-T condition, the standard syllable onset had VOT values of 90, 95, and 105ms; in the high-T condition, the VOTs became 110, 115, and 120ms. The deviant stimulus for both conditions was a [tæ] with a 50ms VOT, as illustrated in Figure 6.

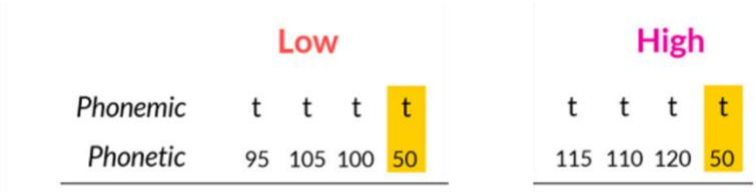


Figure 6: Two experimental conditions in Rhodes et al. (2022). The "Low" condition standards carry 95, 100, and 105ms VOTs. The "High" condition standards carry 110, 115, and 120ms VOTs. The highlighted stimuli represent the infrequent deviants with a 50ms VOT. Adapted from Rhodes et al. (2022).

However, in that experiment, Rhodes et al. only found participants with a VOT threshold above 50ms manifested MMN. The only MMN found was from those who perceived the deviants as /d/ and the standards as /t/ – an across-category MMN. Consequently, the results of that experiment cannot be used to infer a memory trace associated with a within-category MMN. So far, there is no clear evidence that the

[4] An alternative explanation is that the difference between the two conditions was simply too small to observe a MMN difference.

memory trace in the varying-standard paradigm includes gradient information. Experiment 1 aims to fill the gap.

If Experiment 1 does not find a within-category MMN in a varying-standard paradigm, we conclude that a within-category MMN can only be driven by a categorical difference. Alternatively, the lack of an MMN might be because the acoustic contrast between the deviants and the phonetic knowledge is too small to elicit an MMN. To eliminate that possibility, Experiment 1 included a within-category contrast with a fixed stimulus [tæ] serving as standards, i.e., a single-standard condition. The VOT of the single stimulus [tæ] was set to be the mean VOT value of the standard VOTs in the varying-standard condition. Therefore, a lack of within-category MMN in the varying-standard condition is interpretable only if we find a within-category MMN in the single-standard condition.

To anticipate the results, if Experiment 1 found an MMN in both the single-standard and varying-standard conditions, we found evidence supporting a memory trace containing gradient information. If we found an MMN in the single-standard condition but not in the varying-standard condition, then the result is consistent with a memory trace dominated by a discrete category, as shown in the following table.

Table 1: MMN predictions.

	Paradigm	
Memory trace…	**Single-standard**	**Varying-standard**
With gradient information:	MMN	MMN
Without gradient information:	MMN	no MMN

It should be noted that if we do not observe an MMN with a fixed standard stimulus, it could be due to a flaw in the experiment design or equipment failure. To rule out that possibility, Experiment 1 also included conditions where standards and deviants form across-category contrasts as a sanity check.

3.2 Methods

3.2.1 Participants

Sixty-three subjects aged 18-30 were recruited from the University of Delaware. All subjects (55 females[5], mean age = 21, standard deviation = 1) are English monolingual speakers and reported no language impairment history. Subjects received either $20 or extra credit for completing the experiment. The experiment procedure was approved by the University of Delaware Internal Review Board and was compliant with the principles for ethical research established by the Declaration of Helsinki.

3.2.2 Stimuli

3.2.2.1 Creating stimuli set

The experimental stimuli were drawn from a set of resynthesized CV syllables along the /dæ-tæ/ continuum. To create the stimulus set, I started by recording [dæ] and [tæ] syllables produced by a female native speaker of American English, using Zoom H4n Pro Audio Recorder, with a sampling rate of 44100 Hz. The speaker, a

[5] Since there is no evidence of the gender effect on the phonetic MMN amplitude and peak latency in normal-hearing adults (Kasai et al., 2002), we do not consider that the size difference in gender group would cause a problem for the current design.

Ph.D. student in linguistics, first produced multiple samples of [dæ] and [tæ]. Among those samples, I selected two that did not feature a creaky voice and had similar intensity levels. Based on these two syllables, a Praat script (Winn, 2020) was used to generate 146 syllables along the /dæ - tæ/ continuum with VOTs ranging from 0ms to 145m, increasing by a 1ms step. The script adopted the approach of progressive cutback and replacement of the vowel to modify the VOT period. The F0 onset frequency to be constant (279 Hz) across all VOTs, as F0 plays a meaningful role in voicing detection (Kapnoula, Winn, Edwards, & Mcmurray, 2017). For other parameters, the default values were used. Every generated syllable had a duration of 620ms, including a cosine ramp for the first 5ms and the last 80ms. The audible portion of each syllable is about 500ms. Figure 7 shows the waveform and the spectrogram of [tæ] with a 48ms VOT.

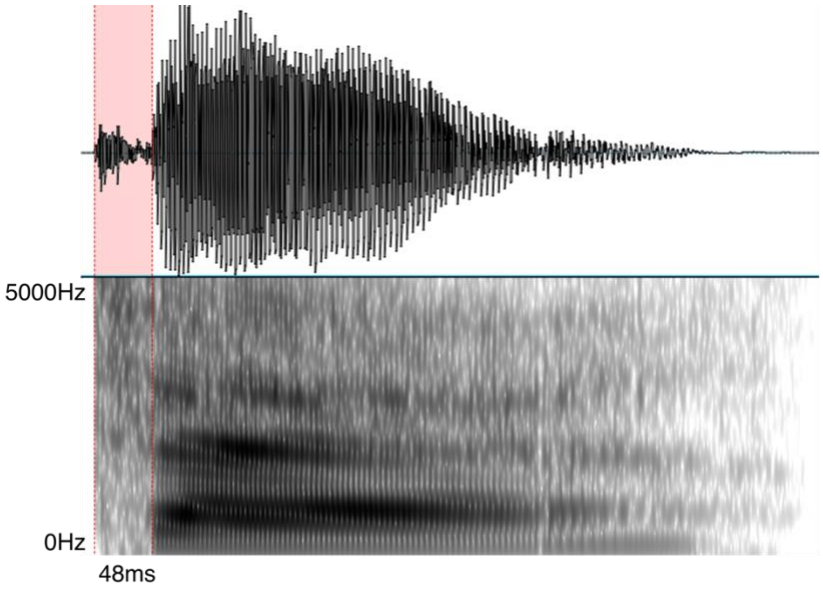


Figure 7: The waveform (upper) and the spectrogram (lower) for a [tæ] with a 48ms VOT. The VOT is highlighted in pink.

The experiment also contains target stimuli to direct participants' attention toward the acoustic properties of the stimuli. The target stimuli were created by replacing a portion of the 48ms VOT [tæ] with a pure tone of 440 Hz. The replaced portion was either the first 150ms after the vowel onset or the last auditable 150ms of the stimulus.

3.2.2.2 VOT selection

Since I expect the brain to extract a regularity based on the various standards, it is crucial that the brain can discriminate between those standards with different VOTs. Processing VOT can be viewed as processing the temporal information of a gap preceded and followed by two audible events. Lister and Tarver (2004) found that the just-noticeable difference for a VOT-type gap (burst-gap-vowel) was around 25ms for younger subjects and 35-55ms for older subjects. Studies also found that the just-noticeable-difference for VOT varies depending on whether the contrasting VOTs belong to the same category: for two across-category stops (i.e., voiced vs. voiceless), subjects were able to capture a VOT difference as small as 10ms; but they could hardly discriminate two within-category stops with a VOT difference as large as 60ms (Aslin, Pisoni, Hennessy, & Perey, 1981; Soli, 1983). However, note that the above behavioral results reflect the conscious judgment of the VOT difference, which does not necessarily correspond to the brain's sensitivity to the subtle VOT difference. Indeed, both fMRI and ERP studies have shown that the auditory system could capture a more subtle difference regardless of conscious awareness (Blumstein et al., 2005; Elangovan & Stuart, 2011; Toscano et al., 2010). In an fMRI study, Blumstein and his colleagues (2005) found that the left inferior frontal gyrus and the cingulate cortex showed different activation levels to stimuli differing in 10ms VOT. In an ERP study,

Toscano et al. (2010) found that the N1 amplitude was sensitive to a VOT difference of 5ms. In contrast to a categorical function typically observed in a VOT identification task, they found a relatively linear decrease in the N1 amplitude as the VOT increased, regardless of the voicing. Also, using stimuli with a minimal VOT difference of 5ms, Hestvik and Durvasula (2016) obtained results consistent with the underspecification hypothesis, which relies on the brain's detection of the within-category VOT difference to extract a phoneme representation. Other MMN studies had adopted a minimal within-category VOT difference of 4ms (Kazanina et al., 2006) and 8ms (Phillips et al., 2000), much smaller than the just-noticeable difference found in behavioral studies. The just-noticeable-difference for VOT detection may be relative to the duration of VOT. In that case, the VOT difference in all the above studies is above 10% of the largest VOT used for the stimuli. In the current experiment, I also set the minimal VOT difference to be more than 10% of the largest VOT of standards.

Another thing to notice is that the previous studies all set the VOT steps on a linear scale. The VOT difference was constant in milliseconds between any two consecutive steps along a continuum. The perception of VOT may be better modeled on a logarithmic scale, which is consistent with Weber's law. That is, a physical property with a larger magnitude requires a proportionally larger difference to be equally discriminable. Previously studies have found that a logarithmic scale is a better fit for the intuition of numbers (Dehaene, Izard, Spelke, & Pica, 2008) and for the neuronal activities responding to auditory and visual stimuli (Scheler, 2017). Logarithmic modeling even works on the grouping of audio events separated by gaps with different durations (Ren, Allenmark, Müller, & Shi, 2020), which might share the same mechanism as VOT perception. Measuring the acoustic properties in the

logarithmic scale is consistent with the nature of the just-noticeable-difference in music and speech perception, which is proportionate to the reference sensory level (Ekman, 1959). In the current study, VOT values were determined on a logarithmic scale, such that the VOT values increase faster for longer VOTs. This decision is based on the following consideration: If the VOT perception indeed follows Weber's law, our design would yield more accurate results than the previous studies. If the VOT perception is linear, the acoustic difference between the standards and deviants would be larger in the within-category condition than in the across-category condition. This potential discrepancy, however, does not invalidate our design as long as the acoustic difference between standards and deviants in the within-category condition is no smaller than that in the across-category condition.

To select the stimuli with specific VOTs for the experiment, I started with a 48ms VOT, then determined the VOT one step away from the 48ms VOT towards the /d/ end to be 42ms. The remaining VOTs were determined such that the difference between every two steps was the same as the difference between 42ms VOT and 48ms VOT on a logarithmic scale. The VOT selection procedure resulted in stimuli with the following VOTs in milliseconds: 19, 22, 25, 33, 37, 42, 48, 55, 63, 72, 82, 93, 105, and 119. Among those stimuli, I chose the stimuli with 42ms, 48ms, and 55ms VOTs to serve as standards, and the stimulus with a 19ms VOT and a 119ms VOT to serve as the deviant for the cross-category condition and the within-category condition, respectively. If the VOT perception conforms to a logarithmic scale, the 48ms VOT is auditorily equidistant from the 19ms VOT and the 119ms VOT.

After the experiment, participants were not symmetrically asked about how they perceived sounds in the experiment, but a few reported that the sound was natural.

3.2.3 Design

The current experiment contains four oddball blocks and one control block. Below I introduce them separately.

3.2.3.1 Oddball blocks

The oddball blocks included two standard types, corresponding to a single-standard condition and a varying-standard condition. In the single-standard condition, all the standards were realized by one single stimulus – a [tæ] syllable with a 48ms VOT; in the varying-standard condition, standards were realized by the three different [tæ]s with VOTs of 42ms, 48ms, and 55ms. Each standard type was associated with two deviant types, corresponding to a within-category condition and an across-category condition. In the within-category condition, the deviant was realized by a [tæ] with a 119ms VOT, belonging to the same /t/ category as the standards; in the across-category condition, the deviant was realized by a 19ms VOT [dæ], yielding an across-category contrast. The combination of the two standard types and the two deviant types resulted in four conditions corresponding to four oddball blocks. Table 2 summarizes the stimulus VOTs in each block.

Table 2: Standard and deviant VOTs in each block.

		VOT (ms)	
Standard type	**Deviant type**	**Standards**	**Deviants**

single-standard	across-category	48	19
single-standard	within-category	48	119
varying-standard	across-category	42, 48, 55	19
varying-standard	within-category	42, 48, 55	119

In each block, the number of deviants was fixed to be 100. Each deviant was preceded by a train of standards, with the number of standards ranging between 3 and 9. The number of the standards in each train was drawn from a uniform distribution of integers ranging between 3 and 9 with the MATLAB function *randi()*. I chose the uniform distribution to prevent subjects from predicting where a deviant would occur after a certain amount of standards, which could attenuate the MMN magnitude, as the MMN is highly context-based (Sussman, Chen, Sussman-Fort, & Dinces, 2014). The number of standards in each train was randomly drawn from the distribution, so the total number of standards varied across participants and blocks, ranging between 578 to 647. Figure 8 shows an example distribution of the number of standards in each train in different blocks for a randomly picked subject. The overall standard-to-deviant ratio is about 6:1. For each participant, after the stimulus list of standards and deviants was determined, I randomly interspersed each oddball block with 33 target stimuli. The complete stimulus list was fed into the E-Prime program before each experiment session.

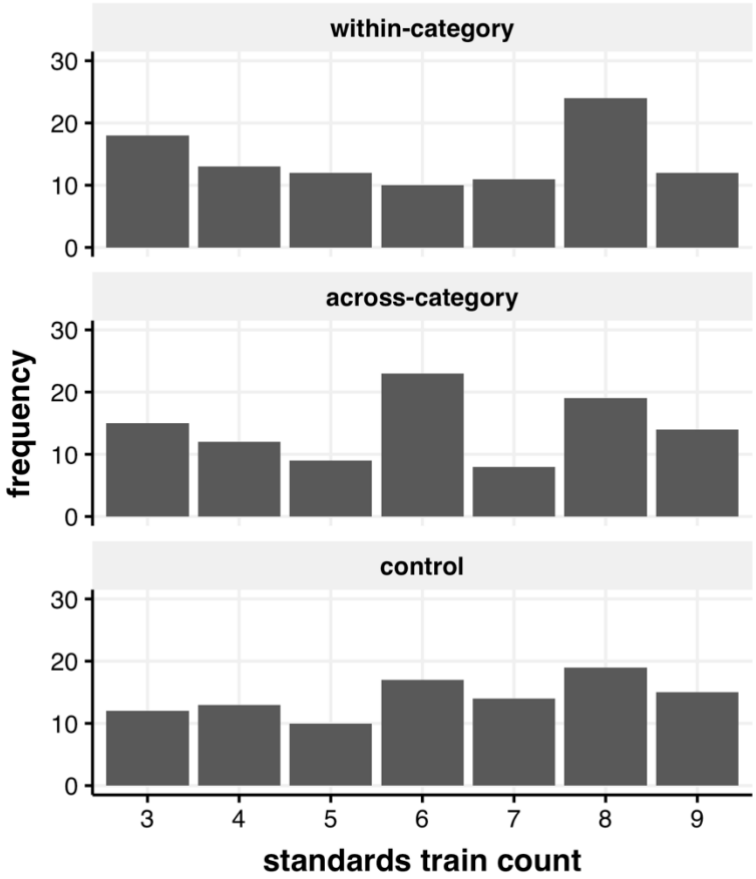


Figure 8: Frequency distributions of standards train counts in different conditions. Data from Subject 10.

3.2.3.2 Roving-standards control block

As discussed above, the current study measures MMN using the identity MMN approach, by subtracting the ERP response to the stimulus serving as deviants from the ERP response to the same stimulus serving as standards. Since the comparison is between two identical stimuli, it is believed this the magnitude obtained following this approach precludes an impact of the difference in the physical properties of the different stimuli. The four oddball blocks in the current experiment are sufficient for a

non-identity approach to computing MMN. But to obtain an identity MMN, we need a control block.

The stimuli used in the control block were a 19ms VOT [dæ] and a 119ms VOT [tæ]. The stimulus presentation in the control block followed a roving-standards paradigm (Bader, Schröger, & Grimm, 2017; Cowan, Winkler, Teder, & Näätänen, 1993), which alternated between a train of 19ms VOT [dæ]s and a train of 119ms VOT [tæ]s. After each alternation, the first token in a train was considered a deviant, and the following tokens within the same train were considered standards. There were altogether 100 deviants in the control block, including 50 [dæ]s of a 19ms VOT and 50 [tæ]s of a 119ms VOT[6]. As in the oddball blocks, each deviant in the control block was also preceded by a train of standards with the number of standards drawn from a uniform distribution of integers between 3 and 9. The overall standard-to-deviant ratio in the control block was also about 6:1. After the standards and deviants were generated, I randomly interspersed the control block with 34 target stimuli.

To keep one experiment session reasonably short, the deviant type was made a within-subject variable but the standard type a between-subject variable. Therefore, one participant would go through both an across-category block and a within-category block, either in a single-standard setting (corresponding to a single-standard group) or in a varying-standard setting (corresponding to a varying-standard group) (Figure 9).

---

[6] This roving-standard control block allows us to look at the MMN responses elicited by the 19ms VOT and the 119ms VOT in the same block. Note that the trial number of each stimulus in the control block differs from the same stimulus in the oddball block, which hinders a direct comparison between the control block's MMN and the oddball blocks' MMN. Nonetheless, I included a separate analysis for the control block's MMN response in Appendix.

Besides the two oddball blocks, each participant also completed the same control block. Since each of the two oddball blocks contained 33 target tokens, and the control block contained 34 target tokens, one participant would encounter 100 target tokens in one experiment session. The inter-stimulus interval was set to vary between 650ms and 920ms, increasing by a 30ms step, to prevent participants from anticipating the onset of the following stimulus.

The identity MMN was computed by comparing the 19ms VOT [dæ]s and the 119ms VOT [tæ]s serving as deviants in the oddball blocks to the same stimuli serving as standards in the control block. Figure 9 illustrates the design and the comparing strategy of the identity MMN.

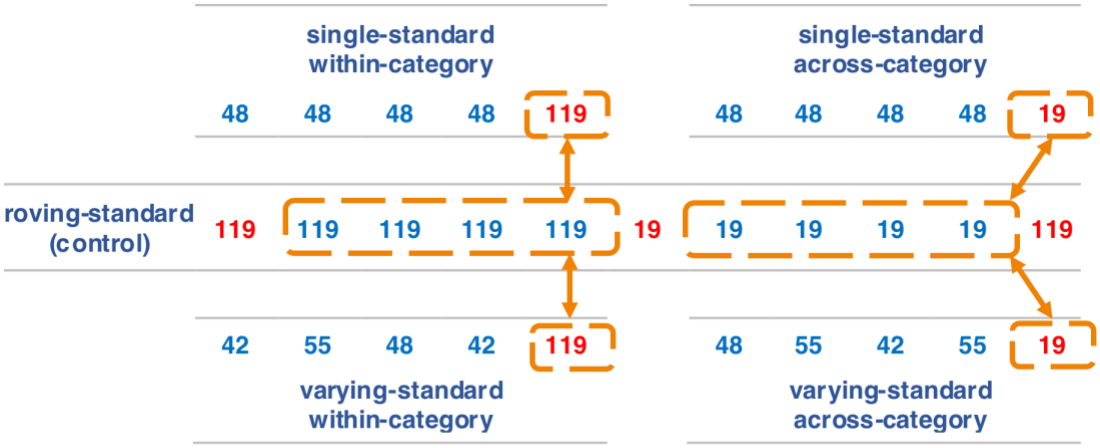


Figure 9: Illustration of the experimental design and how I compute an identity MMN. The experiment contains four oddball blocks and one control block. Each subject completed the control block and two oddball blocks (either two within-category blocks or two across-category blocks). In each block, standard VOT values are in blue and deviant VOT values are in red. To compute an identity MMN, I subtract the ERP of the standards in the roving-standard control block from the ERP of the deviants in the oddball blocks (double-sided arrows).

The across-category conditions served as a sanity check and should always elicit an MMN regardless of the standard type. The core of the design is the within-category conditions. If the memory trace retains gradient information along with a category representation, we should obtain a within-category MMN in both the single-standard and varying-standard blocks. If the memory representation does not retain gradient information, we should obtain a within-category MMN in the single-standard block but not in the varying-standard block.

3.2.3.3 Phoneme identification task

After the EEG session, subjects completed a phoneme identification task. The task used a two-alternative forced-choice procedure where a subject identified whether the perceived CV syllable (from the stimuli set of the /dæ-tæ/) started with a /d/ or /t/. The phoneme identification task served two purposes: First, I expect our stimuli to have a clear VOT to elicit a phoneme representation. That means that the judgment for each VOT value should be consistent and show little variability. Second, since the smallest VOT used in our experiment is a 42ms VOT in the varying-standard conditions, to ensure each subject who participated in the varying-standard conditions perceived the 42ms VOT as belonging to /t/, the subject's VOT threshold should be below 42ms. Thus, I use the VOT threshold of each subject as an exclusion criterion, which resulting in excluding one subject.

The phoneme identification task was conducted after the EEG session to ensure that constant exposure to long VOTs would not shift their perceptual threshold to above 42ms. Studies have shown that repeated exposure to stimuli belonging to one category increases the probability that the same type of stimuli will be perceived as belonging to the opposite category (Eimas & Corbit, 1973). The current experiment

did not aim to assess whether a perceptual boundary shift would occur. Instead, the data is interpretable as long as the shifted perceptual boundary (if any) is still below 42ms.

The stimuli used in the phoneme identification task were the 15 CV syllables chosen from the constructed stimuli set as detailed above. The VOT values of those syllables ranged from 0 to 70 with an increment step of 5ms. The task consisted of six blocks. In each block, all 15 syllables were randomly presented. There were 90 trials in total.

3.2.4 Procedure

Each subject was assigned to either the single-standard group or the varying-standard group according to their subject ID. Both groups went through one control block, one across-category block, and one within-category block. Before the EEG recording, participants went through a practice session to familiarize the experiment procedure. In the practice session, participants were informed that they would hear simple CV syllables throughout the experiment. They would also hear target syllables containing a beep. Their task was to press a button with their most frequently used hand once they heard a target syllable. A picture of a jigsaw puzzle piece would show up on the screen at a button press response or one second after the target syllable onset if no button press was detected. The picture would be colored if the button press response was made within one second after the target syllable onset and would be in greyscale otherwise, as illustrated in Figure 10. If all the 100 jigsaw puzzle pieces were colored, participants would get a bonus of 3 dollars. The inclusion of the jigsaw puzzle game is to keep participants from boredom and from eliciting alpha waves.

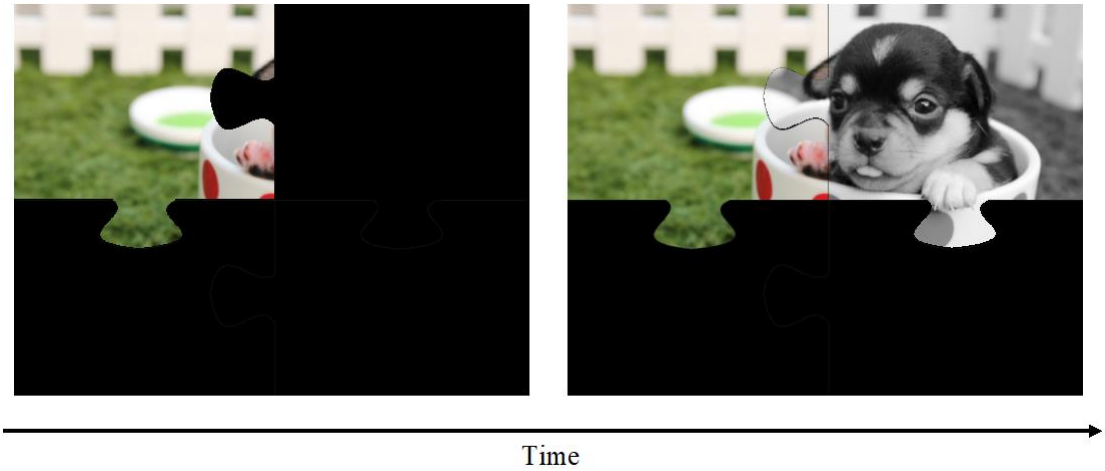


Figure 10: Demo of jigsaw puzzle task. Subjects were asked to press a button when they hear a target syllable. A picture of a jigsaw puzzle piece would show up on the screen at a button press response or one second after the target syllable onset if no button press was detected. The picture would be colored (left) if the button press response was made within one second after the target syllable onset and would be in greyscale (right) otherwise.

During the EEG recording, participants were seated in a sound-attenuating booth. The stimuli were presented via two free-field speakers at the intensity of 65dB, measured at the location of a participant's forehead with the RadioShack Sound Level Meter. For each experiment session, the control block was always presented first. This is because the two stimuli in the control block, a 19ms VOT [dæ] and a 119ms VOT [tæ], were presented with the same probability and thus should not bias participants' perception towards one category or the other. After the control block, the order of the two oddball blocks was counterbalanced across participants. Each block took about 15 minutes. After each block, participants took a break for about 5 minutes. One experiment session took about 2 hours, including the EEG net placement, instruction, breaks, and the EEG net removal.

After the EEG session, the phoneme identification task was conducted. Subjects were told that they would hear a simple CV syllable starting either with /d/ or /t/. They were instructed to press button 1 for /t/ and button 5 for /t/ and to rely on their first impression to make the judgment. The whole task takes about 5 minutes to complete.

3.2.5 Apparatus, data acquisition, and data processing

The experiment was programmed using E-Prime 2 and MATLAB 2021. An E-Prime Extension package for Net Station was used for the EEG acquisition. The continuous EEG data were recorded with the 64-channel HydroCel Geodesic Sensor Nets. Before the EEG recording, the impedance of each channel was lowered to below 50kΩ. During the EEG recording, the incoming analog signal underwent an online 125 Hz low-pass filter to prevent aliasing and was digitized with a sampling rate of 250 Hz. A participant's electro-ocular activity was recorded from 4 bipolar channels around the eyes. Channel E65 (corresponding to Cz in the 10-10 system), placed on the vertex of the scalp, was used as a reference channel.

The recorded data were passed through a first-order high-pass filter of 0.1Hz to remove slow drifts. Then a finite impulse response low-pass filter of 40 Hz with a roll-off of 2Hz was applied to the data to remove line noise and any frequency above 40Hz. The filtered data were segmented into epochs of 1000ms in duration, time-locking to the stimulus onset, including a 200ms pre-stimulus time window. The 200ms pre-stimulus period was also used as a baseline. The segmented data were submitted to an automated process of eyeblink subtraction using ICA with the ERP PCA toolkit (Dien, 2010). An eyeblink template was automatically generated for each subject. An ICA component was marked as an eyeblink component and was subtracted

from the data if it was correlated at r = .9 or greater with the eyeblink template. After the eyeblink subtraction, the data were submitted to the artifact correction procedure to remove bad channels and movement artifacts. For each trial, a channel was marked bad if its best absolute correlation with its neighboring channels fell below r = .4 across all time points. Bad channels were replaced via a spline interpolation from surrounding good channels. If a channel was marked bad in over 20% of trials, it was considered bad in all trials. A trial was marked bad and was dropped if it contained more than 10% bad channels. Table 3 shows the mean percentage of bad trials in each condition.

Table 3: Percentage of bad trials in each condition

		Percentage of bad trials (%)	
Standard type	**Deviant type**	**Standards**	**Deviants**
single-standard	across-category	2.7%	2.7%
single-standard	within-category	2.9%	3.3%
varying-standard	across-category	2.5%	2.3%
varying-standard	within-category	2.4%	2.4%
Control		3.1%	3.3%

3.2.6 Planned signal processing

3.2.6.1 Deciding time window and channels for MMN

To identify the time window and the channels for the MMN analysis, I applied a temporal principal component analysis (PCA) to decompose the data (Dien, 2012).

Separate PCAs were run for the across-category contrast and the within-category contrast. For the PCA input to the across-category contrast, I first computed one deviant ERP for each subject by averaging all the 19ms VOT deviant responses in the oddball blocks, collapsing the single-standard group and varying-standard group. Then computed one standard ERP was computed for each subject by averaging all the 19ms VOT standard responses in the control condition, collapsing the single-standard group and varying-standard group. One difference ERP was then derived for each subject by subtracting the standard ERP from the deviant ERP. The difference ERP was used as the input to the temporal PCA for the across-category conditions. The same procedure was applied to obtain the PCA input to the within-category contrast, except that all ERP responses were from the 119ms VOT. Because the PCA is based on the difference ERP, it should only unravel the ERP components (e.g., MMN) that can be derived in a difference ERP. Any ERP component with the same amplitude and latency in standards and deviants should be ignored.

The temporal PCA procedure computed a covariance matrix of the difference ERP data by treating each time point as a variable and each participant-channel combination as an observation. From the covariance matrix, latent temporal factors were extracted. Those temporal factors were ranked based on the total variance they accounted for. In principle, the temporal factors that reflected that genuine ERP component should account for a fair amount of variance and thus have a relatively high rank. Each temporal factor can be considered a linear combination of the ERP amplitudes at all time points, with each time point contributing differently to a given temporal factor. The contribution of a time point, or the weight of that time point for a given temporal factor, was indexed by a factor loading. For a given temporal factor,

the time point with the highest factor loading was determined to be the peak latency, meaning that that time point contributes the most energy to that temporal factor compared to other time points.

The time window for ERP analysis was determined based on those temporal factors and the factor loadings of time points. Specifically, I looked through each temporal factor to examine whether the peak latency (the time point with the largest factor loading) of a given factor fell within 100-300ms after the stimulus onset because MMN typically peaks at about 100-250ms after the deviance onset and could peak at about 200-300ms for barely discriminable contrast (Näätänen & Alho, 1995; Näätänen, Pakarinen, Rinne, & Takegata, 2004). The time window for analysis comprised the time points centered around the peak latency and carried a factoring loading of 0.6 or higher. I then chose the channels for analysis based on the peak channel, which showed the greatest negativity at the peak latency. Since a typical spatial distribution of MMN exhibits a frontocentral negativity topographically, a proper peak channel should be located in the frontocentral area. For data analysis, the dependent measure was the mean ERP averaged over the selected time window and channels for each participant.

3.2.6.2 Statistical analysis

The statistical analyses were conducted using the R software (R Core Team, 2021). I built linear mixed-effects models to analyze the data using the *lmer* function from the *lme4* package (Bates, Mächler, Bolker, & Walker, 2015). Separate models were built for the across-category contrast and the within-category contrast. This is because the across-category serves as a sanity check and thus is not the core part of the logic of the experiment. The dependent measure was the mean ERP amplitude

averaged over the PCA-delimited time window and the channels for each subject. The independent variables included Group (single-standard vs. varying-standard), Stimulus (standard vs. deviant), and an interaction between the two. The model included Subject as a random intercept. For each fixed factor, I constructed an orthogonal contrast for the factor levels. Since each factor contained only two levels, the coefficients computed with the orthogonal contrasts were indicative of the main effect (and the interaction). The model's explanatory power was obtained from the *report* package (Makowski, Ben-Shachar, Patil, & Lüdecke, 2021), and the model's parameter coefficients using the *report* function and the *model_parameters* function from the *parameters* package (Lüdecke, Ben-Shachar, Patil, & Makowski, 2020). After obtaining the main effect, I further examined the simple contrast between the standard ERP and the deviant ERP for each group using the *emmeans* function from the *emmeans* package (Lenth, 2021). For the effect size, I calculated the partial eta-squared (η_p^2) for the main effect and Cohen's d for the simple contrast using the *effectsize* package (Ben-Shachar, Lüdecke, & Makowski, 2020).

3.3 Results

3.3.1 Behavioral Results: Phoneme identification task

For each participant, the percentage of /t/ response on each VOT value was calculated. A sigmoid function was then used to determine each subject's perpetual boundary for the /d – t/ continuum. This is done by computing the log odds of a /t/ response versus a /d/ response for each VOT value and estimating the parameters for a sigmoid function. The perceptual boundary was determined at the VOT value, where

the function estimates a 50% response of /t/ (where the log odds = 0). Figure 11 shows the frequency distribution of the boundary VOTs obtained from all the participants.

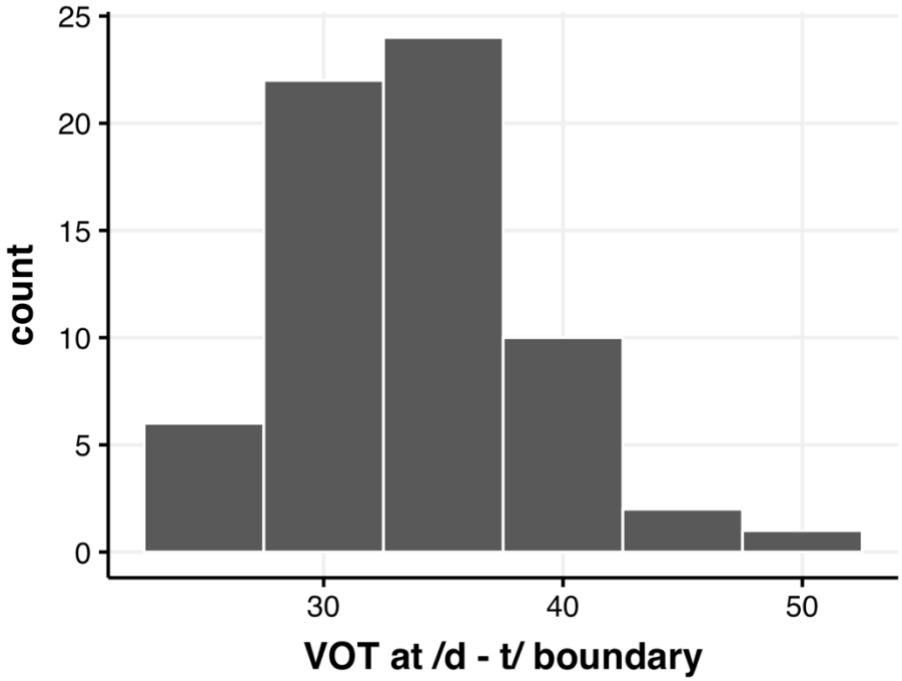


Figure 11: Histogram of perceptual boundary VOTs for /d – t/ continuum. Each subject's perceptual boundary was determined at the VOT value where the sigmoid function estimates a 50% response of /t/. The histogram shows a normal distribution (passing both the normality and log-normality test) with a mean of 34ms and an SD of 5ms.

One subject from the varying-standard group had a threshold VOT greater than 42ms. That means they made a /d/ response when they heard the 42ms VOT stimulus. I excluded that participant from our analysis to ensure all the sounds were perceived as /t/. Figure 12 shows the perceptual function averaged over all participants.

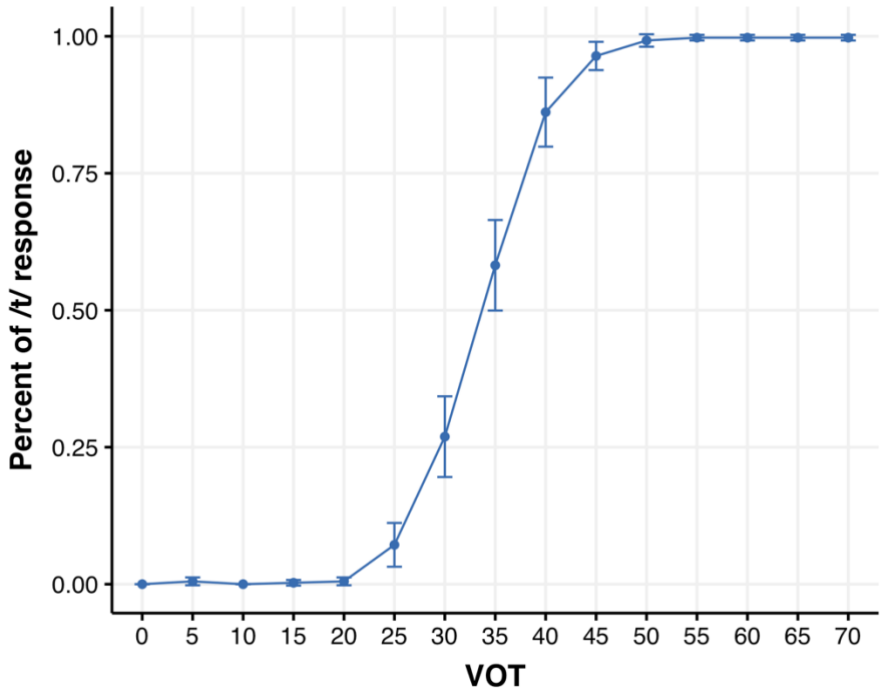


Figure 12: Percent of /t/ response at each VOT. The percentage values were averaged across the 63 subjects. The function shows a clear categorical trend with a boundary at 34ms VOT.

3.3.2 ERP results: MMN

Here I report the PCA solutions and the ERP results for the across-category contrast and the within-category contrast separately. The results for the across-category contrast is reported first. Readers more interested in the within-category contrast results can jump to 3.3.2.2.

3.3.2.1 Across-category MMN

3.3.2.1.1 PCA solution

Following the PCA procedure specified above, I ran a temporal PCA with a Kaiser weighting (kappa = 3). To determine the factors to retain, I used a scree plot in combination with a Parallel Test (Horn, 1965), which compared the factors extracted from the original data to those from a randomized dataset. Twelve temporal factors were retained as they accounted for more variance than the factors extracted from the randomized dataset (Figure 13).

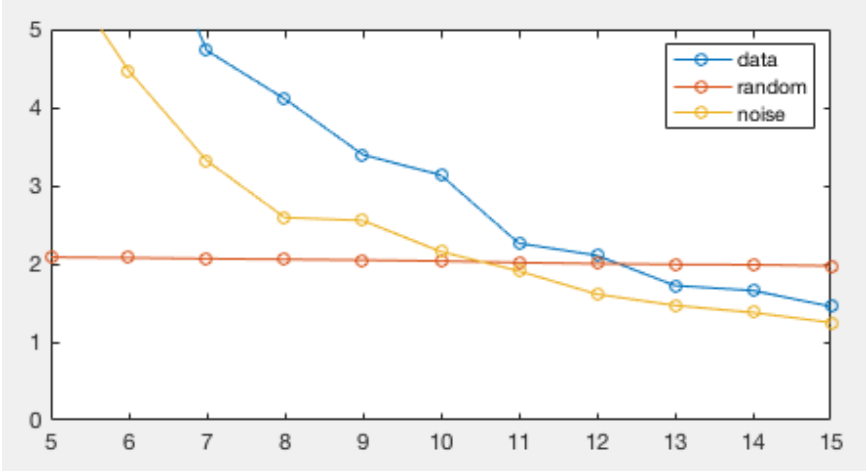


Figure 13: Scree plot with Parallel Test. The parallel test compared the factors extracted from the original data to those from a randomized dataset with the same dimensions. The plot suggests retaining 12 temporal factors (up to which the blue curve is above the red curve).

The 12 factors altogether accounted for 94% of the total variance of the data. To determine the temporal factors that reflected an MMN, I first selected among the 12 temporal factors the ones individually accounting for more than 6% of the total

variance. The first five temporal factors were thus selected: The first temporal factor (TF1) had an energy distribution peaking at 384ms and accounted for 28% of the total variance; the second temporal factor (TF2) peaked at 656ms and accounted for 26% of the total variance; the third temporal factor (TF3) peaked at 528ms and accounted for 10% of the total variance; the fourth temporal factor (TF4) peaked at 224ms and accounted for 7% of the total variance; the fifth temporal factor (TF4) peaked at 296ms and accounted for 6% of the total variance.

The temporal factor to retain was determined by considering the peak latency and the topography at the peak latency (Figure 14). The MMN exhibits a peak latency between 100-300ms and frontocentral negativity. Following this criterion, TF1, TF2, and TF3 were discarded because their peak latencies fell outside the pre-defined time window of 100-300ms. TF5 was also discarded as the topography at its peak latency showed a frontocentral positivity. TF4 had a peak latency at 224ms and exhibited a frontocentral negativity and was thus retained as a latent temporal factor for the MMN.

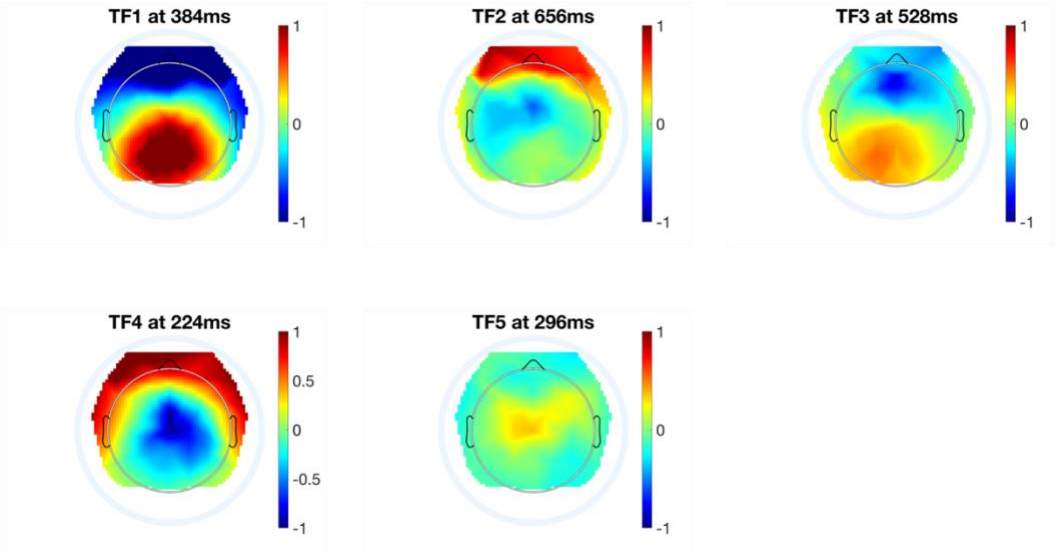


Figure 14: Topography at temporal factor's peak latency.

Having determined TF4 to retain, I then moved on to determine the time window for analysis by selecting time points with a factor loading over 0.6 in TF4. This step yielded a time window of 192-256ms, which was taken as the time window for measuring the MMN.

For the spatial region, I selected the channel (E65) that had peak negativity at the peak latency (at 224ms) of TF4 and its surrounding channels, resulting in 8 channels: E4, E7, E15, E16, E21, E51, E54, E65. Figure 15 shows the position of the selected channels on the layout of a 64-channel HydroCel Geodesic Sensor Net.

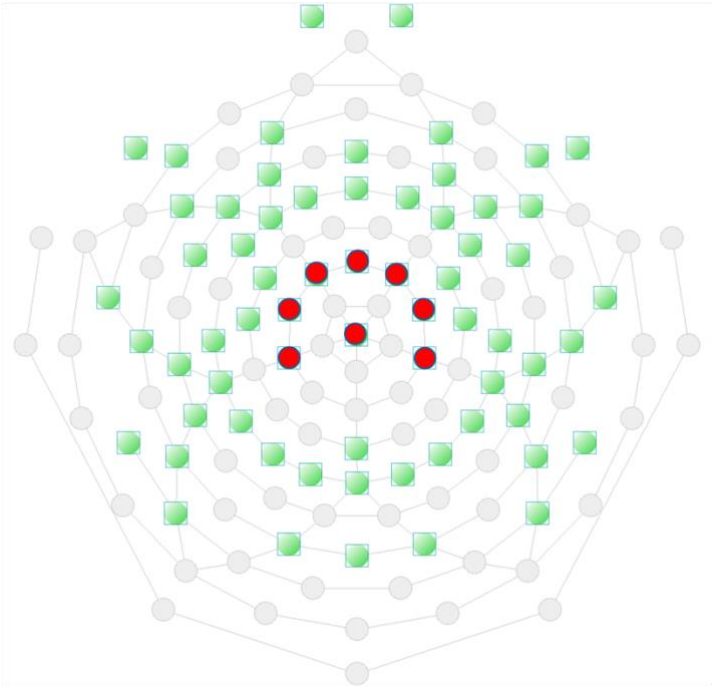


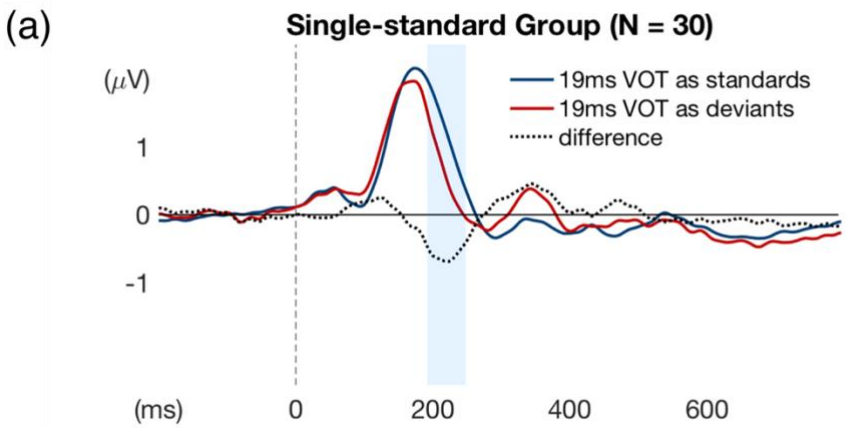
Figure 15: Position of the selected channels in 64-channel HydroCel Geodesic Sensor Net. The eight selected channels are: E4, E7, E15, E16, E21, E51, E54, E65.

To summarize, our PCA solution resulted in a time window of 192-264ms and eight frontocentral channels for measuring the MMN magnitude.

3.3.2.1.2 Statistics

The identity MMN for the across-category contrast was measured by the difference between the 19ms VOT [dæ] serving as standards in the control block and the same 19ms VOT [dæ] serving as deviants in the across-category oddball blocks. The magnitude of the MMN was determined by the ERP amplitudes averaged over the 192-256ms time window and the eight frontocentral channels. Figure 16 shows the

waveforms (averaged over subjects and the eight channels) of the same 19ms VOT [tæ] serving as standards and deviants.



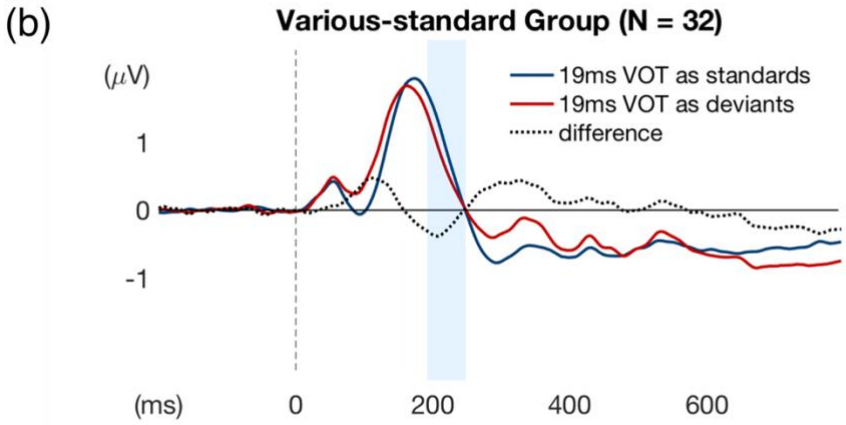


Figure 16: ERP waveforms averaged over subjects and the eight channels. Blue shaded area indicates the time window for analysis (192-256ms).

Figure 17 presents a violin plot showing the ERP averaged over the selected time window and channels for each subject and each condition.

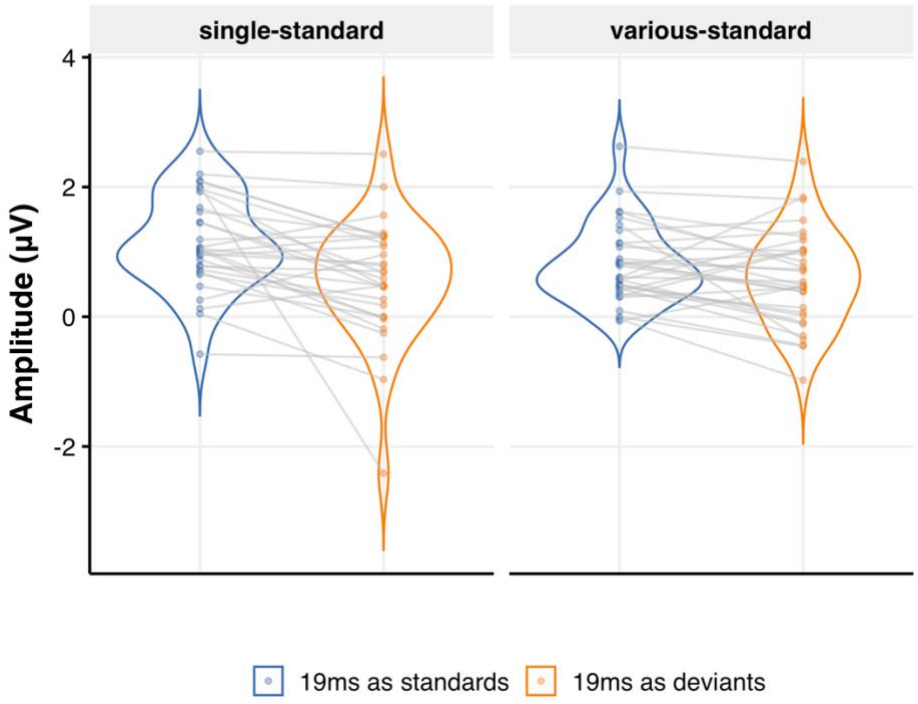


Figure 17: Individual ERPs averaged over the selected time window and the selected channels as a function of condition. Each dot represents one subject's data for a given condition. The grey line connects two data points from the same subject, indicating the amplitude change between standards and deviants. The shape of the violin plots indicates the data distribution.

For the statistical analysis, I used the averaged ERP amplitude as the dependent measure. I built a mixed-effects model with three fixed factors: Group (single-standard vs. varying-standard), Stimulus (standard vs. deviant), and the

interaction Group × Stimulus. The model also included Subject as a random intercept, as indicated by the following R pseudocode:

$$\text{Amplitude} = \text{Group} + \text{Stimulus} + \text{Group} \times \text{Stimulus} + (1|\text{Subject})$$

The model explained a substantial amount of total variance (conditional R^2 = 0.55). However, the part explained by the fixed factors alone (marginal R^2) is only 0.08. This is due to the large inter-subject variability shown in the above violin plot. I report below the fixed factors' coefficients, the corresponding standard errors (SE), 95% confidence intervals (CI), t values, and p values.

Table 4: Model summary

Fixed factors	**Coefficient**	**SE**	**95% CI**	**t(118)**	**p**
(Intercept)	0.78	0.08	[0.61, 0.94]	9.17	< .001***
Stimulus	-0.40	0.10	[-0.59, -0.21]	-4.13	< .001***
Group	-0.15	0.17	[-0.49, 0.18]	-0.88	0.380
Stimulus × Group	0.36	0.19	[-0.03, 0.74]	1.85	0.067

Since the analysis only involved the across-category blocks, both the single-standard group and the varying-standard group should show an MMN response, whether the memory trace contains gradient information or not. That is, there should be smaller deviant amplitudes than standard amplitudes in both conditions – a main effect of Stimulus. Since I coded the two levels of Stimulus with an orthogonal contrast, a main effect could be reflected as a coefficient significantly different from 0. Indeed, the coefficient of Stimulus is significant [t(118) = -4.13, p < .001]. The partial eta squared (η_p^2) associated with the effect is 0.13, indicating a medium effect size. To further examine the difference between the standard ERP and the deviant ERP for

each group, I ran one-tailed t-tests on the effect of Stimulus within each level of Group. There was a large effect of Stimulus in the single-standard group [$t(60) = 4.16$, $p < 0.001$, Cohen's $d = 0.54$], while the effect of Stimulus was marginal in the varying-standard group [$t(60) = 1.64$, $p = 0.053$, Cohen's $d = 0.21$]. Neither the effect of Group nor the interaction (Stimulus × Group) reached significance. Figure 18 shows the bar plot of the averaged ERP for each condition.

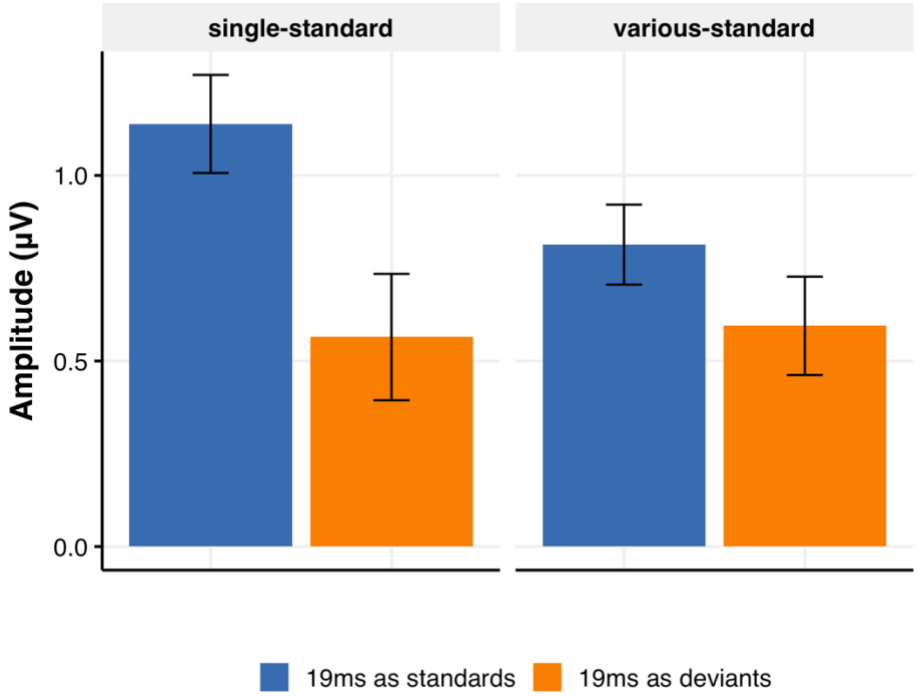


Figure 18: ERP amplitude averaged over subjects, selected time window and channels. Error bar indicates standard error.

As expected, we found the MMN when the standards and deviants belong to different categories. This gave the evidence that our experimental design and materials were valid, and we could continue to examine the within-category contrast.

3.3.2.2 Within-category MMN

3.3.2.2.1 PCA solution

Following the PCA procedure specified in 3.2.6.1, I first ran a temporal PCA with a Kaiser weighting (kappa = 3). Using a scree plot in combination with a Parallel Test, 13 temporal factors were retained as they accounted for more variance than the factors extracted from the randomized dataset (Figure 19).

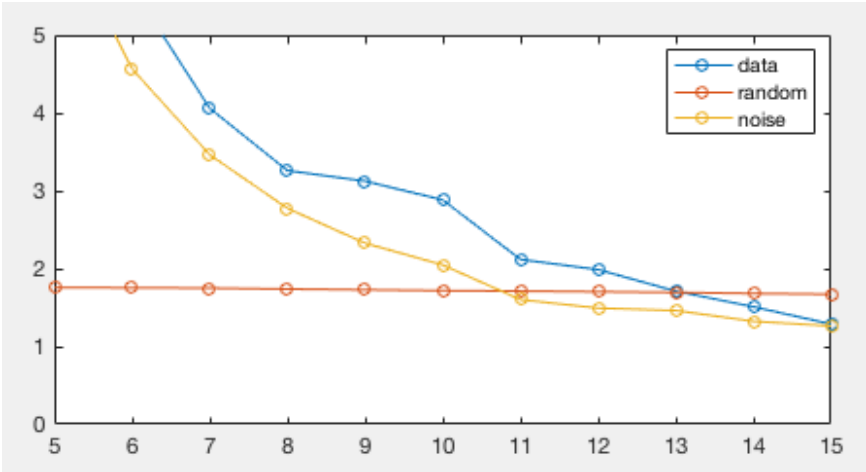


Figure 19: Scree plot with Parallel Test. The parallel test compared the factors extracted from the original data to those from a randomized dataset with the same dimensions. The plot suggests retaining 13 temporal factors (up to which the blue curve is above the red curve).

The 13 factors altogether accounted for 94% of the total variance of the data. Among the 13 temporal factors, I selected the ones individually accounting for more than 6% of the total variance to be the potential factors reflecting an MMN effect. The first five temporal factors were thus selected: The 1st temporal factor (TF1) had an energy distribution peaking at 540ms and accounted for 24% of the total variance; the 2nd temporal factor (TF2) peaked at 784ms and accounted for 23% of the total variance; the 3rd temporal factor (TF3) peaked at 360ms and accounted for 17% of the total variance; the 4th temporal factor (TF4) peaked at 216ms and accounted for 8% of the total variance; the 5th temporal factor (TF5) peaked at 112ms and accounted for 7% of the total variance

To determine which temporal factor to retain, I considered the latency of the peak time as well as the topography at the peak latency (Figure 20). The temporal factor reflecting an MMN response should have a peak latency falling in the pre-defined 100-300ms time window and have a frontocentral negativity. Following this criterion, TF1, TF2, and TF3 were discarded because their peak latencies fell outside the pre-defined time window of 100-300ms. TF5 was also discarded as the topography at its peak latency showed a frontocentral positivity. TF4, peaking at 216ms, featured a central negativity and was retained as a latent temporal factor for the MMN.

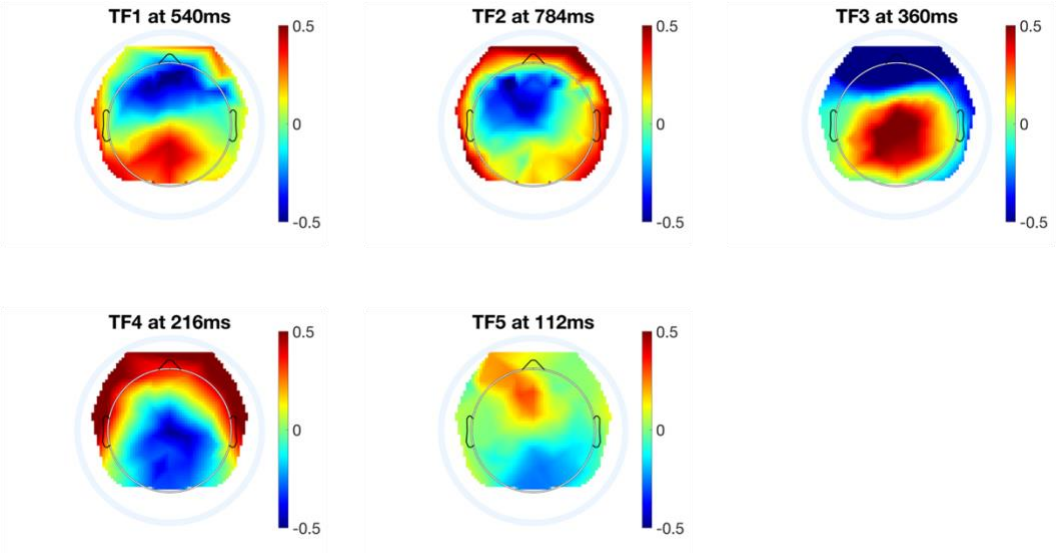


Figure 20: Topography at temporal factor's peak latency.

For the analysis time window, I selected the time points with a factor loading over 0.6 in TF4. This step yielded a time window of 176-248ms. I took this time window as the time window of the MMN response.

For the spatial region, I selected the channel (E65) that showed the most negative peak at the peak latency (at 216ms) of TF4 as well as its surrounding channels, which resulted in 8 channels: E4, E7, E15, E16, E21, E51, E54, E65. Figure 21 shows the position of the selected channels on the layout of a 64-channel HydroCel Geodesic Sensor Net.

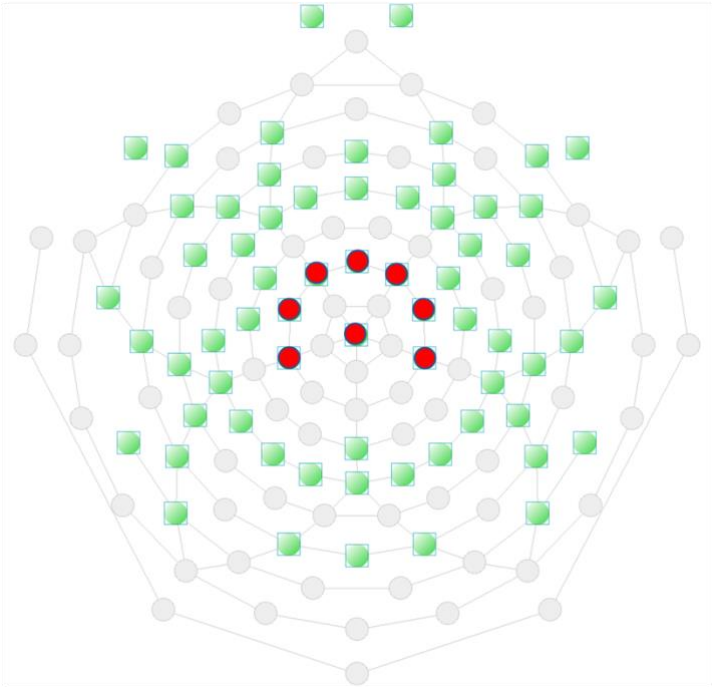


Figure 21: Position of the selected channels in 64-channel HydroCel Geodesic Sensor Net. The eight selected channels are: E4, E7, E15, E16, E21, E51, E54, E65.

To summarize, our PCA solution resulted in a time window of 176-248ms and eight frontal-central channels for the MMN analysis.

3.3.2.2.2 Statistics

The within-category identity MMN was measured by the difference between the 119ms VOT [tæ] serving as standards in the control block and the same 119ms VOT [tæ] serving as deviants in the within-category oddball blocks. The magnitude of the MMN was determined by the ERP averaged over the 176-248ms time window and the eight frontocentral channels. Figure 22 shows the waveforms (averaged over

subjects and the eight channels) of the same 119ms VOT [tæ] serving as standards and deviants.

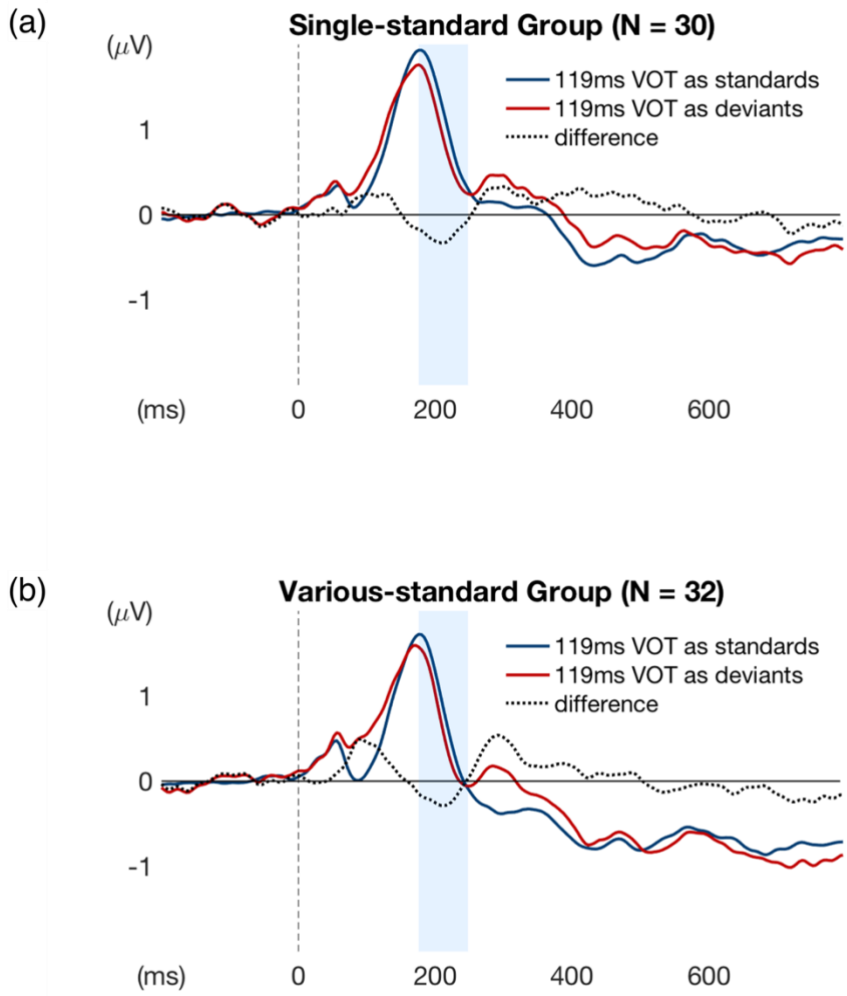


Figure 22: ERP waveforms averaged over subjects and the eight channels. Blue shaded area indicates the time window for analysis (176-248ms).

Figure 23 presents a violin plot showing the ERP averaged over the selected time window and channels for each subject and each condition.

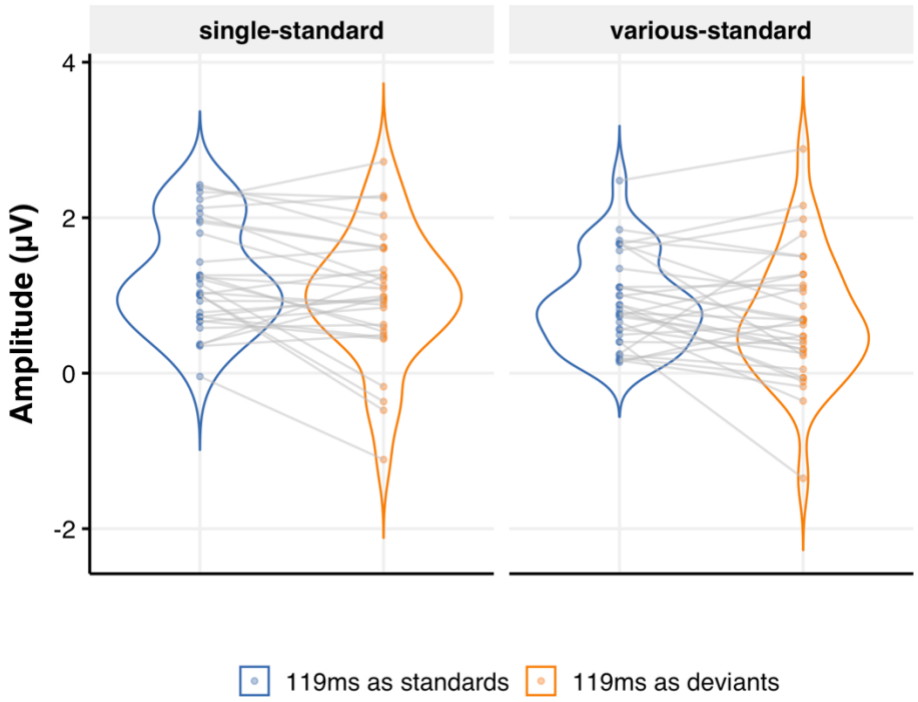


Figure 23: Individual ERPs averaged over the selected time window and the selected channels as a function of condition. Each dot represents one subject's data for a given condition. The grey line connects two data points from the same subject, indicating the amplitude change between standards and deviants. The shape of the violin plots indicates the data distribution.

For the statistical analysis, I used the ERP amplitude as the dependent measure. The mixed-effects model with three fixed factors: Group (single-standard vs. varying-standard), Stimulus (standard vs. deviant), and the interaction Group ×

Stimulus. The model also included Subject as a random intercept as indicated by the following R pseudocode:

$$\text{Amplitude} = \text{Group} + \text{Stimulus} + \text{Group} \times \text{Stimulus} + (1|\text{Subject})$$

Our fixed factors and the random factor together explained a substantial amount of total variance (conditional $R^2 = 0.67$), but the part explained by the fixed factors alone (marginal R^2) is only 0.06. This is again due to the large inter-subject variability, as shown in the above violin plot. I report below the fixed factors' coefficients, the corresponding standard errors (SE), 95% confidence intervals (CI), t values, and p values.

Table 5: Model summary

Fixed factors	**Coefficient**	**SE**	**95% CI**	**t(118)**	**p**
(Intercept)	0.96	0.09	[0.79, 1.13]	11.17	< .001***
Stimulus	-0.22	0.08	[-0.38, -0.07]	-2.82	0.006**
Group	-0.30	0.17	[-0.64, 0.04]	-1.76	0.080
Stimulus × Group	0.04	0.16	[-0.27, 0.35]	0.26	0.792

Note that both Stimulus and Group contained only two levels and that I applied an orthogonal contrast coding, so we could evaluate the effect of each fixed factor by examining whether its coefficient is significantly different from 0. The above table shows that the coefficient of Stimulus is significantly different from 0 [t(118) = -2.82, p = .006], suggesting a main effect of Stimulus. The partial eta squared (η_p^2) associated with the effect is 0.06, indicating a medium effect size. No other main effect or interaction was found. In line with the main effect of Stimulus, the bar plot in figure 24 shows that the mean deviant amplitude is smaller than the mean standard

amplitude in both groups. As planned tests, I ran one-tailed t-tests on the effect of Stimulus within each level of Group. An MMN response with a small effect size was found in both the single-standard group [t(60) = 2.143, p = 0.018, Cohen's d = 0.28] as well as the varying-standard group [t(60) = 1.834, p = 0.036, Cohen's d = 0.24].

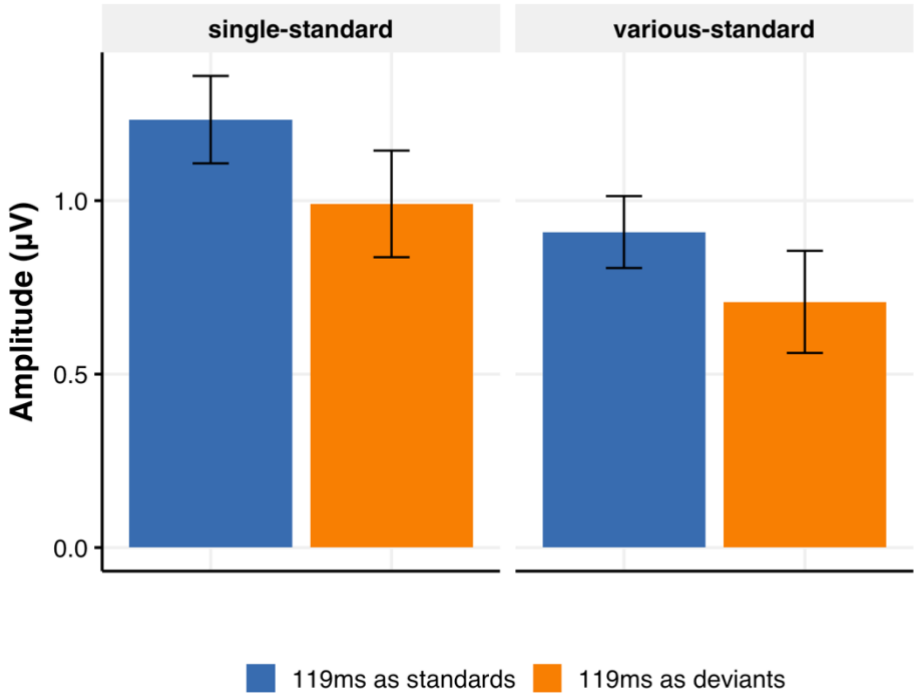


Figure 24: ERP amplitude averaged over subjects, selected time window and channels. Error bar indicates standard error.

3.3.2.3 Post-hoc analysis

So far, we have obtained an MMN response in both the single-standard group and the varying-standard group, whether the standard-deviant contrast is across-categorical or within-categorical. As a post-hoc analysis, I compared the size of the

MMN responses obtained in each condition to explore the effect of the standard type and the standard-deviant contrast on the MMN size (Figure 25).

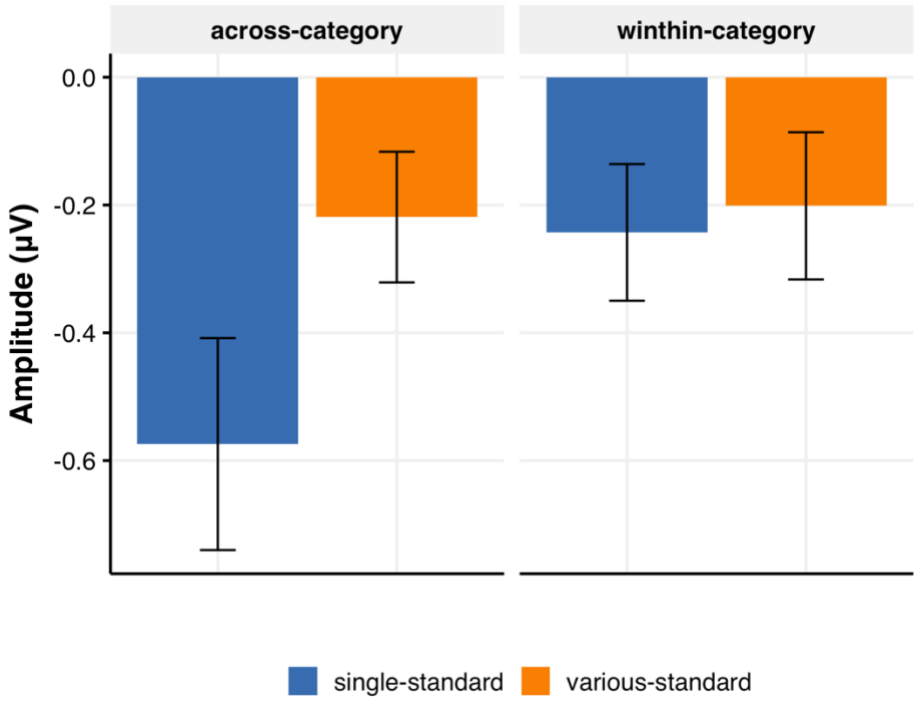


Figure 25: Mean MMN amplitude in each condition. Error bar indicates standard error.

For the statistical analysis, I used the MMN amplitude (subtracting the standard amplitude from the deviant amplitude) as the dependent measure. The model contained three fixed factors: one between-subject factor Group (single-standard vs. varying-standard), one within-subject factor Category (across-category vs. within-

category), and their interaction Group × Category. The model also included Subject as a random intercept, as indicated by the following R pseudocode:

$$\text{MMN} = \text{Group} + \text{Category} + \text{Group} \times \text{Category} + (1|\text{Subject})$$

The model's total explanatory power is moderate (conditional $R^2 = 0.25$). The part related to the fixed factors alone (marginal R^2) is 0.05. I report below the fixed factors' coefficients, the corresponding standard errors (SE), 95% confidence intervals (CI), t values, and p values.

Table 6: Model summary

Fixed factors	**Coefficient**	**SE**	**95% CI**	**t(118)**	**p**
(Intercept)	-0.31	0.07	[-0.44, -0.17]	-4.52	< .001***
Category	0.17	0.11	[-0.04, 0.39]	1.58	0.116
Group	0.20	0.14	[-0.07, 0.47]	1.45	0.150
Category × Group	-0.31	0.22	[-0.75, 0.12]	-1.42	0.157

The intercept is significantly smaller from 0, which is expected as an MMN response (negativity) was observed across conditions. None of the fixed factors reaches the significance level. However, Figure 25 shows that the across-category MMN elicited in the single-standard group is much larger than the MMNs elicited in the other conditions. This pattern becomes significant if we look at pairwise comparisons between every two conditions: the across-category MMN in the single-standard group is significantly larger than the across-category MMN in the varying-standard group [t(115) = -2.021, p = 0.046], the within-category MMN in the single-standard group [t(60) = -2.093, p = 0.041], and the within-category MMN in the varying-standard group [t(115) = -2.121, p = 0.036]. It should be the noted that the

above significance results were obtained with uncorrected p values. Correcting for multiple comparison eliminates the significance results. Below I discuss the possible interpretations for the current results.

3.4 Discussion

3.4.1 Summary of the current experiment

In the time window and the channels delimited by the PCA solutions, a significant MMN response was observed in both the single-standard group and the varying-standard group, whether or not the standards and the deviants belonged to the same category. An across-category MMN (48ms VOT vs. 19ms VOT) is not surprising. Nonetheless, a within-category MMN in the single-standard group (48ms VOT vs. 119ms VOT) suggests that without a categorical difference, the acoustic difference between standards and deviants could still lead to an MMN response. Previous studies typically found a lack of MMN response for a within-category contrast (e.g., Silva et al., 2017). The discrepancy between our results and the previous results might be due to the difference in the stimuli or the analysis techniques. But more likely, it could be that the acoustic difference between the standards and the deviants in the current design was larger than that in the previous studies. The acoustic difference (48ms VOT vs. 119ms VOT) in the current study passed the perceptual threshold and thus led to an MMN. Note that the current experiment had a larger sample size than the previous studies. It is possible that the previous studies simply did not have enough power to detect a small within-category effect.

The current experiment aimed to determine whether the memory trace in the varying-standard paradigm contains gradient information. If the memory trace in the

current experiment is solely categorical, the deviant would not contrast the categorical information in the memory trace as the deviant and standards belong to the same category [t], and thus no MMN is expected. Alternatively, if the memory trace contains gradient information, the memory trace would contain not only the category information ([t]) but also the VOT information, whether it is from the proximal VOT values or the phonetic knowledge about the VOT realizations of a word-initial [t]. If the VOT information is from the proximal stimuli, the brain could simply compute a statistical summary of the VOT values of the proximal stimuli and use that statistical summary as the information in the memory trace used to generate MMN. In the present case, the VOTs of the standards formed a uniform distribution with a mean of around 48ms. Given that we have observed the MMN in the single-standard group with a fixed 48ms standard VOT and a 119ms deviant VOT, it is not surprising that the contrast between the 48ms VOT (the mean VOT value) and the 119ms deviant VOT in the varying-standard group would lead to an MMN. On the other hand, if the gradient information comes from the phonetic knowledge of the category [t], the representation could be a prototype [t] with a 60ms VOT, or a probability distribution with a mean of 60ms. Based on that phonetic knowledge, the brain would predict a [t] with a VOT at around 60ms. Since the deviant VOT is 128ms, it is inconsistent with prediction, and an MMN would also be expected. Indeed, a significant within-category MMN was found in the varying-standard group, suggesting that even with varying standards, the MMN can be generated based on acoustic/phonetic information. Below I discuss the patterns specific to the current experiment.

3.4.2 Prediction error and uncertainty

In the post-hoc analysis, we observed a smaller within-category MMN than the across-category MMN in the single-standard group. This pattern is expected because the across-category MMN has two sources: a categorical contrast ([t] vs. [d]) and an acoustic contrast (48ms vs. 19ms). In contrast, the within-category MMN can only source from an acoustic contrast (48ms vs. 119ms).

For the across-category MMN, we observed a larger (more negative) amplitude in the single-standard group than in the varying-standard group. This pattern can be explained by the Predictive coding framework. The framework states that the brain constructs a model based on the regularity of the presented stimuli and makes predictions based on the model. A prediction error occurs when there is a difference between what is predicted and what is encountered. The prediction error is then fed into the perceptual network and triggers a model update. According to the Predictive coding framework, an MMN reflects an update of the model when there is a prediction error (Friston, 2005). Nonetheless, a prediction error is not the only factor that decides whether the model will get updated. There is also uncertainty associated with the prediction the brain makes. If a prediction is made with high uncertainty, even a large prediction error might not cause the model to update and would thus result in minimal MMN. In contrast, with low uncertainty, even a tiny prediction error can drive the model to update, leading to a robust MMN (Auksztulewicz & Friston, 2016). One source of uncertainty is the variability of the environment. A low variability leads to a low uncertainty.

Regarding the across-category MMN responses, the prediction error came from a categorical contrast and an acoustic contrast between the standards and the deviants; the uncertainty came from the variability of the presented stimuli. The

categorical contrast (standard [t] vs. deviant [d]) was the same in both groups. Their acoustic contrasts were similar in terms of the VOT difference between the mean VOT of standards (~48ms) and the mean VOT of deviants (19ms). Therefore, the single-category group and the varying-standard group shared the same prediction error. However, the two groups differed in the uncertainty: in the single-standard group, the fixed standards (48ms) together with the deviant led to a low variability and thus a low uncertainty associated with the prediction; in the varying-standard group, the various standards (42, 48, and 55ms) together with the deviant led a higher variability and thus a higher uncertainty. As a result, even with the same prediction error, a lower uncertainty in the single-standard group led to a larger MMN response compared to the varying-standard group. Previous studies also found a reduced MMN amplitude when standards were varied (Daikhin & Ahissar, 2012; Korzyukov, Winkler, Gumenyuk, & Alho, 2003; Kujala et al., 2007; Paavilainen et al., 2001). The effect of the variability of the standards holds even when the dimension of the variability is independent of the dimension of the deviance: Winkler et al. (1990) contrasted 650Hz deviant tones to 600Hz standard tones, which varied in intensity. The amplitude of the frequency MMN reduced as the intensity variability increased.

For the within-category MMN, the amplitudes in the single-standard group and the varying-standard group were comparable. Here we do not observe the effect of standards variability, which would otherwise reduce the MMN amplitude in the varying-standard group. Given the small to medium effect size of the within-category MMN, it is possible that the large inter-subject variability masked the even subtler effect of the variability of the standards. Besides the inherent property of a within-category contrast being more difficult to discriminate, there could be three reasons

why the MMN effect size in the current within-category contrast was small. First, the current experiment always presented the roving-standard control block as the first block, which repeated the 19ms and 119ms stimuli serving as deviants in the following oddball blocks. The repeated exposure to the deviant stimuli in the control block might have attenuated the MMN responses in the following oddball block. This primacy effect has been reported in (Hestvik & Durvasula, 2016). Second, the standard-deviant contrast in the control block (19ms vs. 119ms) was across-category and more salient than that in the following oddball within-category blocks (~48ms vs. 119ms). Liu et al. (2022) found a significantly reduced MMN when subjects were previously exposed to a standard-deviant contrast with a larger acoustic difference. Third, the standard-deviant contrast in the control block contained a categorical change (/d/ vs. /t/) which is missing in the following oddball blocks (/t/ vs./t/). Lipski and Mathiak (2008) found that an across-category contrast presented earlier would suppress a later MMN response to a within-category contrast. All these factors together might result in a small MMN effect.

3.4.3 Phonetic knowledge or statistical summary?

A significant within-category MMN in the varying-standard group supports a memory trace containing gradient information. However, the current experiment cannot determine the source of the gradient information. To distinguish whether the observed MMN response was driven by the phonetic knowledge stored in the long-term memory or a statistical summary of the proximal stimuli' acoustic properties, I conducted the second experiment.

Chapter 4

EXPERIMENT 2: PHONETIC KNOWLEDGE VERSUS STATISTICAL SUMMARY

Experiment 1 found a mismatch negativity (MMN) response to the within-category contrast with a varying-standard paradigm. The results of Experiment 1 support a memory trace retaining gradient information along with a category representation. The gradient information could come from the phonetic knowledge associated with the category, which might be a prototype or a probability distribution with a mean of 60ms. Alternatively, the subjects could have formed a representation based on a statistical summary of proximal stimuli's acoustic properties. In the present case, the statistical summary could be a mean and a standard deviation (SD) of the standard voice onset time (VOT) values. Note that the VOT values of the standards in Experiment 1 were uniformly presented as 42, 48, and 55ms. Thus, the gradient information could be a statistical summary of a uniform distribution with those three VOT values. Both types of gradient information would lead to a prediction different from the deviant [t] carrying a VOT of 119ms. Therefore, a question remains: Is the MMN in Experiment 1 driven by the phonetic knowledge retrieved from the long-term memory, or is it driven by a statistical summary of the proximal stimuli? Experiment 2 aims to distinguish between these two possibilities.

Previous studies have found that the MMN response can reflect the brain's implicit tracking of the statistics in the proximal stimuli. Garrido et al. (2013) presented subjects with tones of different frequencies. The tone frequencies were

drawn from one of the two Gaussian distributions, which shared a mean of 500Hz. One distribution was narrower, with an SD of around 700Hz (0.5 octaves above the mean); the other distribution was wider, with an SD of around 1400Hz (1.5 octaves above the mean). Each distribution was interspersed with additional 100 standard tones of 500Hz and 100 deviant tones of 2000Hz (2 octaves above the mean). The subjects were asked to press a button to a luminance change that coincided with each of the 100 standards and the 100 deviants. The behavioral results showed that a deviant tone facilitated recognizing the luminance change, reducing the reaction time. In particular, the facilitation effect was larger when the presented tones were drawn from a narrower Gaussian distribution compared to when the tones were drawn from a wider distribution. More importantly, the MMN response exhibited the same pattern: the 100 deviants elicited a larger MMN in the narrower Gaussian distribution than in the wider Gaussian distribution. The results suggested that the brain was tracking the statistical structure of the presented tones and interpreting the deviants in the narrower distribution as being more unlikely than the deviants in the wider distribution. To our knowledge, no study has tested whether the same pattern would emerge with speech stimuli, which could lead to a categorical representation (e.g., phoneme) that is absent in nonspeech stimuli. Our Experiment 2 will replicate Garrido et al. (2013)'s design. However, instead of manipulating the tone frequency, Experiment 2 focuses on the VOT. The logic of the experiment design is as follows: If the observed within-category MMN was due to the phonetic knowledge retrieved from the long-term memory, the MMN magnitude should be insensitive to the manipulation of the statistical structure of the proximal standard stimuli VOTs; on the other hand, if the

MMN was driven by a statistically summary of the proximal standard stimuli, the MMN magnitude should be modulated by the way the standard VOTs are presented.

4.1 Methods

4.1.1 Participants

Thirty-five subjects from the University of Delaware participated in the experiment. All subjects were monolingual English speakers aged 18 to 35 (23 females, mean age = 21, SD = 3) and reported no history of language impairment. Subjects received either $20 or extra credits for completing the experiment. The experiment procedure was approved by the University of Delaware Internal Review Board and was compliant with the principles for ethical research established by the Declaration of Helsinki.

4.1.2 Stimuli

The stimuli used in the current experiment were extracted from the stimulus set created in Experiment 1. Readers are referred to 3.2.2.1 for detailed procedures and parameters for creating the stimuli. To recap, the stimulus set included 146 resynthesized CV syllables along the /dæ-tæ/ continuum, with the onset VOT ranging from 0ms to 145ms. In addition, a portion of the 48ms VOT syllable was replaced by a 440Hz pure tone to serve as a target stimulus. The target stimulus was not included in the analyses.

To test the effect of the proximal stimuli's statistical structure on the MMN amplitude, I created two distributions for the proximal standard VOTs and manipulated the SD. Garrido et al. (2013) generated the distribution of tone frequencies using the binary logarithmic (log2) scale. Experiment 2 also adopted this

design. The idea was to first create a frequency distribution of VOT values in the log2 scale and convert them back to the linear scale. To make the VOT values easy for the logarithmic-linear conversion, the mean of the two distributions were determined to be 6 (which is $2^6 = 64$ms in the linear scale), approximating the empirical mean (60ms) of the VOTs for a word-initial /t/ (Chodroff & Wilson, 2018). The SD values were set to 0.33 for a wide distribution and 0.11 for a narrow distribution. The selection of the SD values was based on the following two considerations. First, the ratio between the narrow-distribution SD and the wide-distribution SD was set to match the ratio in Garrido et al. (2013). Second, since the VOT range that facilitates a /t/ interpretation is more restricted than the tone range, the VOT range in the narrow distribution should not cover overly short VOTs which would lead to a /d/ interpretation, and the VOT range in the wide distribution should not cover overly long VOTs which sound nonspeech-like. To select the VOT values that conform to the narrow and wide distributions, I used a MATLAB script to sample 840 values (see 4.1.3 for how the number was determined) from a normal distribution with the defined mean (6) and SDs (0.33 and 0.11) in the log2 scale. Each sampled value was then converted to the linear scale and rounded to the nearest integer. VOT values above 145 were replaced with 145. The resulting SDs in the linear scale were 5 for the narrow distribution and 15 for the wide distribution. Figure 26 shows the two distributions in both the log2 scale and the linear scale.

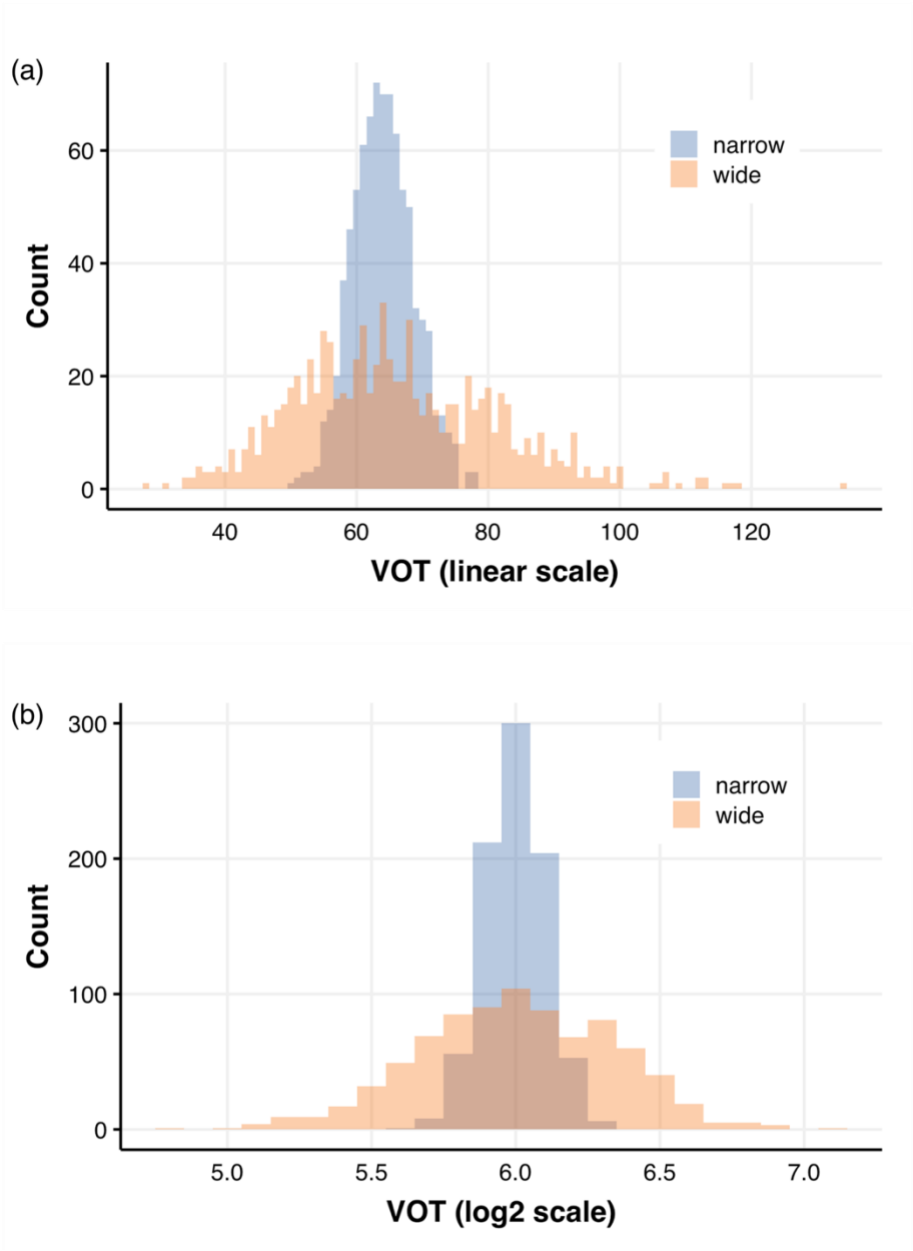


Figure 26: Frequency distribution of the 840 VOT values in the narrow distribution (blue) and the wide distribution (yellow). (a) VOT values on the linear scale. (b) VOT values on the log2 scale.

For each distribution, besides the 840 VOT values that function as part of standards, there were another 105 standards with a 64ms VOT (equal to the distribution mean) and 105 deviants with a 128ms VOT. In the log2 scale, the deviant VOT is 3SD away from the mean in the wide distribution and 9SD away from the mean in the narrow distribution. Those 100 deviants were compared to the 100 standards of a 64ms VOT to compute the MMN.

4.1.3 Design

The statistical distribution examined was a between-subject variable: Group. The Group variable included two levels: a narrow-distribution group and a wide-distribution group, respectively corresponding to the narrow distribution and the wide distribution. For each group, the number of deviants was fixed to 105. Each deviant was preceded by a train of standards with a count being any of the following seven values: 3, 5, 7, 9, 11, 13, 15. Each count was associated with 15 trains of standards, giving $(3 + 5 + 7 + 9 + 11 + 13 + 15) \times 15 = 945$ standards. Note that the 945 standards included the 840 standards I drew from a normal distribution and the 105 standards used for computing an MMN response. The standards-to-deviants ratio is 9:1. After the stimulus list of standards and deviants for each distribution group was determined, I randomly interspersed each list with 32 target stimuli. For each subject, I presented the same distribution twice as two separate blocks. This manipulation also allows us to assess the possible effect of the block order on the MMN amplitude. Therefore, the experiment included two fully crossed independent variables: a between-subject variable Group and a within-subject variable Block.

Note that the current experiment did not compute an identity MMN by comparing the deviant stimulus to itself serving as standards. This is because the focus

was on the amplitude difference between the two MMNs instead of the MMN amplitude itself. In both distributions, the MMN was computed by subtracting the ERP of the 64ms-VOT standards from the ERP of the 128ms-VOT deviants. Therefore, any brain response attributable to a stimulus-specific property would be equivalent in both MMNs and would be canceled out when I compared the two MMNs. Figure 27 illustrates the experimental design and the MMN computing strategy.

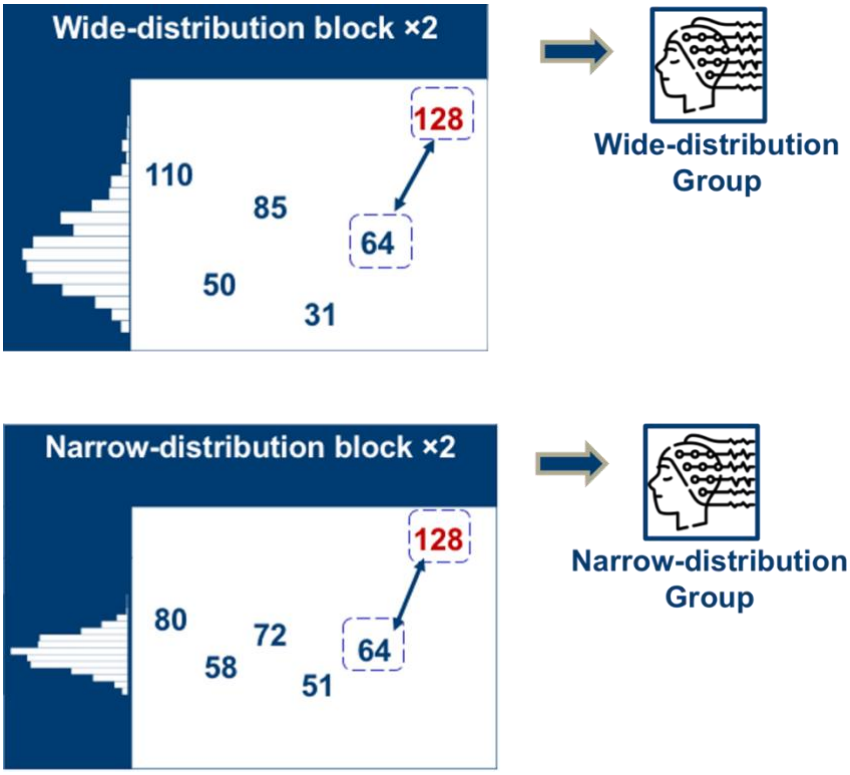


Figure 27: Illustration of the experimental design and how I compute a non-identity MMN. Each subject completed either two wide-distribution blocks or two narrow-distribution. In each block, standard VOT values are in blue, and deviant VOT values are in red. To compute a non-identity MMN, I subtract the ERP averaged over the 105 standards from the ERP averaged over the 105 deviants in the same block (double-sided arrows).

If the MMN in Experiment 1 was driven by a statistical summary based on the proximal stimuli, then the narrow-distribution group should exhibit a larger MMN amplitude than the narrow-distribution group, as shown in Garrido et al. (2013). On the contrary, if the MMN in Experiment 1 was indeed driven by the phonetic knowledge retrieved from the long-term memory, the MMN amplitude should not be modulated by the statistical structure of the proximal stimuli.

4.1.4 Procedure

The procedure of Experiment 2 was similar to that in Experiment 1. Before the EEG recording, participants went through a practice session. They were informed to press a button with their most frequently used hand once they heard a target sound containing a beep. A piece of a jigsaw puzzle would show on the screen after each target sound. Depending on whether they made the response and how fast the response was, the picture would be either colored or in greyscale.

Each subject was assigned to either the narrow-distribution group or the wide-distribution group, depending on their assigned subject ID. Each subject completed two identical blocks, each lasting about 25 minutes. One experiment session took about 1.5 hours, including the EEG net placement, instruction, breaks, and the EEG net removal.

4.1.5 Apparatus, data acquisition, and data processing

The experiment was programmed using E-Prime 2 and the Net Station E-prime extension package for EEG acquisition. The continuous EEG data were recorded with the 64-channel HydroCel Geodesic Sensor Nets. Before the EEG recording, the impedance of each channel was lowered to below 50kΩ. During the EEG recording,

the incoming analog signals underwent an online 125 Hz low-pass filter to prevent aliasing and were digitized with a sampling rate of 250 Hz. Subjects' electro-ocular activity was recorded from 4 bipolar channels around the eyes. Channel E65 (corresponding to Cz in the 10-10 system), placed on the vertex of the scalp, was used as a reference channel.

The recorded data were passed through a first-order high-pass filter of 0.1 Hz to remove slow drifts. Then a finite impulse response low-pass filter of 40 Hz with a roll-off of 2 Hz was applied to the data to remove line noise and any frequency above 40 Hz. From the filtered data, I segmented out the 105 64-ms VOT standards and the 105 128-ms VOT deviants for analysis. Each segment was 1000ms long with a 200ms pre-stimulus period and had the 0ms time-locked to the stimulus onset. The 200ms pre-stimulus period was also used as a baseline. The segmented data were submitted to an automated process of eyeblink subtraction using ICA with the ERP PCA toolkit. An eyeblink template was automatically generated for each subject. An ICA component was marked as an eyeblink component and was subtracted from the data if it was correlated at r = .9 or greater with the eyeblink template. After the eyeblink subtraction, the data were submitted to the artifact correction procedure to remove bad channels and movement artifacts. For each trial, a channel was marked bad if its best absolute correlation with its neighboring channels fell below r = .4 across all time points. Bad channels were replaced via a spline interpolation from surrounding good channels. If a channel was marked bad in over 20% of trials, it was considered bad in all trials. A trial was marked bad and was dropped if it contained more than 10% bad channels. Table 7 shows the mean percentage of bad trials in each condition

Table 7: Percentage of bad trials in each condition

Group	Block	Percentage of bad trials (%)	
		Standards	Deviants
Narrow-distribution	first	3.4%	3.1%
Narrow-distribution	second	2.9%	3.0%
Wide-distribution	first	4.1%	4.3%
Wide-distribution	second	2.2%	2.2%

4.1.6 Planned signal processing

4.1.6.1 Deciding time window and channels for MMN

To identify the time window and the channels for the MMN analysis, I applied a temporal principal component analysis (PCA) to decompose the data. Since both the narrow-distribution group and the wide-distribution group shared the same standard stimuli (64ms) and deviant stimuli (128ms), one temporal PCA was run for both groups. To construct the input to the PCA, I first computed one deviant ERP and one standard ERP for each subject by respectively averaging the 420 (i.e., 105 × two groups × two blocks) deviant responses and the 420 (i.e., 105 × two groups × two blocks) standard responses, collapsing groups and blocks. Then one difference ERP was derived by subtracting the standard ERP from the deviant ERP. This difference ERP was used as the input to the temporal PCA.

The PCA procedure extracted latent temporal factors from the input data. The time window and the channel group for EPR analysis was determined based on those temporal factors. Specifically, I looked through each temporal factor to examine whether the peak latency of the factor fell 100-300ms after the stimulus onset. The

time window for analysis comprised the time points centering around the peak latency and carrying a factoring loading of 0.6 or higher. The channels for analysis were chosen based on the peak channel, which should have the greatest negativity at the peak latency. For data analysis, I used the magnitude of the difference ERP (by subtracting the standard ERP from the deviant ERP). The magnitude of the difference ERP was computed by averaging the amplitudes over the selected time window and channels for each participant.

4.1.6.2 Statistical analysis

All statistical analyses were conducted using the R software and a linear mixed-effects model using the *lmer* function from the *lme4* package. The dependent measure was the mean ERP amplitude averaged over the PCA-delimited time window and the channels for each subject. The model included three fixed factors: Group (narrow-distribution vs. wide-distribution), Block (first vs. second), and the interaction between Group and Block. The model also included Subject as a random intercept. For each fixed factor, I constructed an orthogonal contrast for factor levels and used the coefficients computed with the orthogonal contrasts to indicate the overall effect of each factor. I obtained the model's explanatory power and parameter coefficients using the *report* function from the *report* package and the *model_parameters* function from the *parameters* package. For the effect size, I calculated the partial eta-squared (η_p^2).

If the MMN in Experiment 1 was merely driven by a statistical summary based on the proximal stimuli, then the MMN magnitude should be modulated by Group, that is, a main effect of Group.

4.2 Results

4.2.1 PCA solution

Following the PCA procedure specified above, I ran a temporal PCA with a Kaiser weighting (kappa = 3). To determine the factors to retain, a scree plot in combination with a Parallel Test was used to compare the factors extracted from the original data to those from a randomized dataset. Thirteen temporal factors were retained as they accounted for more variance than those extracted from the randomized dataset (Figure 28).

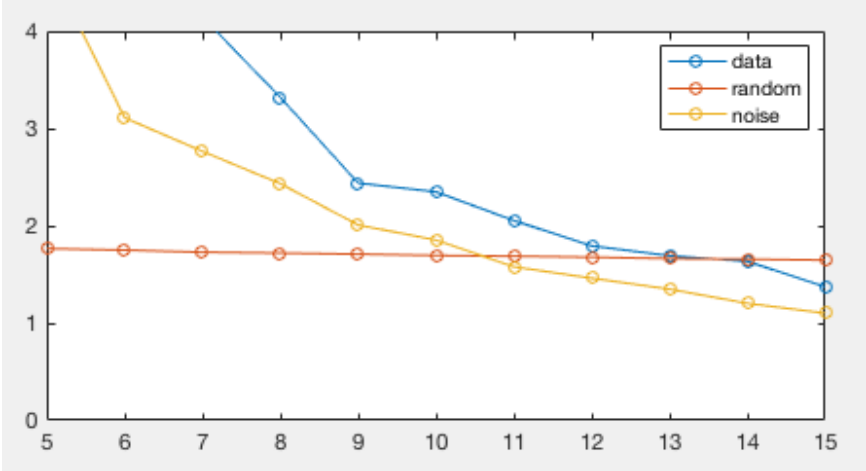


Figure 28: Scree plot with Parallel Test. The parallel test compared the factors extracted from the original data to those from a randomized dataset with the same dimensions. The plot suggests retaining 13 temporal factors (up to which the blue curve is above the red curve)

The 13 factors altogether accounted for 93% of the total variance of the dataset. To determine the temporal factors that reflected an MMN, I first selected among the 13 temporal factors the ones individually accounting for more than 6% of the total

variance. The first four temporal factors were thus selected: The first temporal factor (TF1) had an energy distribution peaking at 692ms and accounted for 29% of the total variance; the second temporal factor (TF2) peaked at 372ms and accounted for 23% of the total variance; the third temporal factor (TF3) peaked at 520ms and accounted for 12% of the total variance; the fourth temporal factor (TF4) peaked at 236ms and accounted for 8% of the total variance.

I then determined which temporal factor to retain by considering the latency of the peak time as well as the topographic map at the peak time (Figure 29). The peak latency of TF1, TF2, and T3 all fell outside the pre-defined peak time window and were thus discarded. I retained TF4, which had a peak latency of 236ms and featured a frontocentral negativity.

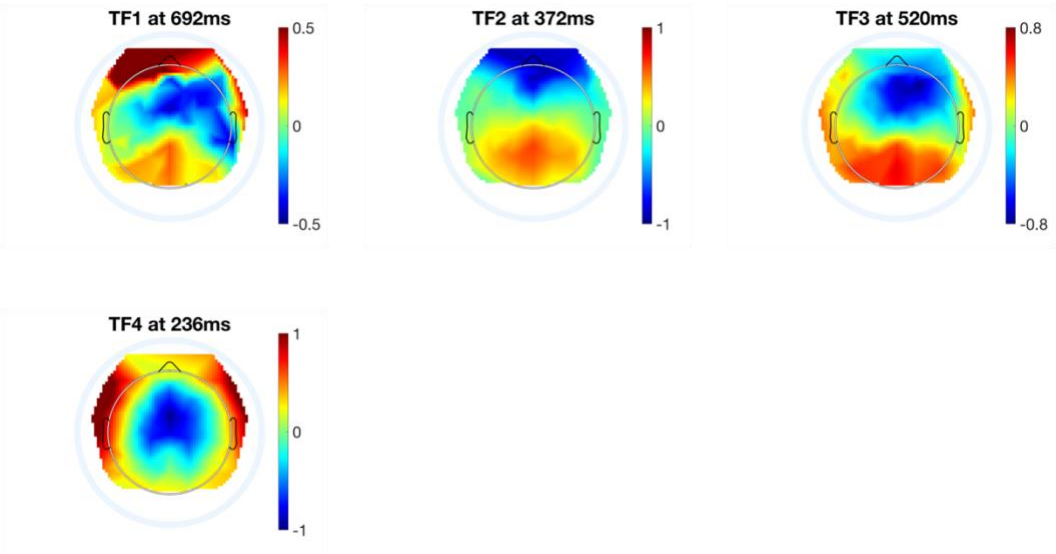


Figure 29: Topography at temporal factor's peak latency.

Having determined TF4 to retain, I then moved on to determine the time window for analysis by selecting time samples with a factor loading over 0.6 in TF4. This step yielded a time window of 208-268ms, which was taken as the time window for the MMN.

For the spatial region, I selected the channels that showed a peak negativity at the peak latency of TF4 (236ms) and its surrounding channels, resulting in 9 channels: E3, E4, E6, E7, E9, E12, E54, E60, E65. Figure 30 shows the position of the selected channels on the layout of a 64-channel HydroCel Geodesic Sensor Net.

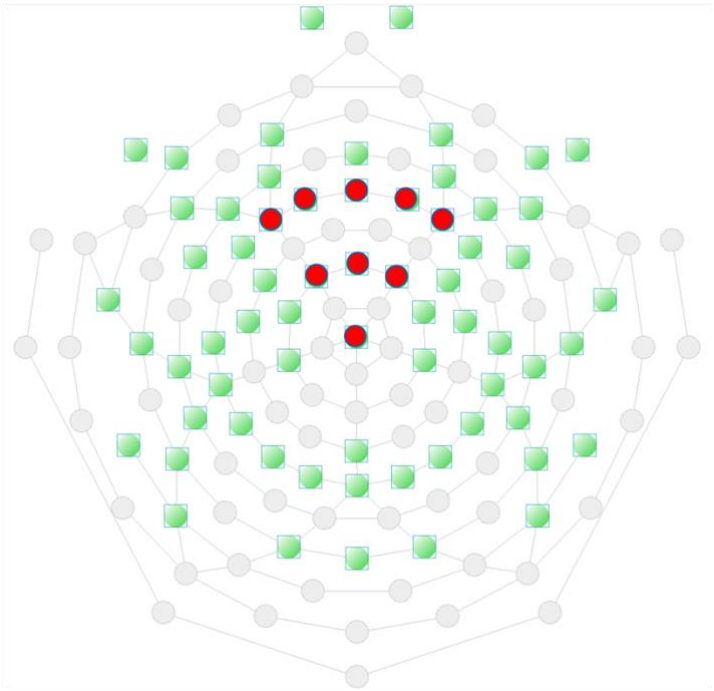


Figure 30: Position of the selected channels in 64-channel HydroCel Geodesic Sensor Net. The night selected channels are: E3, E4, E6, E7, E9, E12, E54, E60, E65.

To summarize, our PCA solution resulted in a time window of 208-268ms and nine frontal-central channels featuring a frontocentral negativity.

4.2.2 Statistics

The current MMN computation employed non-identity MMN by subtracting the ERP of the 64ms-VOT standards from the ERP of the 128ms-VOT deviants. The magnitude of the MMN was measured by the mean amplitude averaged over the amplitudes in the time window of 208-268ms and the selected nine channels. Figure 31 shows the waveforms of the standard ERP and the deviant ERP for both oddball groups, as well as the corresponding topographical map of the difference ERP at the start, peak, and end latency of the selected time window.

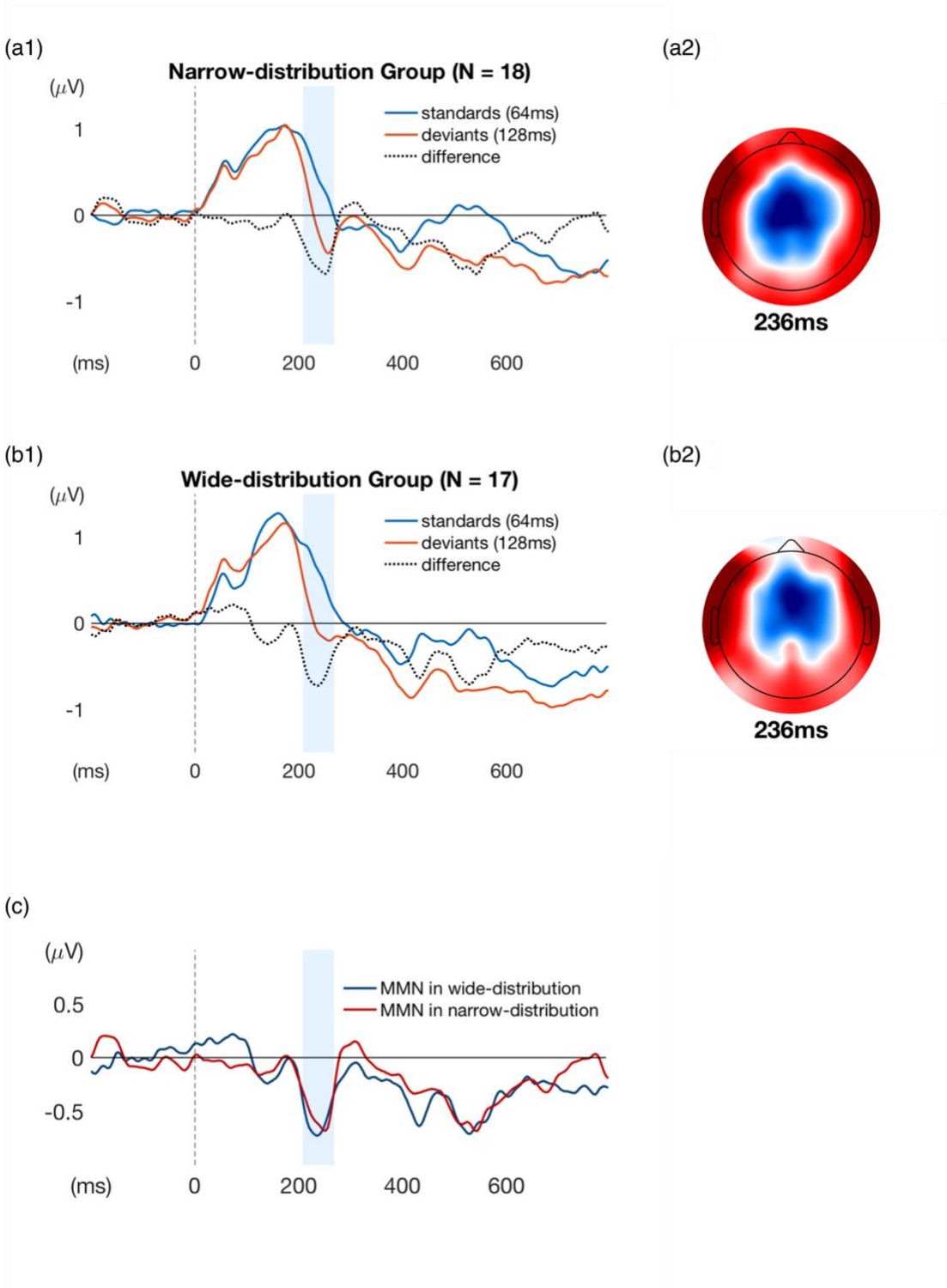

(a1)
Narrow-distribution Group (N = 18)
(μV)
1
0
-1
(ms)
0
200
400
600
standards (64ms)
deviants (128ms)
difference
(a2)
236ms
(b1)
Wide-distribution Group (N = 17)
(μV)
1
0
-1
(ms)
0
200
400
600
standards (64ms)
deviants (128ms)
difference
(b2)
236ms
(c)
(μV)
0.5
0
-0.5
(ms)
0
200
400
600
MMN in wide-distribution
MMN in narrow-distribution

Figure 31: ERP waveform and topography averaged over subjects. (a1, b1) ERP waveforms of standards and deviants in the narrow-distribution group (a1) and wide-distribution group (b1). (a2, b2) Topographical maps of difference ERP (deviants minus standards) at the peak latency (236ms) of TF2 in the narrow-distribution (a2) and wide-distribution group (b2). (c) MMN waveforms (deviants minus standards) in both groups. Blue shaded area indicates the time window for analysis (208-268ms).

The violin plot in Figure 32 shows the distribution of the MMN response in each subject as a function of Block and Group.

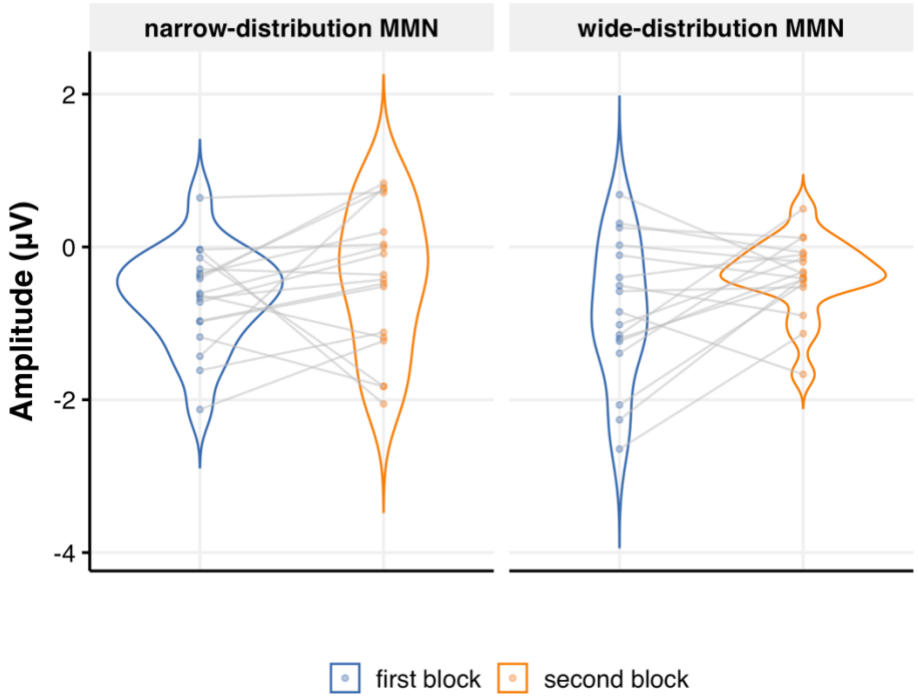


Figure 32: Individual MMN amplitude averaged over the selected time window and the selected channels as a function of condition. Each dot represents one subject's MMN amplitude for a given condition. The grey line connects two data points from the same subject, indicating the MMN magnitude change between the two blocks. The shape of the violin plots indicates the data distribution

For the statistical analysis, the dependent variable was the MMN amplitude. The independent variables were Group (narrow-distribution vs. wide-distribution), Block (block 1 vs. block 2), and their interaction. The model included Subject as a random intercept, as shown in the following R pseudocode.

$$\text{MMN} = \text{Group} + \text{Block} + \text{Group} \times \text{Block} + (1|\text{Subject})$$

The model's total explanatory power is substantial (conditional R^2 = 0.26), and the part related to the fixed factors alone (marginal R^2) is 0.05. I report below the model's coefficients, the corresponding standard error (SE), 95% confidence interval (95% CI), t values, and p values.

Table 8: Model summary

Fixed factors	**Coefficient**	**SE**	**95% CI**	**t(64)**	**p**
(Intercept)	-0.58	0.10	[-0.78, -0.37]	-5.58	< .001***
Group	-0.06	0.21	[-0.47, 0.36]	-0.27	0.788
Block	0.34	0.16	[0.01, 0.67]	2.08	0.042*
Group × Block	0.23	0.33	[-0.43, 0.88]	0.69	0.493

I found a significant intercept [t(64) = -5.58, p < .001, η_p^2 = 0.33 (a large effect)]. This is expected as significance indicated a coefficient significantly smaller than 0 – an MMN effect. Importantly, the effect of Group was not significant [t(64) = -0.27, p = 0.788]. The lack of a Group effect suggested a lack of effect of the statistical structure on the MMN response (Figure 33). In addition, I also found a significant effect of Block [t(64) = 2.08, p = 0.042, η_p^2 = 0.06 (a medium effect)]. In line with the Block effect, Figure 33 shows a reduced (less negative) MMN magnitude in the second block compared to the first block[7].

[7] In addition, I compared the ERPs of the 105 standards across the two distributions. No difference was found.

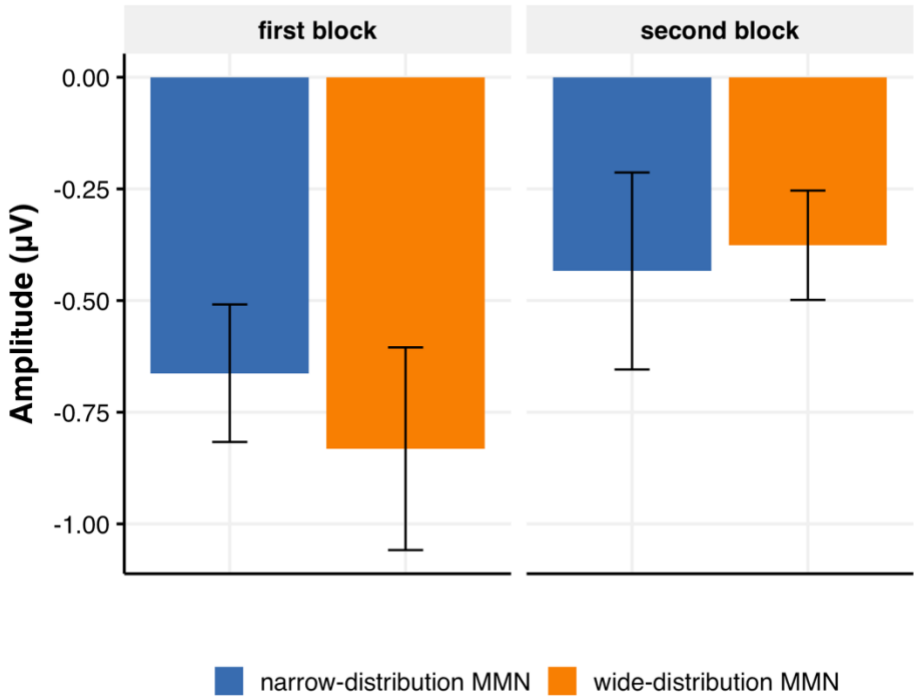


Figure 33: MMN amplitude averaged over subjects in each condition. The error bar indicates standard error.

As post-hoc analyses, I first looked up the simple effect of Group (with one-tailed t-tests) within each block using the *emmeans* package. Again, there was no effect Group for the first block [t(62.9) = 0.64, p = 0.738] or the second block [t(62.9) = -0.22, p = 0.414]. To avoid an interpretation that hinges on a null effect, I measured the likelihood of the Group's null effect by estimating a Bayes factor (Keysers, Gazzola, & Wagenmakers, 2020) using JASP (JASP Team, 2022). I performed a Bayesian mixed ANOVA, which produced inverted Bayes Factor (BF_{01}) ratios. The BF_{01} ratio indexes how much a model improved when the Group effect was excluded

relative to when the Group effect was included. The Bayesian mixed ANOVA produced the following BF_{01} ratios:

Table 9: BF_{01} ratios

Models	**BF_{01}**
Block	1
Block + Group	3.118
Group	5.470
Block + Group + Block × Group	8.315

The results show that a model with a Block effect alone is 3.118 times more likely than a model with a Block and a Group effect, 5.470 times more likely than a model with a Group effect alone, and 8.315 times more likely than a model including both main effects and the interaction.

I also examined the source of the Block effect by looking at how the magnitudes of the standards (64ms) and the deviants (128ms) changed from the first block to the second block. The line plot in Figure 34 showed a change for the deviants but not for the standards. To confirm that the source of the Block effect was the change in the deviant response. I built a model including Stimulus (standard vs. deviants), Block (block 1 vs. block 2), and their interaction. The model included Subject as a random intercept, as shown in the following R pseudocode.

$$\text{Amplitude} = \text{Stimulus} + \text{Block} + \text{Stimulus} \times \text{Block} + (1|\text{Subject})$$

The following results were obtained:

Table 10: Model summary

Fixed factors	**Coefficient**	**SE**	**95% CI**	**t(134)**	**p**
(Intercept)	0.25	0.16	[-0.06, 0.57]	1.58	0.116
Stimulus	-0.58	0.09	[-0.76, -0.39]	-6.07	< .001***
Block	0.13	0.09	[-0.06, 0.31]	1.34	0.182
Stimulus × Block	0.34	0.19	[-0.04, 0.71]	1.79	0.076

Although the interaction between Stimulus and Block did not reach significance, given the trend in Figure 34, I still decomposed the interaction by looking at the simple effect of Block within each level of Stimulus. Running one-tailed t-tests, I found a significantly smaller deviant response in the second block than in the first block [$t(102) = -2.21$, $p = 0.015$], but the standard responses were comparable in both blocks [$t(102) = 0.32$, $p = 0.624$]. The results confirmed that the reduction in the deviant response drove the smaller MMN amplitude in the second block.

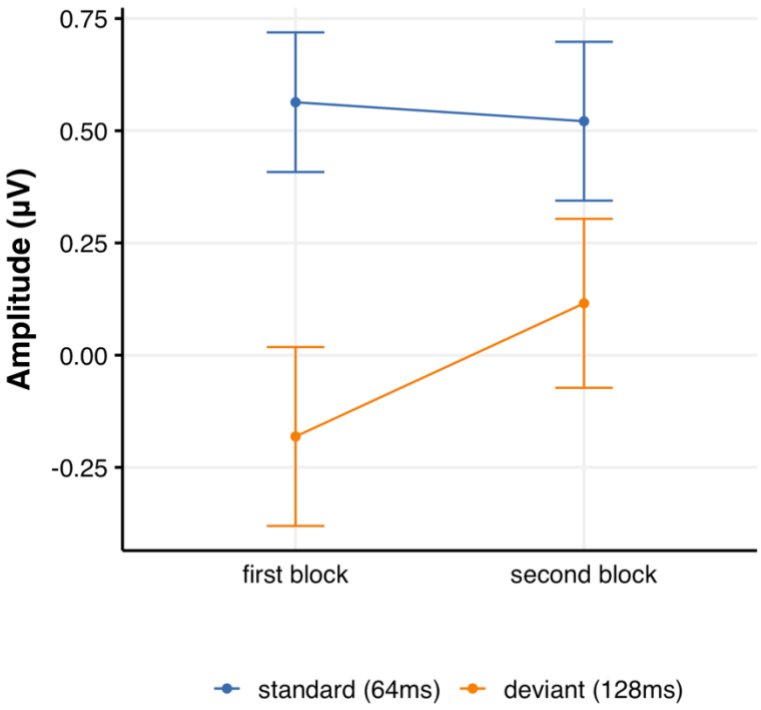


Figure 34: ERP amplitude (pooling across conditions) averaged over subjects, selected time points, and channels. Error bar indicating the standard error.

4.3 Discussion

The null hypothesis test and the Bayesian ANOVA showed that the MMN magnitude did not differ across the two groups exposed to different statistical structures of the proximal stimuli. This result is inconsistent with Garrido et al. (2013)'s finding, where the statistical structure of the proximal stimuli modulated the MMN amplitude. The discrepancy could be explained by the difference between the speech versus the nonspeech perception. During the nonspeech perception, the brain might compute a Bayesian model where learning is essentially about dynamically updating the brain's estimates of conditional probabilities of the upcoming event

(Mathys, 2011). Given that the brain can implicitly trace the statistical structure of the proximal stimuli, a Bayesian-learning brain in Garrido et al. (2013)'s experiment would know that the presence of tone frequencies higher than the deviant (2000Hz) has a lower probability in the narrow distribution ($p \approx 0$) than in the wide distribution ($p = 0.09$). Therefore, the brain was more surprised to encounter a deviant frequency in the narrow distribution than the same deviant frequency in the wide distribution, which was reflected as a larger (more negative) MMN amplitude in the narrow distribution than in the wide distribution. However, in the current experiment, the two distributions elicited comparable MMN magnitudes. It is possible that the brain in the current experiment does more than merely compute a Bayesian model based on the statistical structure of the proximal stimuli due to the nature of the speech stimuli. To test this possibility, we could rerun the experiment with matched nonspeech (spectrally rotated) stimuli. If the discrepancy between the current findings and Garrido et al. (2013)'s results are indeed due to the difference between the speech and nonspeech stimuli, a future experiment with a spectrally-rotated version of the current stimuli experiment should be able to replicate Garrido et al. (2013)'s results.

Alternatively, the Bayesian-learning brain in the current experiment might be insensitive to the difference in the statistical structure between the narrow and wide distributions. Although the deviant VOT in the narrow distribution was more unlikely to occur than the same deviant VOT in the wide distribution, the deviant VOT was clearly outside the range of the standards' variability in both distributions. The brain response to the deviant in both distributions might have already been saturated. In Garrido et al. (2013), the deviant tone frequency (2000Hz) in the narrow distribution was 1.33SD away from the mean (500Hz), and the same deviant frequency in the wide

distribution was about 4SD away from the mean. In contrast, in the current experiment, the deviant was 3SD away from the mean in the wide distribution and 9SD away from the mean in the narrow distribution. The question was whether the MMN response got saturated when a deviant was 3SD away from the standard mean. Daikhin and Ahissar (2012) found that the MMN amplitude in the varying-standard paradigm was modulated by how much the deviant deviated from the standards. Using pure tones as stimuli, they found the smallest MMN amplitude when the deviant tone (1080Hz) was deviant by 8% from the mean of standard tones (1000Hz), a larger MMN amplitude when the deviant tone (1400Hz) was deviant by 40%, and the largest MMN for a 100% deviance (2000Hz). Note that their degree of deviance was measured by a linear-scale frequency difference between a deviant tone and the mean of standard tones, and their various standards formed a uniform distribution. To make their stimulus parameters comparable to the current study, I converted their frequency values to the log2 scale and calculated the SD of their uniformly presented standard tones using the formula for calculating an SD of a uniform distribution:

$$\sigma = \frac{\max - \min}{\sqrt{12}}$$

Their findings could thus be translated as: they found the smallest MMN amplitude when the deviant was 7SD away from the mean of standards, a larger MMN for a 29SD deviance, and the largest MMN for a 60SD deviance[8]. Since an MMN amplitude modulation survived with a difference between a 7SD deviance versus and a 39SD difference, there is no reason that our 3SD versus 9SD would lead to a ceiling

[8] If we have adopted the linear scale, their three levels of deviance would be 7SD, 35SD, and 87SD, which would not change our conclusion.

effect. That being said, it is another question whether their results can provide insights into the current experiment as the nonspeech pure tone and the speech VOT involve different neuronal populations leading to different EEG responses.

4.3.1 Reduced MMN in the second block

In the current experiment, we also observed a significant main effect of Block ($p = 0.042$), suggesting that the MMN magnitude was larger (more negative) in the first block than in the second block. A smaller deviant response in the second block drove the reduction in the MMN magnitude. Previous studies have found that if a deviant stimulus was immediately repeated, the MMN to the second deviant would show attenuation by about 50% in magnitude (Müller, Widmann, & Schröger, 2005; Sams, Alho, & Näätänen, 1984). This effect was called a "deviance-repetition effect" and is believed to be a short-term habituation effect of the MMN generator process. In the current experiment, the second block was a repetition of the first block, and we also found about a 50% reduction in the MMN magnitude. The effect resembles a deviant-repetition effect, but the underlying mechanism cannot be a short-term habituation of the MMN generator as the effect lasted to the second block. Another possibility is that the magnitude reduction is a stimulus-specific long-term neuronal habituation. The long-term habituation effect has been reported in animal studies. For example, Lu et al. (2018) did single-neuron recordings on a ferret. They found that the ferret exhibits long-term habituation in the secondary auditory cortex for at least 20 minutes. However, a stimulus-specific effect would predict a magnitude reduction in both the standards and the deviants. However, our results showed a significant magnitude reduction for the deviants only, raising the possibility of a more top-down adaptation instead of a passive bottom-up habituation process. Repeated exposure to

the same stimulus as a deviant constantly might update the brain's expectation about when a deviant will appear. Under the predictive coding framework, the MMN reflects a process of updating a model about the environment when the reality differs from the model based on past experience (Friston, 2005). When deviants are first encountered, the mental model is primarily based on the information of standards and has little expectation about when a deviant would appear. Thus, the first encountered deviants would form a sharp contrast to the model's prediction. However, the brain keeps monitoring its occurrence over a long time. In our experiment, after the repeated presentation of the deviants, the model started to include the information that a deviant would occur after 3 – 15 standards, thus adjusting the expectation about the occurrence of a deviant. Therefore, the deviants encountered in the second block would form a less sharp contrast to the model's prediction, reducing the MMN magnitude.

In sum, Experiment 2 suggests that the within-category MMN observed in Experiment 1 was not simply driven by a statistical summary of the acoustic properties of the proximal stimuli. However, the conclusion was based on a null result which might suffer from the lack of power. Therefore, I conducted a third experiment for decisive evidence.

Chapter 5

EXPERIMENT 3: EVIDENCE FOR STATISTICAL SUMMARY

Experiment 2 found a mismatch negativity (MMN) response in both the narrow and wide distribution conditions. However, the MMN magnitudes of both conditions did not differ. Following the logic of the experimental design, the lack of difference in the MMN magnitude can be explained by the retrieval of identical phonetic knowledge from long-term memory, which could be a prototype or a probability distribution of the VOTs concerning the phonetic realizations of /t/ at the onset position of a stressed syllable. However, the lack of MMN magnitude difference could also be due to the lack of power, i.e., a Type II error. To avoid an unrealistic power requirement, Experiment 3 examines the MMN magnitude itself as evidence for or against phonetic knowledge as the gradient information retained in the memory trace.

The design of Experiment 3 is the same as the narrow-distribution condition in Experiment 2, except that Experiment 3 exchanged the standard and deviant stimuli used in Experiment 2. As a recap, in the narrow-distribution condition in Experiment 2, standard tokens formed a normal distribution with a mean of 6 and an SD of 0.11, and the deviant stimulus has a VOT of 7 (in log2 scale). In Experiment 3, the standard tokens will form a normal distribution with a mean of 7 and an SD of 0.11; the deviant will have a VOT of 6 (in log2 scale). That way, the standard VOTs are now atypical realizations of /t/. At the same time, the deviant VOT is a typical realization, approximately equivalent to the mean (60ms) of the empirical VOTs (Chodroff &

Wilson, 2018). The deviant VOT is also an outlier to the probability distribution of VOTs presented in the experiment. If the deviant VOT (64ms) results in an MMN effect, the MMN can only result from the memory trace being a statistical summary of the acoustic properties of the presented stimuli. On the other hand, if the varying standards invoke a representation about a prototype or a probability distribution with a mean VOT of about 60ms, then the deviant would be identical to the memory trace, and no MMN is expected.

5.1 Methods

5.1.1 Participants

Twenty-five subjects from the University of Delaware participated in the experiment. All subjects were monolingual English speakers aged 18 to 35 (23 females, mean age = 20, SD = 3) and reported no history of language impairment. Subjects received either $20 or extra credit for completing the experiment. The experiment procedure was approved by the University of Delaware Internal Review Board and was compliant with the principles for ethical research established by the Declaration of Helsinki.

5.1.2 Stimuli

The stimuli used in the current experiment were extracted from the same stimulus set used in Experiment 1 and Experiment 2. Readers are referred to 3.2.2.1 for detailed procedures and stimulus creation parameters.

For the distribution of standard VOTs, I used a MATLAB script to sample 840 VOT values from a normal distribution with a mean of 7 (on a log2 scale) and an SD of 0.11. Each sampled value was then converted to the linear scale and rounded to the

nearest integer. VOT values above 145 were replaced with 145. Figure 35 shows the distribution in both the log2 scale and the linear scale.

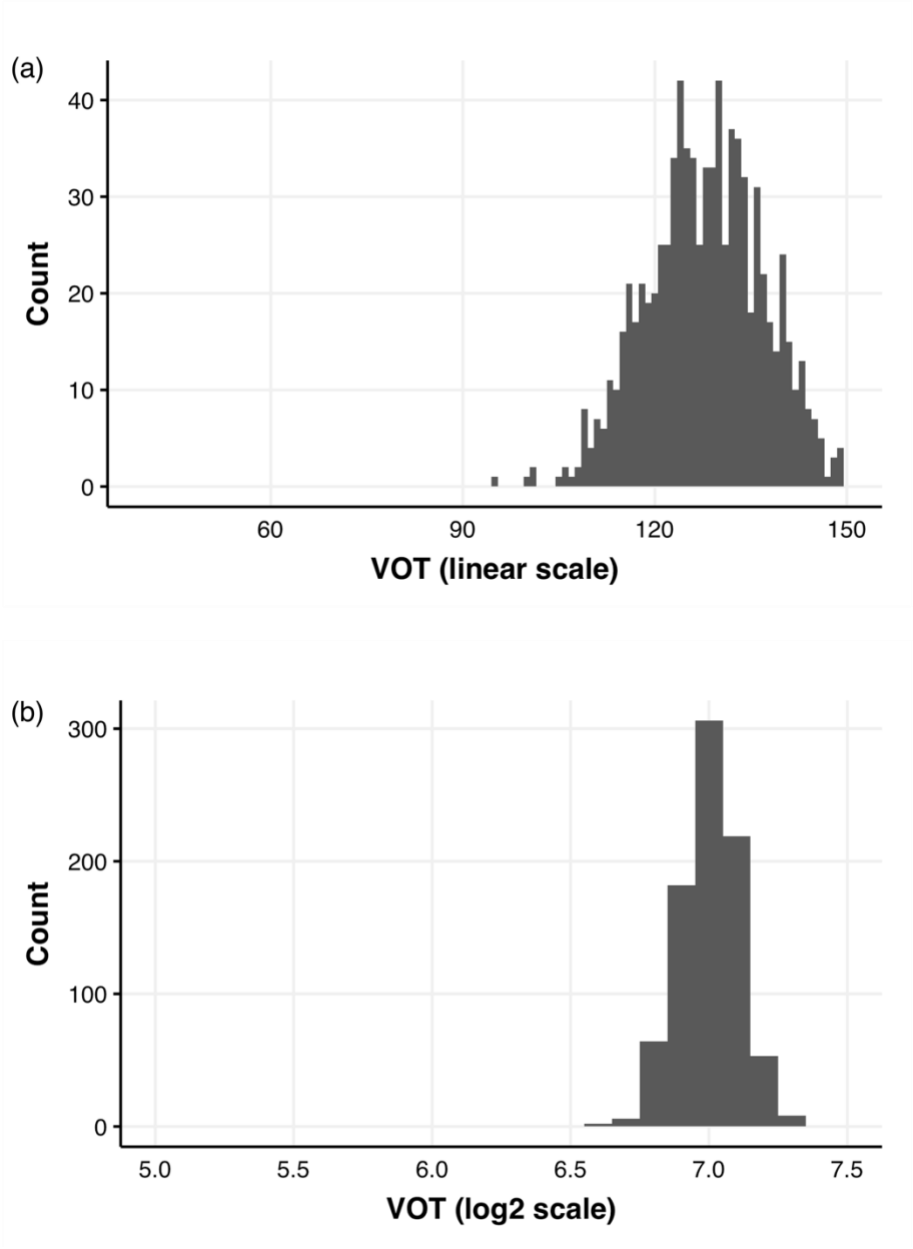


Figure 35: Frequency distribution of the 840 VOT values. (a) VOT values on the linear scale. (b) VOT values on the log2 scale.

In addition to the 840 VOT values, there were another 105 standards with a 128ms VOT (equal to the mean of the distribution) and 105 deviants with a 64ms VOT. In the log2 scale, the deviant VOT is 9SD away from the mean in the narrow distribution. The 100 deviants of a 128ms VOT were compared to the 100 standards of a 64ms VOT to compute the MMN.

5.1.3 Design

As in Experiment 2, each deviant was preceded by a set of standards with a count of the following seven values: 3, 5, 7, 9, 11, 13, 15. Each count was associated with 15 trains of standards, resulting in 945 standards. The 945 standards included the 840 standards drawn from a normal distribution and the 105 standards used for computing an MMN response. The stimulus list also included 32 randomly interspersed target stimuli. Each subject was exposed to the same stimulus list twice as two separate blocks. Nonetheless, the analysis was done by collapsing both blocks. The design has only one independent variable: a within-subject variable Stimulus (standard vs. deviant). Figure 36 illustrates the experimental design and the MMN computing strategy.

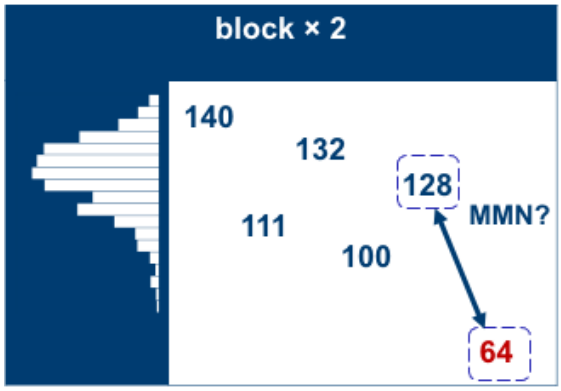


Figure 36: Illustration of the experimental design and computation of a non-identity MMN. Standard VOT values are in blue, and deviant VOT values are in red. To compute a non-identity MMN, I subtract the standards' ERP from the deviants' ERP in the same block (double-sided arrows).

If the comparable MMN magnitude in Experiment 2 was driven by phonetic knowledge, then the atypical standards would also retrieve the same prototype or the same probability distribution as typical standards did. When the deviant carries the most typical VOT value, there is no clash between the deviant VOT and the phonetic knowledge. Thus, we should not observe an MMN response. In contrast, if an MMN is found in the current experiment, it would be the evidence for a statistical summary based on the acoustic properties of the presented stimuli.

5.1.4 Procedure, apparatus, data acquisition and preprocessing

The procedure of Experiment 3 was the same as that of Experiment 2. Readers are referred to 4.1.4 for details. Each subject went through the instruction, practice, EEG net placement, two blocks of EEG acquisition, and EEG net removal. The whole session took about 1.5 hours.

The equipment and the data acquisition and preprocessing were the same as in Experiment 2. Readers are referred to 4.1.5 for details. The extracted trials for analysis

include the 105 standards of a 128ms VOT and the 105 deviants with a 64ms VOT. After the preprocessing, Table 11 shows the mean percentage of bad trials in each condition

Table 11: Percentage of bad trials

	Percentage of bad trials (%)	
Block	**Standards**	**Deviants**
first	5.1%	5.6%
second	3.7%	3.5%

5.1.5 Planned signal processing

5.1.5.1 Deciding time window and channels for MMN

To identify the time window and the channels for the MMN analysis, a temporal principal component analysis (PCA) was used to decompose the data. To construct the input to the PCA, I first computed one deviant ERP and one standard ERP for each subject by respectively averaging the 210 (i.e., 105 × two blocks) deviant responses and the 210 (i.e., 105 × two blocks) standard responses, collapsing the two blocks. Then one difference ERP was derived by subtracting the standard ERP from the deviant ERP. This difference ERP was used as the input to the temporal PCA.

The PCA procedure extracted latent temporal factors from the input data. I then looked through each temporal factor to examine whether the peak latency of the factor fell 100-300ms after the stimulus onset. The time window for analysis

comprised the time points centering around the peak latency and carrying a factoring loading of 0.6 or higher. The channels for analysis were chosen based on the peak channel, which should have the greatest negativity at the peak latency. For data analysis, I averaged the amplitudes over the selected time points and channels for each participant.

5.1.5.2 Statistical analysis

All statistical analyses were conducted using the R software and a linear mixed-effects model using the *lmer* function from the *lme4* package. The dependent measure was the mean ERP amplitude averaged over the PCA-delimited time window and the channels for each subject. The model included only one fixed factor: Stimulus (standard vs. deviant). The model also included Subject as a random intercept. For the fixed factor of Stimulus, I constructed an orthogonal contrast for the two levels and used the coefficients computed with the orthogonal contrasts to indicate the overall effect. The model's explanatory power and parameter coefficients were obtained using the *report* function from the *report* package and the *model_parameters* function from the *parameters* package. For the effect size, I calculated the partial eta-squared (η_p^2).

5.2 Results

5.2.1 PCA solution

Following the PCA procedure specified above, I ran a temporal PCA with a Kaiser weighting (kappa = 3). To determine which factors to retain, a scree plot in combination with a Parallel Test was used to compare the factors extracted from the original data to those from a randomized dataset. Eight temporal factors were retained

as they accounted for more variance than those extracted from the randomized dataset (Figure 37).

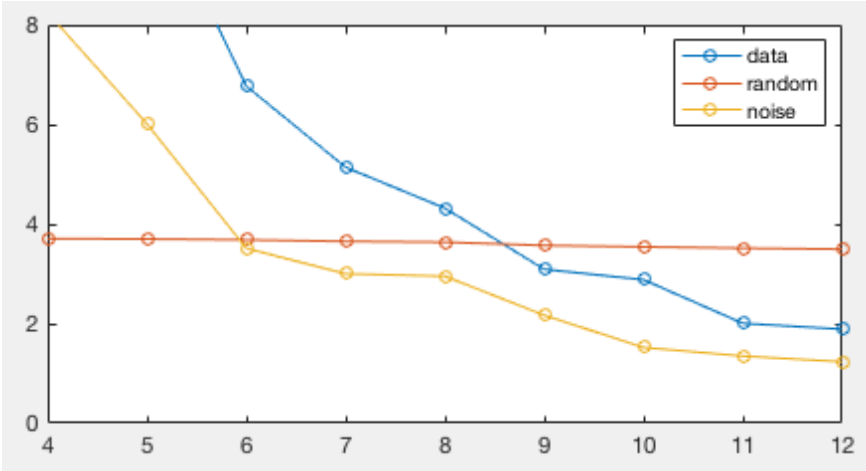


Figure 37: Scree plot with Parallel Test. The parallel test compared the factors extracted from the original data to those from a randomized dataset with the same dimensions. The plot suggests retaining eight temporal factors (up to which the blue curve is above the red curve)

The eight factors accounted for 95% of the total variance of the dataset. To determine the temporal factors that reflected an MMN, I first selected the ones individually accounting for more than 6% of the total variance among the eight temporal factors. The first four temporal factors were thus selected: The first temporal factor (TF1) had an energy distribution peaking at 476ms and accounted for 52% of the total variance; the second temporal factor (TF2) peaked at 772ms and accounted for 15% of the total variance; the third temporal factor (TF3) peaked at 272ms and accounted for 9% of the total variance; the fourth temporal factor (TF4) peaked at 364ms and accounted for 6% of the total variance.

The temporal factor retained was determined by considering the peak time's latency and the topographic map at the peak time (Figure 38). The peak latency of TF1, TF2, and TF4 all fell outside the pre-defined peak time window and were thus discarded. I thus retained TF3, which had a peak latency of 272ms and featured a frontocentral negativity.

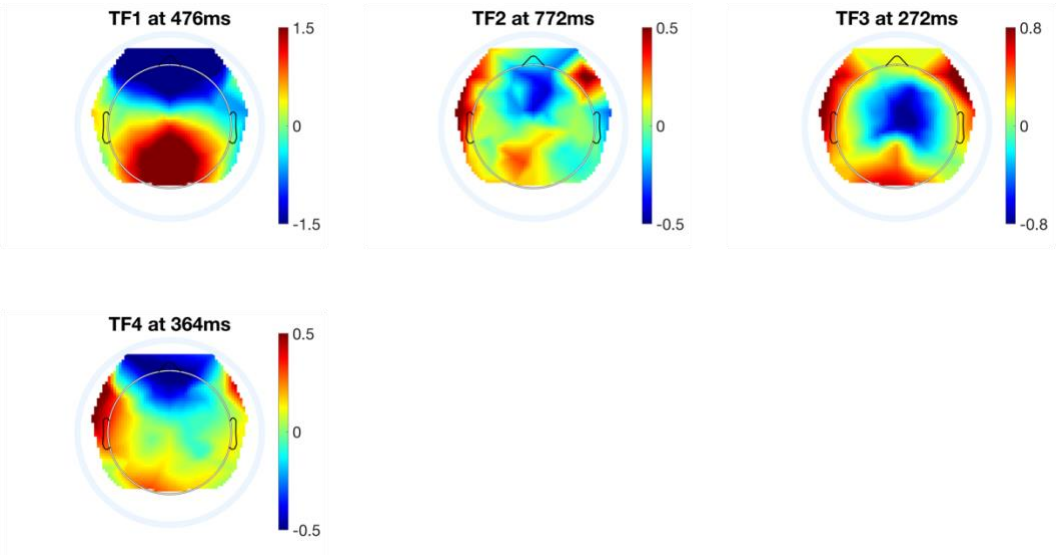


Figure 38: Topography at temporal factor's peak latency.

Having determined TF3 to be retained, the time window for analysis was then determined by selecting time samples with a factor loading over 0.6 in TF3. This step yielded a time window of 200-308ms, which was taken as the time window for the MMN.

For the spatial region, I selected the channels that showed a peak negativity at the peak latency of TF3 (272ms) and its surrounding channels, resulting in six

channels: E41, E50, E51, E53, E54, and E65. Figure 39 shows the position of the selected channels on the layout of a 64-channel HydroCel Geodesic Sensor Net.

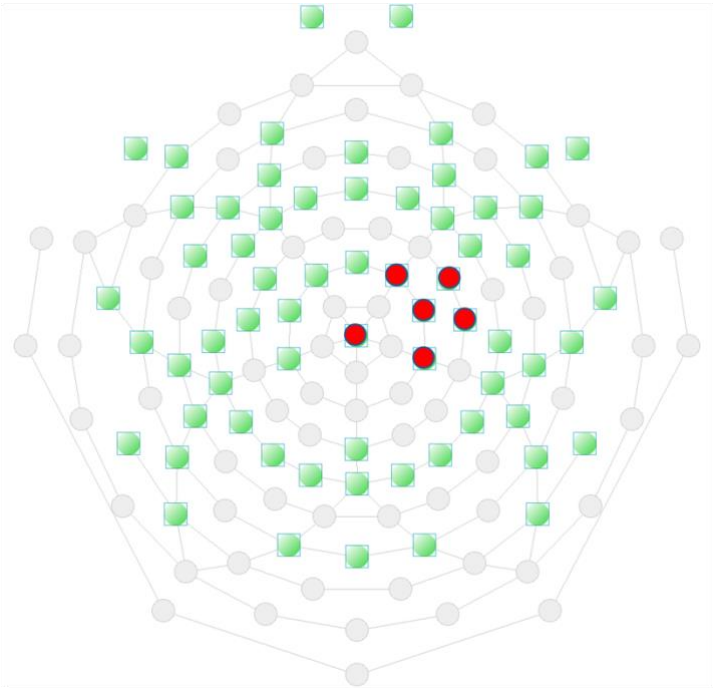


Figure 39: Position of the selected channels in 64-channel HydroCel Geodesic Sensor Net. The six selected channels are E41, E50, E51, E53, E54, and E65.

To summarize, our PCA solution resulted in a time window of 200-308ms and nine frontal-central channels featuring a frontocentral negativity.

5.2.2 Statistics

The current MMN computation employed non-identity MMN by comparing the ERP of the 128ms-VOT standards to the ERP of the 64ms-VOT deviants. The ERP magnitude was measured by the mean amplitude averaged over the amplitudes in

the time window of 200-308ms and the selected six channels. Figure 40 shows the waveforms and the topographical map of the difference ERP at the peak latency of the selected time window.

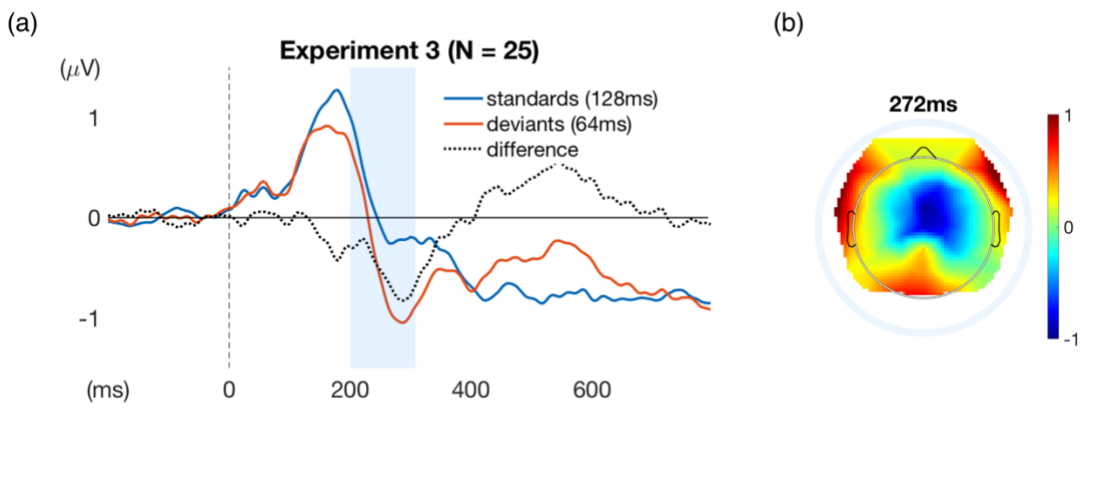


Figure 40: ERP waveform and topography averaged over subjects. (a) ERP waveforms of standards, deviants, and the difference. (b) Topographical maps of the difference ERP (deviants minus standards) at the peak latency (272ms) of TF2. The Blue shaded area indicates the time window for analysis (200-308ms).

The violin plot in Figure 41 shows the distribution of the ERP response in each subject as a function of Stimulus, collapsing the two blocks.

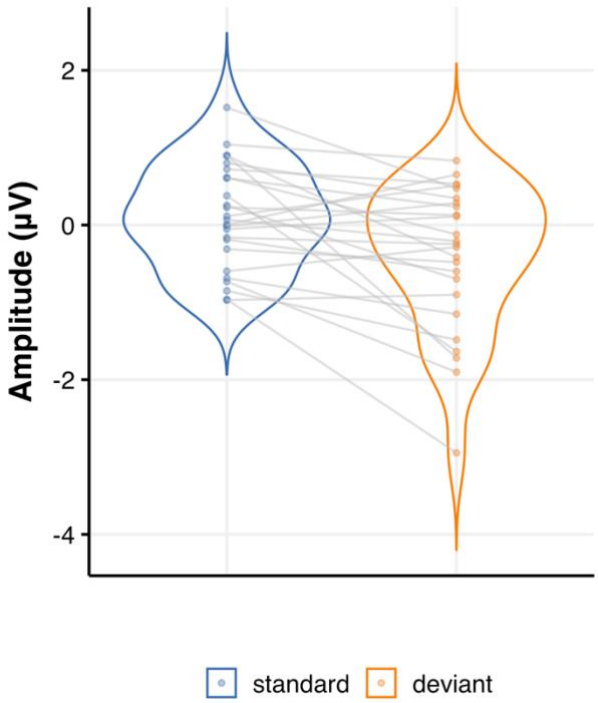


Figure 41: Individual MMN amplitude averaged over the selected time window and the selected channels as a function of Stimulus. Each dot represents one subject's ERP amplitude. The shape of the violin plots indicates the data distribution.

For the statistical analysis, the dependent variable was the MMN amplitude. The independent variables were Stimulus. The model included Subject as a random intercept, as shown in the following R pseudocode.

$$\text{MMN} = \text{Stimulus} + (1|\text{Subject})$$

The model's total explanatory power is substantial (conditional $R^2 = 0.57$), and the part related to the fixed factors alone (marginal R^2) is 0.1. Table 12 presents the model's coefficients, the corresponding standard error (SE), 95% confidence interval (95% CI), t values, and p values.

Table 12: Model summary

Fixed factors	Coefficient	SE	95% CI	t(46)	p
(Intercept)	-0.16	0.14	[-0.45, 0.13]	-1.13	0.266
Stimulus	-0.53	0.16	[-0.86, -0.21]	-3.31	0.002**

The significant effect of Stimulus [t(46) = -3.31, p = .002, η_p^2 = 0.19 (a large effect)] is a clear indication of a mismatch negativity effect, as shown in Figure 42. The robust MMN observed in Experiment 3 provides evidence for a statistical summary of the proximal standards.

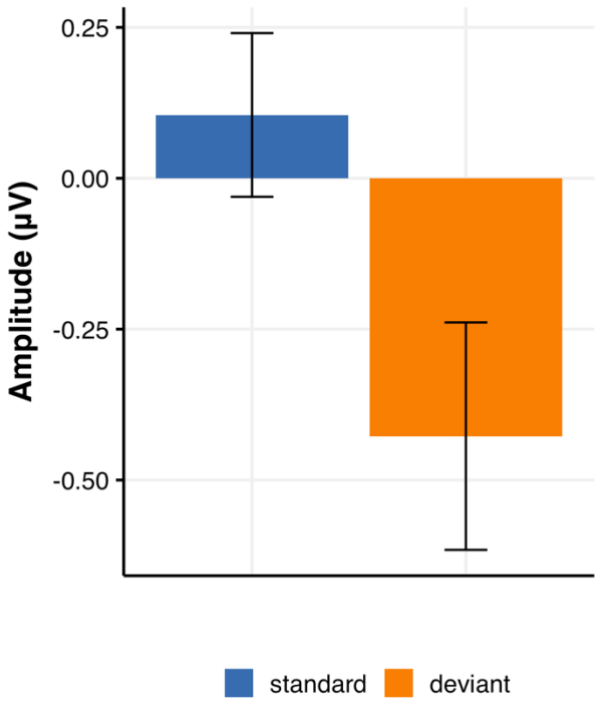


Figure 42: MMN amplitude averaged over subjects. The error bar indicates standard error.

Chapter 6

GENERAL DISCUSSION

Natural speech is gradient and full of inter- and intra-speaker variability, but listeners can extract invariant categorical information out of the variability. Listeners also know how a phoneme category could be acoustically and articulatorily realized in a specific environment – a knowledge the current dissertation refers to as phonetic knowledge. With multiple MMN studies showing the effect of a categorical contrast, it is an open question whether a difference in the gradient information could otherwise drive an MMN. By examining this issue, we are also focusing attention on whether the memory trace in the varying-standard paradigm retains gradient information alone with a category representation. The current dissertation conducted three experiments to address the question. Below I summarize the main findings and discuss their implications for speech perception.

6.1 What have we found?

The three experiments used the varying-standard oddball paradigm, which is assumed to elicit a discrete category representation. I focused on the mismatch negativity (MMN) component, which reflects the brain's pre-attentive novelty detection process. In the first experiment, the standards and the deviants were realized by syllable /tæ/ with various VOTs for the onset /t/. The VOTs of the standard /t/ were in the typical range (i.e., 42, 48, and 55ms), while the deviant /ta/ was an atypical realization (i.e., 119ms). I specifically asked whether a within-category MMN could

emerge with the various standards compared to the non-varying standards. A within-category MMN would indicate that the memory trace must retain gradient VOT information such that the brain treats the deviant VOT as violating the prediction based on the standard VOTs. The first experiment did show a within-category MMN, entailing pre-attentive access to gradient information. However, the observed within-category MMN could come from two sources: The MMN could be driven by phonetic knowledge (i.e., a prototype or a distribution with a mean of 60ms) or by a statistical summary of the VOT values of the proximal stimuli (i.e., a uniform distribution with a mean of 48ms and a range between 42 and 55ms), which would also lead to contrast to the deviant VOT. To distinguish between these two possibilities, Experiment 2 manipulated the statistical structure of the proximal VOTs such that the same standard-deviant difference was embedded either in a wide distribution or in a narrow distribution of the proximal VOTs. Since the MMN magnitude can be modulated by the shape of the distribution of proximal acoustic characteristics, if the observed within-category MMN was driven by a statistical summary of the proximal VOTs, we would expect a larger MMN for the narrow distribution than for the wide distribution. The results showed that the two distributions elicited comparable MMNs, suggesting that a statistical summary alone could not account for the within-category MMN observed in Experiment 1. However, the conclusion was based on the null results. Experiment 3 exchanged the standards and deviants used in Experiment 2, such that the standards were in the atypical range while the deviant carried a typical VOT. If the gradient information retained in the memory trace was about phonetic knowledge retrieved from long-term memory, the deviant should not elicit an MMN response as it carried the most typical VOT. However, we did observe an MMN response, indicating

the evidence for a statistical summary based on the proximal stimuli. Taking those findings together, I concluded that the memory trace contains gradient information in the varying-standard MMN paradigm and that gradient information comes from the statistical summary of the presented stimuli.

6.2 Relating to previous findings

Previous studies have failed to find a within-category MMN in the varying-standard paradigm. This is true even when standards and deviants are different allophones of a phoneme. For example, Kazanina, Phillips, and Idsardi (2006) tested Russian and Korean speakers with a VOT contrast for alveolar stops. The standards and deviants formed a phonemic contrast (/d/ vs. /t/) for Russian speakers, but they are allophones of Korean /t/. The contrast elicited an MMN for Russian speakers but not for Korean speakers, indexing, unsurprisingly, an effect of phoneme category on speech perception. What is surprising is the lack of an MMN for Korean speakers. Given that the current studies find a within-category MMN even when the standards and deviants belong to the same allophone, one would expect an MMN for Korean speakers when the standards and deviants belong to different allophones. Yet, a similar pattern was found in Phillips et al. (1995) which tested English and Japanese speakers with the [r – l] contrast. The two sounds are separate phonemes in English but are allophones of Japanese /r/. The contrast elicited an MMN in English speakers but not in Japanese speakers.

The question is why those studies did not observe an MMN when the standards and deviants belonged to different allophonic categories. One possibility is that the auditory difference between the standards and deviants is simply too small to support the sorting of stimuli into standards and deviants, regardless of the allophonic

difference. Following this logic, as long as the acoustic properties of the deviants are sufficiently different from those of standards, an MMN should emerge. Thus, an MMN driven by an allophonic contrast should have a bigger chance to be observed in a vowel contrast than a consonant contrast, since vowel perception exhibits a more gradient pattern than consonant perception (Fry, Abramson, Eimas, & Liberman, 1962). Indeed, Miglietta, Grimaldi, and Calabrese (2013) tested an allophonic contrast [ɛ - e] in a southern variety of Italian, and found an MMN response in native speakers. Note that not all vowel studies found an MMN response driven by an allophonic contrast. For example, in Winkler et al. (1999) the allophonic vowel contrast [e – æ] in Hungarian did not elicit an MMN in native speakers. The discrepancy between the above two studies' results can be explained by the acoustic distance between the standards and deviants: In Miglietta, Grimaldi, and Calabrese (2013), the formant differences between allophonic standards and deviants (130Hz for F1, and 59Hz for F2) were more salient than the those in Winkler et al. (1999) (33Hz for F1, and 81Hz for F2). Consequently, the former study found an MMN, while the latter did not.

6.3 Implications for speech perception

The present findings suggested an important role of gradient information in speech perception. Previous research has shown that the acoustic and idiosyncratic information of speech was remembered and stored for a considerable amount of time (Palmeri, Goldinger, & Pisoni, 1993; Toscano et al., 2010). The present study is the first work showing an MMN driven by a simple within-category acoustic contrast in the varying-standard paradigm, suggesting that gradient information is pre-attentively encoded and retained along with the categorical information. The finding thus bears on three questions in speech perception, as discussed below.

6.3.1 Where is category representation from?

The current results showed that the acoustic properties of the proximal stimuli are pre-attentively encoded and retained with the categorical information. As both the gradient and categorical information are retained in the memory trace, a question that naturally arises from this finding is the following: Is the categorical information simply a statistical summary based on the gradient information of the proximal stimuli, or is it retrieved from long-term memory? The most extreme answer is given in the Rich Phonology approach proposed by Robert Port (2007), which contends that abstraction is computed in real-time based on the memory traces of concrete phonetic tokens whenever needed. Port's view suggests that, in speech perception, gradient information is fundamental, while categorical information is secondary and does not need to enter long-term memory. The current study does not provide conclusive evidence either for or against this claim, however, we may obtain some insight from two lines of research.

The first line of research is about the auditory cortex encoding unified abstract features, as a type of perceptual category, across different dimensions of phonetic properties. For example, Monahan et al. (2022) conducted an MMN study where the standard syllables carried voiceless onsets while the deviant syllables carried voiced onsets. Their crucial manipulation was that the voiceless standards included both stops and fricatives in the same block. The voiceless feature was thus realized in the temporal dimension for stops (VOT) but in the spectral dimension for fricative (periodic low-frequency spectral energy). Despite the different dimensions for voicing realization, the standard-deviant contrast nevertheless elicited an MMN. Since the subjects were naive to the task purpose, an MMN response entails that the memory must pre-attentively encodes a category representation associated with the voiceless

feature. That category representation cannot simply be a statistical summary based on the proximal stimuli. This is because the acoustic properties corresponding to the voicing feature of their stimuli came from two independent dimensions: temporal and spectral. A statistical summary cannot be computed based on acoustic properties across two acoustic dimensions. Therefore, category information about voicing has to be retrieved from long-term memory.

Another line of research is the MMN studies on the Phonological Underspecification Theory. In those studies (e.g. Cornell et al., 2011; Hestvik & Durvasula, 2016), a minimal MMN was observed even for an across-category contrast (e.g., standard /d/ vs. deviant /t/), as long as the standards were mapped to a category representation containing unspecified features (e.g., /d/ has an unspecified voicing feature). The knowledge of features being underspecified has to be stored in long-term memory and retrieved during the experiment so that the unspecified features can be compared to the deviants, leading to an absence of an MMN. If the categorial features are simply extracted from the proximal stimuli, there is no reason for an extracted feature to be unspecified. Note that the underspecification studies posed a challenge to the current finding about the memory encoding gradient information. If the memory trace retains the acoustic properties of the proximal stimuli, then those acoustic properties will contrast the deviants regardless of the unspecified features. It is thus the categorical features but not the gradient acoustic properties that are unspecified, so we should observe an MMN. Future studies are needed to confirm that the lack of MMN is not due to the stimulus design. Nevertheless, if we assume that gradient information is not retained in the memory trace along with an underspecified category representation, we need to stipulate that the retrieval of a feature with unspecified

feature values might impede the retrieval of the relevant phonetic knowledge. Again, this would require prior knowledge of the category representation, which suggests that the category representation has to be stored in long-term memory.

6.3.2 Where does gradient information go?

In speech perception, there is a step where listeners match gradient speech signals to category representations through the process of speaker normalization. An extreme version of speech representation claims that after the speaker normalization, the idiosyncratic or acoustic characteristics of speech never enter the long-term memory (Halle, 2003). However, evidence has shown that gradient information is not always "normalized" and discarded. For example, Schmidt and Toscano (2017) presented subjects with pronouns that sounded in between "he" and "she". Those pronouns induced an ambiguous interpretation of the gender identity of the referent. When the subjects later encountered words that disambiguated the referent and contrasted their original interpretation, the amount of time the subjects took to recover from the misinterpretation (as indexed by the time of the eye gaze fixating on the target referent) increased with the distance between the original acoustic token and the endpoint corresponding to the target referent along the continuum. Of course, one could still argue that the results can be explained by the gradient acoustic information retained in the sensory memory or the working memory. However, studies have also shown a gradient sensitivity to the episodic traces stored in long-term memory. For example, Palmeri, Goldinger, and Pisoni (1993) performed a word-recognition task where subjects listened to a list of words and indicated whether each word was a new item or a repetition of a previously presented word. The results showed that participants could recognize a repeated word better if a repeated word were produced

by the same speaker. This speaker-specific voice facilitation persisted even after 64 intervening words when the speaker-specific information of the previously presented word had to be encoded in the long-term memory. Goldinger (1996) also found that the episodic traces of the speaker-specific information was stored with impressive details and lasted at least one week.

A question that now arises is how the gradient information is organized after it enters long-term memory. The current results support a statistical summary based on the proximal stimuli, but no evidence was found either for a statistical summary (a probabilistic distribution) of empirical phonetic realizations or for a prototype-like representation. It is possible that the gradient information was faithfully encoded in long-term memory without abstraction or summary statistics. This view aligns with Exemplar Theory (Nosofsky, 1988). Exemplar Theory suggests that human beings categorize their experiences in terms of exemplar clouds, which are clusters of the remembered episodes of each individual experience. The speech perception research falling in the framework of the Exemplar Theory believes that the idiosyncratic information of speech is encoded and retained in long-term memory (e.g., Miller, 1994; Pierrehumbert, 2001). When listeners perceive speech sounds, they remember each individual occurrence of phonetic tokens just as they are perceived. Of course, more research is needed to investigate the organization of gradient information in long-term memory. However, if long-term memory encodes gradient information as the Exemplar Theory predicts, a question naturally arises from this regarding the nature of the mental lexicon, as discussed below.

6.3.3 Does the mental lexicon contain exemplars?

One central issue for the mental lexicon is in which form of a word is represented. If a word is represented with abstract codes, a set of computations is needed to map the acoustic signal to the abstract representations. Going to the other extreme, if a word is represented as exemplar clouds consisting of all empirical realizations, then no computation is needed. It is also possible that a word representation contains some exemplars or some prototypes, in which case some computation is still needed. The question is whether the mental lexicon is purely abstract or in fact contains some exemplars. Given that long-term memory encodes phonetically detailed information, we could further suspect that the mental lexicon also, by nature, consists of exemplars of meaningful units. Previous studies on word recognition have shown that the brain can retain and recognize highly detailed information, including the idiosyncratic characteristics of individual speakers, consistent with an exemplary view of lexical access (Goldinger, 1996, 1998; Jacoby, 1983; Jacoby & Brooks, 1984; Jacoby & Hayman, 1987; Palmeri et al., 1993). Admittedly, those studies are about behavioral measures and thus cannot serve as decisive evidence for the mental lexicon encoding episodic experiences. However, I would like to emphasize the work by Goldinger (1996, 1998), which assessed the validity of an exemplar model, i.e., MINERVA 2, initially proposed by Hintzman (1986, 1988). The model takes the effect of episodic memory on speech recognition to a logical extreme, assuming that speech recognition is achieved using only exemplars. The model builds separate memory traces for each experience. A retrieval cue contacts all memory traces simultaneously, and each memory trace is activated to a degree depending on its similarity to the retrieval cue. Those activations are summed together to determine the behavior. The model successfully predicted human subjects' response

patterns in a word recall test after short (five minutes) and long (one day, and one week) delays (Goldinger, 1996). The model also predicted subjects' reaction times and the acoustic manifestations in a nonword shadowing task, where the idiosyncratic details of a subject's production attenuated as the interval between the shadowing material and the production increased (Goldinger, 1998). Those results suggest that speech perception *can* be achieved using only exemplars without an abstract category, suggesting that the role of episodic memory traces in speech perception might be more important than we previously thought.

6.4 Conclusion and future direction

With three experiments using the varying-standard oddball paradigm where the standards vary alone VOT, the current dissertation provides evidence for an MMN response solely driven by an acoustic contrast when the standards and deviants belong to the same category. The previous work has been focusing on the MMN driven by a categorical contrast, but the current work suggests that the memory trace in the varying-standard paradigm also retains phonetically gradient information of the presented stimuli which could also drive an MMN response. The finding aligns with a speech perception process that maps from gradient acoustic signals to binary category representations. During this process, both gradient and categorical information were utilized and retained as part of the mental representation. The conclusion points to the important role of episodic memory traces in speech perception.

Note that the gradient information found in the current study is about a statistical summary based on the acoustic properties of the presented stimuli. The current study did not find evidence for an effect of phonetic knowledge stored in long-term memory, whose content is also gradient. It is possible that the experimental

design is not sensitive enough to capture the subtle impact of phonetic knowledge. Future experiments using a spectrally rotated version of the stimuli presented in Experiment 2 and Experiment 3 should be able to tease apart the effect of phonetic knowledge and the effect of the proximal stimuli. This is because spectral-rotated stimuli preserve the acoustic difference between the standards and deviants in the speech condition but remove the linguistic information (Marklund, Lacerda, & Schwarz, 2018). Thus, the experiment with the spectrally rotated stimuli could better inform us of the roles the proximal stimuli and long-term memory play in speech perception.

REFERENCES

Alain, C., Achim, A., & Woods, D. L. (1999). Separate memory-related processing for auditory frequency and patterns. *Psychophysiology*, *36*(6), 737–744. https://doi.org/10.1111/1469-8986.3660737

Alain, C., Woods, D. L., & Knight, R. T. (1998). A distributed cortical network for auditory sensory memory in humans. *Brain Research*, *812*(1–2), 23–37. https://doi.org/10.1016/S0006-8993(98)00851-8

Allen, J., Kraus, N., & Bradlow, A. (2000). Neural representation of consciously imperceptible speech sound differences. *Perception & Psychophysics*, *62*(7), 1383–1393. https://doi.org/10.3758/BF03212140

Andruski, J. E., Blumstein, S. E., & Burton, M. (1994). The effect of subphonetic differences on lexical access. *Cognition*, *52*(3), 163–187. https://doi.org/10.1016/0010-0277(94)90042-6

Aslin, R. N., Pisoni, D. B., Hennessy, B. L., & Perey, A. J. (1981). Discrimination of Voice Onset Time by Human Infants: New Findings and Implications for the Effects of Early Experience. *Child Development*, *52*(4), 1135–1145. https://doi.org/10.1111/j.1467-8624.1981.tb03159.x

Auksztulewicz, R., & Friston, K. (2016). Repetition suppression and its contextual determinants in predictive coding. *Cortex*, *80*, 125–140. https://doi.org/10.1016/j.cortex.2015.11.024

Bader, M., Schröger, E., & Grimm, S. (2017). How regularity representations of short sound patterns that are based on relative or absolute pitch information establish over time: An EEG study. *PLOS ONE*, *12*(5), e0176981. https://doi.org/10.1371/journal.pone.0176981

Barrios, S. L., Namyst, A. M., Lau, E. F., Feldman, N. H., & Idsardi, W. J. (2016). Establishing New Mappings between Familiar Phones: Neural and Behavioral Evidence for Early Automatic Processing of Nonnative Contrasts. *Frontiers in Psychology*, *7*(JUN), 1–16. https://doi.org/10.3389/fpsyg.2016.00995

Barry, R. J., Cocker, K. I., Anderson, J. W., Gordon, E., & Rennie, C. (1992). Does the N100 evoked potential really habituate? Evidence from a paradigm

appropriate to a clinical setting. *International Journal of Psychophysiology*, *13*(1), 9–16. https://doi.org/10.1016/0167-8760(92)90014-3

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Ben-Shachar, M. S., Lüdecke, D., & Makowski, D. (2020). {e}ffectsize: Estimation of Effect Size Indices and Standardized Parameters. *Journal of Open Source Software*, *5*(56), 2815. https://doi.org/10.21105/joss.02815

Blumstein, S. E., Myers, E. B., & Rissman, J. (2005). The Perception of Voice Onset Time: An fMRI Investigation of Phonetic Category Structure. *Journal of Cognitive Neuroscience*, *17*(9), 1353–1366. https://doi.org/10.1162/0898929054985473

Brown-Schmidt, S., & Toscano, J. C. (2017). Gradient acoustic information induces long-lasting referential uncertainty in short discourses. *Language, Cognition and Neuroscience*, *32*(10), 1211–1228. https://doi.org/10.1080/23273798.2017.1325508

Budd, T. ., Barry, R. J., Gordon, E., Rennie, C., & Michie, P. . (1998). Decrement of the N1 auditory event-related potential with stimulus repetition: habituation vs. refractoriness. *International Journal of Psychophysiology*, *31*(1), 51–68. https://doi.org/10.1016/S0167-8760(98)00040-3

Butler, R. A., Spreng, M., & Keidel, W. D. (1969). Stimulus Repetition Rate Factors Which Influence the Auditory Evoked Potential in Man. *Psychophysiology*, *5*(6), 665–672. https://doi.org/10.1111/j.1469-8986.1969.tb02869.x

Callaway, E. (1973). Habituation of Averaged Evoked Potentials in Man. In *Physiological Substrates* (pp. 153–174). Elsevier. https://doi.org/10.1016/B978-0-12-549802-9.50011-X

Chang, E. F., Rieger, J. W., Johnson, K., Berger, M. S., Barbaro, N. M., & Knight, R. T. (2010). Categorical speech representation in human superior temporal gyrus. *Nature Neuroscience*, *13*(11), 1428–1432. https://doi.org/10.1038/nn.2641

Chodroff, E., & Wilson, C. (2018). Predictability of stop consonant phonetics across talkers: Between-category and within-category dependencies among cues for place and voice. *Linguistics Vanguard*, *4*(s2). https://doi.org/10.1515/lingvan-2017-0047

Conway, C. M. (2020). How does the brain learn environmental structure? Ten core

principles for understanding the neurocognitive mechanisms of statistical learning. *Neuroscience & Biobehavioral Reviews*, *112*(August 2019), 279–299. https://doi.org/10.1016/j.neubiorev.2020.01.032

Cornell, S. A., Lahiri, A., & Eulitz, C. (2011). "What you encode is not necessarily what you store": Evidence for sparse feature representations from mismatch negativity. *Brain Research*, *1394*, 79–89. https://doi.org/10.1016/j.brainres.2011.04.001

Cornell, S. A., Lahiri, A., & Eulitz, C. (2013). Inequality across consonantal contrasts in speech perception: Evidence from mismatch negativity. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(3), 757–772. https://doi.org/10.1037/a0030862

Cowan, N. (1984). On short and long auditory stores. *Psychological Bulletin*, *96*(2), 341–370. https://doi.org/10.1037/0033-2909.96.2.341

Cowan, N., Winkler, I., Teder, W., & Näätänen, R. (1993). Memory prerequisites of mismatch negativity in the auditory event-related potential (ERP). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(4), 909–921. https://doi.org/10.1037/0278-7393.19.4.909

Cummings, A., Madden, J., & Hefta, K. (2017). Converging evidence for [coronal] underspecification in English-speaking adults. *Journal of Neurolinguistics*, *44*, 147–162. https://doi.org/10.1016/j.jneuroling.2017.05.003

Daikhin, L., & Ahissar, M. (2012). Responses to deviants are modulated by subthreshold variability of the standard. *Psychophysiology*, *49*(1), 31–42. https://doi.org/10.1111/j.1469-8986.2011.01274.x

Dehaene-Lambertz, G. (1997). Electrophysiological correlates of categorical phoneme perception in adults. *Neuroreport*, *8*(4), 919–924. https://doi.org/10.1097/00001756-199703030-00021

Dehaene-Lambertz, G., & Pena, M. (2001). Electrophysiological evidence for automatic phonetic processing in neonates. *Neuroreport*, *12*(14), 3155–3158. https://doi.org/10.1097/00001756-200110080-00034

Dehaene, S., Izard, V., Spelke, E., & Pica, P. (2008). Log or linear? Distinct intuitions of the number scale in western and Amazonian indigene cultures. *Science*, *320*(5880), 1217–1220. https://doi.org/10.1126/science.1156540

Dien, J. (2010). The ERP PCA Toolkit: An open source program for advanced statistical analysis of event-related potential data. *Journal of Neuroscience*

Methods, *187*(1), 138–145. https://doi.org/10.1016/j.jneumeth.2009.12.009

Dien, J. (2012). Applying principal components analysis to event-related potentials: A tutorial. *Developmental Neuropsychology*, *37*(6), 497–517. https://doi.org/10.1080/87565641.2012.697503

Eimas, P. D., & Corbit, J. D. (1973). Selective adaptation of linguistic feature detectors. *Cognitive Psychology*, *4*(1), 99–109. https://doi.org/10.1016/0010-0285(73)90006-6

Ekman, Gös. (1959). Weber's Law and Related Functions. *The Journal of Psychology*, *47*(2), 343–352. https://doi.org/10.1080/00223980.1959.9916336

Elangovan, S., Cranford, J. L., Walker, L., & Stuart, A. (2005). A Comparison of the mismatch negativity and a differential waveform response. *International Journal of Audiology*, *44*(11), 637–646. https://doi.org/10.1080/00222930500271564

Elangovan, S., & Stuart, A. (2011). A cross-linguistic examination of cortical auditory evoked potentials for a categorical voicing contrast. *Neuroscience Letters*, *490*(2), 140–144. https://doi.org/10.1016/j.neulet.2010.12.044

Escera, C., & Malmierca, M. S. (2014). The auditory novelty system: An attempt to integrate human and animal research. *Psychophysiology*, *51*(2), 111–123. https://doi.org/10.1111/psyp.12156

Eulitz, C., & Lahiri, A. (2004). Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition. *Journal of Cognitive Neuroscience*, *16*, 577–583. https://doi.org/10.1162/089892904323057308

Fernández-Alfonso, T., & Ryan, T. A. (2004). The Kinetics of Synaptic Vesicle Pool Depletion at CNS Synaptic Terminals. *Neuron*, *41*(6), 943–953. https://doi.org/10.1016/S0896-6273(04)00113-8

Fitzgerald, K., & Todd, J. (2020). Making Sense of Mismatch Negativity. *Frontiers in Psychiatry*, *11*(June), 1–19. https://doi.org/10.3389/fpsyt.2020.00468

Fox, N. P., Leonard, M., Sjerps, M. J., & Chang, E. F. (2020). Transformation of a temporal speech cue to a spatial neural code in human auditory cortex. *ELife*, *9*, 1–43. https://doi.org/10.7554/eLife.53051

Friedrich, C. K., Eulitz, C., & Lahiri, A. (2006). Not every pseudoword disrupts word recognition: An ERP study. *Behavioral and Brain Functions*, *2*, 1–10. https://doi.org/10.1186/1744-9081-2-36

Friedrich, C. K., Lahiri, A., & Eulitz, C. (2008). Neurophysiological Evidence for Underspecified Lexical Representations: Asymmetries With Word Initial Variations. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(6), 1545–1559. https://doi.org/10.1037/a0012481

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1456), 815–836. https://doi.org/10.1098/rstb.2005.1622

Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M. (1962). The Identification and Discrimination of Synthetic Vowels. *Language and Speech*, *5*(4), 171–189. https://doi.org/10.1177/002383096200500401

Garrido, M. I., Kilner, J. M., Kiebel, S. J., Stephan, K. E., & Friston, K. J. (2007). Dynamic causal modelling of evoked potentials: A reproducibility study. *NeuroImage*, *36*(3), 571–580. https://doi.org/10.1016/j.neuroimage.2007.03.014

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

Garrido, M. I., Sahani, M., & Dolan, R. J. (2013). Outlier Responses Reflect Sensitivity to Statistical Structure in the Human Brain. *PLoS Computational Biology*, *9*(3), e1002999. https://doi.org/10.1371/journal.pcbi.1002999

Giard, M.-H., Perrin, F., Pernier, J., & Bouchet, P. (1990). Brain Generators Implicated in the Processing of Auditory Stimulus Deviance: A Topographic Event-Related Potential Study. *Psychophysiology*, *27*(6), 627–640. https://doi.org/10.1111/j.1469-8986.1990.tb03184.x

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(5), 1166–1183. https://doi.org/10.1037/0278-7393.22.5.1166

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

Grimm, S., & Escera, C. (2012). Auditory deviance detection revisited: Evidence for a hierarchical novelty system. *International Journal of Psychophysiology*, *85*(1), 88–92. https://doi.org/10.1016/j.ijpsycho.2011.05.012

Gross, C. G., Bender, D. B., & Rocha-Miranda, C. E. (1969). Visual Receptive Fields

of Neurons in Inferotemporal Cortex of the Monkey. *Science*, *166*(3910), 1303–1306. https://doi.org/10.1126/science.166.3910.1303

Gross, C. G., Schiller, P. H., Wells, C., & Gerstein, G. L. (1967). Single-unit activity in temporal association cortex of the monkey. *Journal of Neurophysiology*, *30*(4), 833–843. https://doi.org/10.1152/jn.1967.30.4.833

Grotheer, M., & Kovács, G. (2016). Can predictive coding explain repetition suppression? *Cortex*, *80*(February), 113–124. https://doi.org/10.1016/j.cortex.2015.11.027

Haenschel, C., Vernon, D. J., Dwivedi, P., Gruzelier, J. H., & Baldeweg, T. (2005). Event-Related Brain Potential Correlates of Human Auditory Sensory Memory-Trace Formation. *Journal of Neuroscience*, *25*(45), 10494–10501. https://doi.org/10.1523/JNEUROSCI.1227-05.2005

Halle, M. (2003). Speculations about the Representations of Words in Memory. In *From Memory to Speech and Back* (pp. 122–136). Berlin, Boston: DE GRUYTER. https://doi.org/10.1515/9783110871258.122

Hari, R., Kaila, K., Katila, T., Tuomisto, T., & Varpula, T. (1982). Interstimulus interval dependence of the auditory vertex response and its magnetic counterpart: Implications for their neural generation. *Electroencephalography and Clinical Neurophysiology*, *54*(5), 561–569. https://doi.org/10.1016/0013-4694(82)90041-4

Heilbron, M., & Chait, M. (2018). Great Expectations: Is there Evidence for Predictive Coding in Auditory Cortex? *Neuroscience*, *389*, 54–73. https://doi.org/10.1016/j.neuroscience.2017.07.061

Hestvik, A., & Durvasula, K. (2016). Neurobiological evidence for voicing underspecification in English. *Brain and Language*, *152*, 28–43. https://doi.org/10.1016/j.bandl.2015.10.007

Hintzman, D. L. (1986). “Schema abstraction” in a multiple-trace memory model. *Psychological Review*, *93*(4), 411–428. https://doi.org/10.1037/0033-295X.93.4.411

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, *95*(4), 528–551. https://doi.org/10.1037/0033-295X.95.4.528

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*(2), 179–185. https://doi.org/10.1007/BF02289447

Horváth, J., Czigler, I., Jacobsen, T., Maess, B., Schröger, E., & Winkler, I. (2008). MMN or no MMN: No magnitude of deviance effect on the MMN amplitude. *Psychophysiology*, *45*(1), 60–69. https://doi.org/10.1111/j.1469-8986.2007.00599.x

Jääskeläinen, I. P., Ahveninen, J., Bonmassar, G., Dale, A. M., Ilmoniemi, R. J., Levänen, S., … Belliveau, J. W. (2004). Human posterior auditory cortex gates novel sounds to consciousness. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(17), 6809–6814. https://doi.org/10.1073/pnas.0303760101

Jacobsen, T., Horenkamp, T., & Schröger, E. (2003). Preattentive Memory-Based Comparison of Sound Intensity. *Audiology and Neurotology*, *8*(6), 338–346. https://doi.org/10.1159/000073518

Jacobsen, T., & Schröger, E. (2001). Is there pre-attentive memory-based comparison of pitch? *Psychophysiology*, *38*(4), S0048577201000993. https://doi.org/10.1017/S0048577201000993

Jacobsen, T., Schröger, E., Horenkamp, T., & Winkler, I. (2003). Mismatch negativity to pitch change: varied stimulus proportions in controlling effects of neural refractoriness on human auditory event-related brain potentials. *Neuroscience Letters*, *344*(2), 79–82. https://doi.org/10.1016/S0304-3940(03)00408-7

Jacoby, L. L. (1983). Remembering the data: analyzing interactive processes in reading. *Journal of Verbal Learning and Verbal Behavior*, *22*(5), 485–508. https://doi.org/10.1016/S0022-5371(83)90301-8

Jacoby, L. L., & Brooks, L. R. (1984). Nonanalytic Cognition: Memory, Perception, and Concept Learning. In *Psychology of Learning and Motivation - Advances in Research and Theory* (Vol. 18, pp. 1–47). https://doi.org/10.1016/S0079-7421(08)60358-8

Jacoby, L. L., & Hayman, C. A. G. (1987). Specific Visual Transfer in Word Identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*(3), 456–463. https://doi.org/10.1037/0278-7393.13.3.456

JASP Team. (2022). JASP (Version 0.16.4)[Computer software]. Retrieved from https://jasp-stats.org/

Javit, D. C., Steinschneider, M., Schroeder, C. E., Vaughan, H. G., & Arezzo, J. C. (1994). Detection of stimulus deviance within primate primary auditory cortex: intracortical mechanisms of mismatch negativity (MMN) generation. *Brain Research*, *667*(2), 192–200. https://doi.org/10.1016/0006-8993(94)91496-6

Joanisse, M. F., Robertson, E. K., & Newman, R. L. (2007). Mismatch negativity reflects sensory and phonetic speech processing. *NeuroReport*, *18*(9), 901–905. https://doi.org/10.1097/WNR.0b013e3281053c4e

Kapnoula, E. C., Winn, M. B., Edwards, J., & Mcmurray, B. (2017). Supplemental Material for Evaluating the Sources and Functions of Gradiency in Phoneme Categorization: An Individual Differences Approach. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(9), 1594–1611. https://doi.org/10.1037/xhp0000410.supp

Kasai, K., Nakagome, K., Iwanami, A., Fukuda, M., Itoh, K., Koshida, I., & Kato, N. (2002). No effect of gender on tonal and phonetic mismatch negativity in normal adults assessed by a high-resolution EEG recording. *Cognitive Brain Research*, *13*(3), 305–312. https://doi.org/10.1016/S0926-6410(01)00125-2

Kazanina, N., Phillips, C., & Idsardi, W. (2006). The influence of meaning on the perception of speech sounds. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(30), 11381–11386. https://doi.org/10.1073/pnas.0604821103

Keysers, C., Gazzola, V., & Wagenmakers, E. (2020). Using Bayes factor hypothesis testing in neuroscience to establish evidence of absence. *Nature Neuroscience*, *23*(7), 788–799. https://doi.org/10.1038/s41593-020-0660-4

Kirihara, K., Tada, M., Koshiyama, D., Fujioka, M., Usui, K., Araki, T., & Kasai, K. (2020). A Predictive Coding Perspective on Mismatch Negativity Impairment in Schizophrenia. *Frontiers in Psychiatry*, *11*(July), 1–8. https://doi.org/10.3389/fpsyt.2020.00660

Korzyukov, O. A., Winkler, I., Gumenyuk, V. I., & Alho, K. (2003). Processing abstract auditory features in the human auditory cortex. *NeuroImage*, *20*(4), 2245–2258. https://doi.org/10.1016/j.neuroimage.2003.08.014

Korzyukov, O., Alho, K., Kujala, A., Gumenyuk, V., Ilmoniemi, R. J., Virtanen, J., … Näätänen, R. (1999). Electromagnetic responses of the human auditory cortex generated by sensory-memory based processing of tone-frequency changes. *Neuroscience Letters*, *276*(3), 169–172. https://doi.org/10.1016/S0304-3940(99)00807-1

Kronrod, Y., Coppess, E., & Feldman, N. H. (2016). A unified account of categorical effects in phonetic perception. *Psychonomic Bulletin & Review*, *23*(6), 1681–1712. https://doi.org/10.3758/s13423-016-1049-y

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. https://doi.org/10.1016/j.biopsycho.2006.06.001

Lecaignard, F., Bertrand, O., Gimenez, G., Mattout, J., & Caclin, A. (2015). Implicit learning of predictable sound sequences modulates human brain responses at different levels of the auditory hierarchy. *Frontiers in Human Neuroscience*, *9*(September), 1–14. https://doi.org/10.3389/fnhum.2015.00505

Lenth, R. V. (2021). emmeans: Estimated Marginal Means, aka Least-Squares Means. Retrieved from https://cran.r-project.org/package=emmeans

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*(5), 358–368. https://doi.org/10.1037/h0044417

Lipski, S. C., & Mathiak, K. (2008). Auditory mismatch negativity for speech sound contrasts is modulated by language context. *NeuroReport*, *19*(10), 1081–1084. https://doi.org/10.1097/WNR.0b013e3283056378

Lister, J., & Tarver, K. (2004). Effect of Age on Silent Gap Discrimination in Synthetic Speech Stimuli. *Journal of Speech, Language, and Hearing Research*, *47*(2), 257–268. https://doi.org/10.1044/1092-4388(2004/021)

Liu, L., Yuan, C., Ong, J. H., Tuninetti, A., Antoniou, M., Cutler, A., & Escudero, P. (2022). Learning to Perceive Non-Native Tones via Distributional Training: Effects of Task and Acoustic Cue Weighting. *Brain Sciences*, *12*(5), 559. https://doi.org/10.3390/brainsci12050559

Lu, K., Liu, W., Zan, P., David, S. V., Fritz, J. B., & Shamma, S. A. (2018). Implicit memory for complex sounds in higher auditory cortex of the ferret. *Journal of Neuroscience*, *38*(46), 9955–9966. https://doi.org/10.1523/JNEUROSCI.2118-18.2018

Lüdecke, D., Ben-Shachar, M., Patil, I., & Makowski, D. (2020). Extracting, Computing and Exploring the Parameters of Statistical Models using R. *Journal of Open Source Software*, *5*(53), 2445. https://doi.org/10.21105/joss.02445

Makowski, D., Ben-Shachar, M. S., Patil, I., & Lüdecke, D. (2021). Automated Results Reporting as a Practical Tool to Improve Reproducibility and Methodological Best Practices Adoption. *CRAN*. Retrieved from https://github.com/easystats/report

Marklund, E., Lacerda, F., & Schwarz, I.-C. (2018). Using rotated speech to

approximate the acoustic mismatch negativity response to speech. *Brain and Language*, *176*(October 2017), 26–35. https://doi.org/10.1016/j.bandl.2017.10.006

Mathys, C. (2011). A Bayesian foundation for individual learning under uncertainty. *Frontiers in Human Neuroscience*, *5*(MAY), 9. https://doi.org/10.3389/fnhum.2011.00039

May, P J C, & Tiitinen, H. (2004). Auditory scene analysis and sensory memory: the role of the auditory N100m. *Neurology & Clinical Neurophysiology : NCN*, *2004*, 19. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16015713

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

McCloskey, M. E., & Glucksberg, S. (1978). Natural categories: Well defined or fuzzy sets? *Memory & Cognition*, *6*(4), 462–472. https://doi.org/10.3758/BF03197480

McMurray, B., Aslin, R. N., Tanenhaus, M. K., Spivey, M. J., & Subik, D. (2008). Gradient sensitivity to within-category variation in words and syllables. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(6), 1609–1631. https://doi.org/10.1037/a0011747

McMurray, B., Tanenhaus, M. K., Aslin, R. N., & Spivey, M. J. (2003). Probabilistic constraint satisfaction at the lexical/phonetic interface: evidence for gradient effects of within-category VOT on lexical access. *Journal of Psycholinguistic Research*, *32*(1), 77–97. https://doi.org/10.1023/a:1021937116271

Mervis, C. B., & Rosch, E. (1981). Categorization of Natural Objects. *Annual Review of Psychology*, *32*(1), 89–115. https://doi.org/10.1146/annurev.ps.32.020181.000513

Miglietta, S., Grimaldi, M., & Calabrese, A. (2013). Conditioned allophony in speech perception: An ERP study. *Brain and Language*, *126*(3), 285–290. https://doi.org/10.1016/j.bandl.2013.06.001

Miller, E., & Desimone, R. (1994). Parallel neuronal mechanisms for short-term memory. *Science*, *263*(5146), 520–522. https://doi.org/10.1126/science.8290960

Miller, E. K., Li, L., & Desimone, R. (1991). A Neural Mechanism for Working and Recognition Memory in Inferior Temporal Cortex. *Science*, *254*(5036), 1377–1379. https://doi.org/10.1126/science.1962197

Miller, EK, Li, L., & Desimone, R. (1993). Activity of neurons in anterior inferior temporal cortex during a short- term memory task. *The Journal of Neuroscience*, *13*(4), 1460–1478. https://doi.org/10.1523/JNEUROSCI.13-04-01460.1993

Miller, J. L. (1994). On the internal structure of phonetic categories: a progress report. *Cognition*, *50*(1–3), 271–285. https://doi.org/10.1016/0010-0277(94)90031-0

Monahan, P. J., Schertz, J., Fu, Z., & Pérez, A. (2022). Unified Coding of Spectral and Temporal Phonetic Cues: Electrophysiological Evidence for Abstract Phonological Features. *Journal of Cognitive Neuroscience*, *34*(4), 618–638. https://doi.org/10.1162/jocn_a_01817

Müller, D., Widmann, A., & Schröger, E. (2005). Deviance-repetition effects as a function of stimulus feature, feature value variation, and timing: a mismatch negativity study. *Biological Psychology*, *68*(1), 1–14. https://doi.org/10.1016/j.biopsycho.2004.03.018

Näätänen, R., Gaillard, A. W. K., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica*, *42*(4), 313–329. https://doi.org/10.1016/0001-6918(78)90006-9

Näätänen, R., Kujala, T., Escera, C., Baldeweg, T., Kreegipuu, K., Carlson, S., & Ponton, C. (2012). The mismatch negativity (MMN) - A unique window to disturbed central auditory processing in ageing and different clinical conditions. *Clinical Neurophysiology*, *123*(3), 424–458. https://doi.org/10.1016/j.clinph.2011.09.020

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. https://doi.org/10.1016/j.clinph.2007.04.026

Näätänen, Risto. (1990). The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. *Behavioral and Brain Sciences*, *13*(2), 201–233. https://doi.org/10.1017/S0140525X00078407

Näätänen, Risto. (1992). *Attention and Brain Function* (1st ed.). Routledge. https://doi.org/10.4324/9780429487354

Näätänen, Risto, & Alho, K. (1995). Mismatch negativity-a unique measure of sensory processing in audition. *International Journal of Neuroscience*, *80*(1–4), 317–337. https://doi.org/10.3109/00207459508986107

Näätänen, Risto, & Alho, K. (1997). Higher-order processes in auditory-change detection. *Trends in Cognitive Sciences*, *1*(2), 44–45. https://doi.org/10.1016/S1364-6613(97)01013-9

Näätänen, Risto, Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., … Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, *385*(6615), 432–434. https://doi.org/10.1038/385432a0

Näätänen, Risto, Paavilainen, P., Alho, K., Reinikainen, K., & Sams, M. (1989). Do event-related potentials reveal the mechanism of the auditory sensory memory in the human brain? *Neuroscience Letters*, *98*(2), 217–221. https://doi.org/10.1016/0304-3940(89)90513-2

Näätänen, Risto, Pakarinen, S., Rinne, T., & Takegata, R. (2004). The mismatch negativity (MMN): towards the optimal paradigm. *Clinical Neurophysiology*, *115*(1), 140–144. https://doi.org/10.1016/j.clinph.2003.04.001

Näätänen, Risto, & 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. https://doi.org/10.1111/j.1469-8986.1987.tb00311.x

Näätänen, Risto, Schröger, E., Karakas, S., Tervaniemi, M., & Paavilainen, P. (1993). Development of a memory trace for a complex sound in the human brain. *NeuroReport*, *4*(5), 503–506. https://doi.org/10.1097/00001756-199305000-00010

NELKEN, I. (2004). Processing of complex stimuli and natural scenes in the auditory cortex. *Current Opinion in Neurobiology*, *14*(4), 474–480. https://doi.org/10.1016/j.conb.2004.06.005

Nosofsky, R. M. (1988). Similarity, frequency, and category representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(1), 54–65. https://doi.org/10.1037/0278-7393.14.1.54

O'Shea, R. P. (2015). Refractoriness about adaptation. *Frontiers in Human Neuroscience*, *9*(December), 139. https://doi.org/10.3389/fnhum.2015.00038

PAAVILAINEN, P., SIMOLA, J., JARAMILLO, M., NÄÄTÄNEN, R., & WINKLER, I. (2001). Preattentive extraction of abstract feature conjunctions from auditory stimulation as reflected by the mismatch negativity (MMN). *Psychophysiology*, *38*(2), S0048577201000920. https://doi.org/10.1017/S0048577201000920

Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(2), 309–328. https://doi.org/10.1037/0278-7393.19.2.309

Peter, V., McArthur, G., & Thompson, W. F. (2010). Effect of deviance direction and calculation method on duration and frequency mismatch negativity (MMN). *Neuroscience Letters*, *482*(1), 71–75. https://doi.org/10.1016/j.neulet.2010.07.010

Pettigrew, C. M., Murdoch, B. M., Kei, J., Chenery, H. J., Sockalingam, R., Ponton, C. W., … Alku, P. (2004). Processing of English Words with Fine Acoustic Contrasts and Simple Tones: A Mismatch Negativity Study. *Journal of the American Academy of Audiology*, *15*(01), 047–066. https://doi.org/10.3766/jaaa.15.1.6

Phillips, C., Marantz, A., McGinnis, M., Pesetsky, D., Wexler, K., & Yellin, E. (1995). Brain Mechanisms of Speech Perception: A Preliminary Report. *Massachusetts Institute of Technology Working Papers in Linguistics*, *26*(Sept), 191.

Phillips, C., Pellathy, T., Marantz, A., Yellin, E., Wexler, K., Poeppel, D., … Roberts, T. (2000). Auditory cortex accesses phonological categories: an MEG mismatch study. *Journal of Cognitive Neuroscience*, *12*(6), 1038–1055. https://doi.org/10.1162/08989290051137567

Pierrehumbert, J. B. (2001). Exemplar dynamics. In J. Bybee & P. Hopper (Eds.), *Frequency and the Emergence of Linguistic Structure* (p. 137). Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.45.08pie

Pisoni, D. B., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics*, *15*(2), 285–290. https://doi.org/10.3758/BF03213946

Politzer-Ahles, S., Schluter, K., Wu, K., & Almeida, D. (2016). Asymmetries in the perception of Mandarin tones: Evidence from mismatch negativity. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(10), 1547–1570. https://doi.org/10.1037/xhp0000242

Port, R. F. (2007). The graphical basis of phones and phonemes (pp. 349–365). https://doi.org/10.1075/lllt.17.29por

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, *77*(3, Pt.1), 353–363.

https://doi.org/10.1037/h0025953

R Core Team. (2021). R: A Language and Environment for Statistical Computing. Vienna, Austria. Retrieved from https://www.r-project.org/

Rao, R. P. N., & Ballard, D. H. (1999). Hierarchical Predictive Coding Model Hierarchical Predictive Coding of Natural Images. *Nature Neuroscience*, *2*(1), 79–87. Retrieved from http://neurosci.nature.com

Ren, Y., Allenmark, F., Müller, H. J., & Shi, Z. (2020). Logarithmic encoding of ensemble time intervals. *Scientific Reports*, *10*(1), 18174. https://doi.org/10.1038/s41598-020-75191-6

Rentzsch, J., Shen, C., Jockers-Scherübl, M. C., Gallinat, J., & Neuhaus, A. H. (2015). Auditory Mismatch Negativity and Repetition Suppression Deficits in Schizophrenia Explained by Irregular Computation of Prediction Error. *PLOS ONE*, *10*(5), e0126775. https://doi.org/10.1371/journal.pone.0126775

Rhodes, R., Avcu, E., Han, C., & Hestvik, A. (2022). Auditory predictions are phonological when phonetic information is variable. *Language, Cognition and Neuroscience*, 1–16. https://doi.org/10.1080/23273798.2022.2043395

Rhodes, R., Han, C., & Hestvik, A. (2019). Phonological memory traces do not contain phonetic information. *Attention, Perception, & Psychophysics*, *81*(4), 897–911. https://doi.org/10.3758/s13414-019-01728-1

Rinne, T., Särkkä, A., Degerman, A., Schröger, E., & Alho, K. (2006). Two separate mechanisms underlie auditory change detection and involuntary control of attention. *Brain Research*, *1077*(1), 135–143. https://doi.org/10.1016/j.brainres.2006.01.043

Ritter, W., Vaughan, H. G., & Costa, L. D. (1968). Orienting and habituation to auditory stimuli: A study of short terms changes in average evoked responses. *Electroencephalography and Clinical Neurophysiology*, *25*(6), 550–556. https://doi.org/10.1016/0013-4694(68)90234-4

Rosburg, T., Trautner, P., Boutros, N. N., Korzyukov, O. A., Schaller, C., Elger, C. E., & Kurthen, M. (2006). Habituation of auditory evoked potentials in intracranial and extracranial recordings. *Psychophysiology*, *43*(2), 137–144. https://doi.org/10.1111/j.1469-8986.2006.00391.x

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, *104*(3), 192–233. https://doi.org/10.1037/0096-3445.104.3.192

Rosch, E. (1988). Principles of Categorization. In *Readings in Cognitive Science* (pp. 312–322). Elsevier. https://doi.org/10.1016/B978-1-4832-1446-7.50028-5

Rosch, E. H. (1973). Natural categories. *Cognitive Psychology*, *4*(3), 328–350. https://doi.org/10.1016/0010-0285(73)90017-0

Saarinen, J., Paavilainen, P., Schöger, E., Tervaniemi, M., & Näätänen, R. (1992). Representation of abstract attributes of auditory stimuli in the human brain. *NeuroReport*, *3*(12), 1149–1151. https://doi.org/10.1097/00001756-199212000-00030

Sams, M., Alho, K., & Näätänen, R. (1984). Short-Term Habituation and Dishabituation of the Mismatch Negativity of the ERP. *Psychophysiology*, *21*(4), 434–441. https://doi.org/10.1111/j.1469-8986.1984.tb00223.x

Scharinger, M., Bendixen, A., Trujillo-Barreto, N. J., & Obleser, J. (2012). A sparse neural code for some speech sounds but not for others. *PLoS ONE*, *7*(7). https://doi.org/10.1371/journal.pone.0040953

Scharinger, M., Merickel, J., Riley, J., & Idsardi, W. J. (2011). Neuromagnetic evidence for a featural distinction of English consonants: Sensor- and source-space data. *Brain and Language*, *116*(2), 71–82. https://doi.org/10.1016/j.bandl.2010.11.002

Scharinger, M., Monahan, P. J., & Idsardi, W. J. (2012). Asymmetries in the Processing of Vowel Height. *Journal of Speech Language and Hearing Research*, *55*(3), 903. https://doi.org/10.1044/1092-4388(2011/11-0065)

Scharinger, M., Monahan, P. J., & Idsardi, W. J. (2016). Linguistic category structure influences early auditory processing: Converging evidence from mismatch responses and cortical oscillations. *NeuroImage*, *128*, 293–301. https://doi.org/10.1016/j.neuroimage.2016.01.003

Scheler, G. (2017). Logarithmic distributions prove that intrinsic learning is Hebbian. *F1000Research*, *6*, 1222. https://doi.org/10.12688/f1000research.12130.2

Schluter, K., Politzer-Ahles, S., Al-Kaabi, M., & Almeida, D. (2017). Laryngeal Features are Phonetically Abstract. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2017.00746

Schluter, K., Politzer-Ahles, S., & Almeida, D. (2016). No place for /h/: an ERP investigation of English fricative place features. *Language, Cognition and Neuroscience*, *31*(6), 728–740. https://doi.org/10.1080/23273798.2016.1151058

Schröger, E. (1996). The influence of stimulus intensity and inter-stimulus interval on the detection of pitch and loudness changes. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *100*(6), 517–526. https://doi.org/10.1016/S0168-5597(96)95576-8

Schröger, E., Paavilainen, P., & Näätänen, R. (1994). Mismatch negativity to changes in a continuous tone with regularly varying frequencies. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *92*(2), 140–147. https://doi.org/10.1016/0168-5597(94)90054-X

Schröger, E., & Wolff, C. (1996). Mismatch response of the human brain to changes in sound location. *NeuroReport*, *7*(18), 3005–3008. https://doi.org/10.1097/00001756-199611250-00041

Shafer, V. L., Kresh, S., Ito, K., Hisagi, M., Vidal, N., Higby, E., … Strange, W. (2021). The neural timecourse of American English vowel discrimination by Japanese, Russian and Spanish second-language learners of English. *Bilingualism: Language and Cognition*, *24*(4), 642–655. https://doi.org/10.1017/S1366728921000201

Sharma, A., Kraus, N., McGee, T., Carrell, T., & Nicol, T. (1993). Acoustic versus phonetic representation of speech as reflected by the mismatch negativity event-related potential. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *88*(1), 64–71. https://doi.org/10.1016/0168-5597(93)90029-O

Sharma, Anu, & Dorman, M. F. (1999). Cortical auditory evoked potential correlates of categorical perception of voice-onset time. *The Journal of the Acoustical Society of America*, *106*(2), 1078–1083. https://doi.org/10.1121/1.428048

Shestakova, A., Brattico, E., Huotilainen, M., Galunov, V., Soloviev, A., Sams, M., … Näätänen, R. (2002). Abstract phoneme representations in the left temporal cortex: magnetic mismatch negativity study. *NeuroReport*, *13*(14), 1813–1816. https://doi.org/10.1097/00001756-200210070-00025

Silva, D. M. R., Melges, D. B., & Rothe-Neves, R. (2017). N1 response attenuation and the mismatch negativity (MMN) to within- and across-category phonetic contrasts. *Psychophysiology*, *54*(4), 591–600. https://doi.org/10.1111/psyp.12824

Silva, D. M. R., Rothe-Neves, R., & Melges, D. B. (2020). Long-latency event-related responses to vowels: N1-P2 decomposition by two-step principal component analysis. *International Journal of Psychophysiology*, *148*(November 2019), 93–102. https://doi.org/10.1016/j.ijpsycho.2019.11.010

Sittiprapaporn, W., Tervaniemi, M., Chindaduangratn, C., & Kotchabhakdi, N. (2005).

Preattentive discrimination of across-category and within-category change in consonant–vowel syllable. *NeuroReport*, *16*(13), 1513–1518. https://doi.org/10.1097/01.wnr.0000175618.46677.07

Soli, S. D. (1983). The role of spectral cues in discrimination of voice onset time differences. *The Journal of the Acoustical Society of America*, *73*(6), 2150–2165. https://doi.org/10.1121/1.389539

Stefanics, G., Kremláček, J., & Czigler, I. (2014). Visual mismatch negativity: a predictive coding view. *Frontiers in Human Neuroscience*, *8*(September), 1–19. https://doi.org/10.3389/fnhum.2014.00666

Summerfield, C., Wyart, V., Johnen, V. M., & de Gardelle, V. (2011). Human Scalp Electroencephalography Reveals that Repetition Suppression Varies with Expectation. *Frontiers in Human Neuroscience*, *5*(JULY), 1–13. https://doi.org/10.3389/fnhum.2011.00067

Sussman, E. S., Chen, S., Sussman-Fort, J., & Dinces, E. (2014). The Five Myths of MMN: Redefining How to Use MMN in Basic and Clinical Research. *Brain Topography*, *27*(4), 553–564. https://doi.org/10.1007/s10548-013-0326-6

Tervaniemi, M., Maury, S., & Näätänen, R. (1994). Neural representations of abstract stimulus features in the human brain as reflected by the mismatch negativity. *NeuroReport*, *5*(7), 844–846. https://doi.org/10.1097/00001756-199403000-00027

Thompson, R. F., & Spencer, W. A. (1966). Habituation: A model phenomenon for the study of neuronal substrates of behavior. *Psychological Review*, *73*(1), 16–43. https://doi.org/10.1037/h0022681

Todorovic, A., & de Lange, F. P. (2012). Repetition Suppression and Expectation Suppression Are Dissociable in Time in Early Auditory Evoked Fields. *Journal of Neuroscience*, *32*(39), 13389–13395. https://doi.org/10.1523/JNEUROSCI.2227-12.2012

Toscano, J. C., McMurray, B., Dennhardt, J., & Luck, S. J. (2010). Continuous perception and graded categorization: Electrophysiological evidence for a linear relationship between the acoustic signal and perceptual encoding of speech. *Psychological Science*, *21*(10), 1532–1540. https://doi.org/10.1177/0956797610384142

Ulanovsky, N., Las, L., & Nelken, I. (2003). Processing of low-probability sounds by cortical neurons. *Nature Neuroscience*, *6*(4), 391–398. https://doi.org/10.1038/nn1032

Wacongne, C., Changeux, J.-P., & Dehaene, S. (2012). A Neuronal Model of Predictive Coding Accounting for the Mismatch Negativity. *Journal of Neuroscience*, *32*(11), 3665–3678. https://doi.org/10.1523/JNEUROSCI.5003-11.2012

Winkler., Istvan, Paavilainen, P., Alho, K., Reinikainen, K., Sams, M., & Naatanen, R. (1990). The Effect of Small Variation of the Frequent Auditory Stimulus on the Event-Related Brain Potential to the Infrequent Stimulus. *Psychophysiology*, *27*(2), 228–235. https://doi.org/10.1111/j.1469-8986.1990.tb00374.x

WINKLER, I. 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), S0048577299981908. https://doi.org/10.1017/S0048577299981908

Winkler, István. (2007). Interpreting the Mismatch Negativity. *Journal of Psychophysiology*, *21*(3–4), 147–163. https://doi.org/10.1027/0269-8803.21.34.147

Winkler, István, Reinikainen, K., & Näätänen, R. (1993). Event-related brain potentials reflect traces of echoic memory in humans. *Perception & Psychophysics*, *53*(4), 443–449. https://doi.org/10.3758/BF03206788

Winkler, István, & Schröger, E. (1995). Neural representation for the temporal structure of sound patterns. *NeuroReport*, *6*(4), 690–694. https://doi.org/10.1097/00001756-199503000-00026

Winn, M. B. (2020). Manipulation of voice onset time in speech stimuli: A tutorial and flexible Praat script. *The Journal of the Acoustical Society of America*, *147*(2), 852–866. https://doi.org/10.1121/10.0000692

Yabe, H., Tervaniemi, M., Reinikainen, K., & Näätänen, R. (1997). Temporal window of integration revealed by MMN to sound omission. *NeuroReport*, *8*(8), 1971–1974. https://doi.org/10.1097/00001756-199705260-00035

Ylinen, S., Shestakova, A., Huotilainen, M., Alku, P., & Näätänen, R. (2006). Mismatch negativity (MMN) elicited by changes in phoneme length: A cross-linguistic study. *Brain Research*, *1072*(1), 175–185. https://doi.org/10.1016/j.brainres.2005.12.004

Yu, Y. H., Shafer, V. L., & Sussman, E. S. (2017). Neurophysiological and Behavioral Responses of Mandarin Lexical Tone Processing. *Frontiers in Neuroscience*, *11*(March), 1–19. https://doi.org/10.3389/fnins.2017.00095

Appendix A

MMN IN ROVING-STANDARD CONTROL BLOCK

Here I briefly report the two MMN responses elicited by the 19ms VOT and the 119ms VOT in the roving-standard control block. The MMN was computed by comparing a deviant stimulus (19ms or 119ms) to the same stimulus serving as the standards in the same block. Separate PCAs were run for the 19ms VOT and the 119ms VOT. The PCA solution for the 19ms VOT selected a time window of 524-588ms and five channels: E3, E4, E6, E8, and E9. The PCA solution for the 119ms VOT selected a time window of 496-664ms and five channels: E2, E3, E5, E59, and E60. Note that both time windows were later than a typical MMN response, probably reflecting a late discriminative negativity.

The following figure shows the waveforms (averaged over subjects and delimited channels) of the same stimulus serving as standards and deviants.

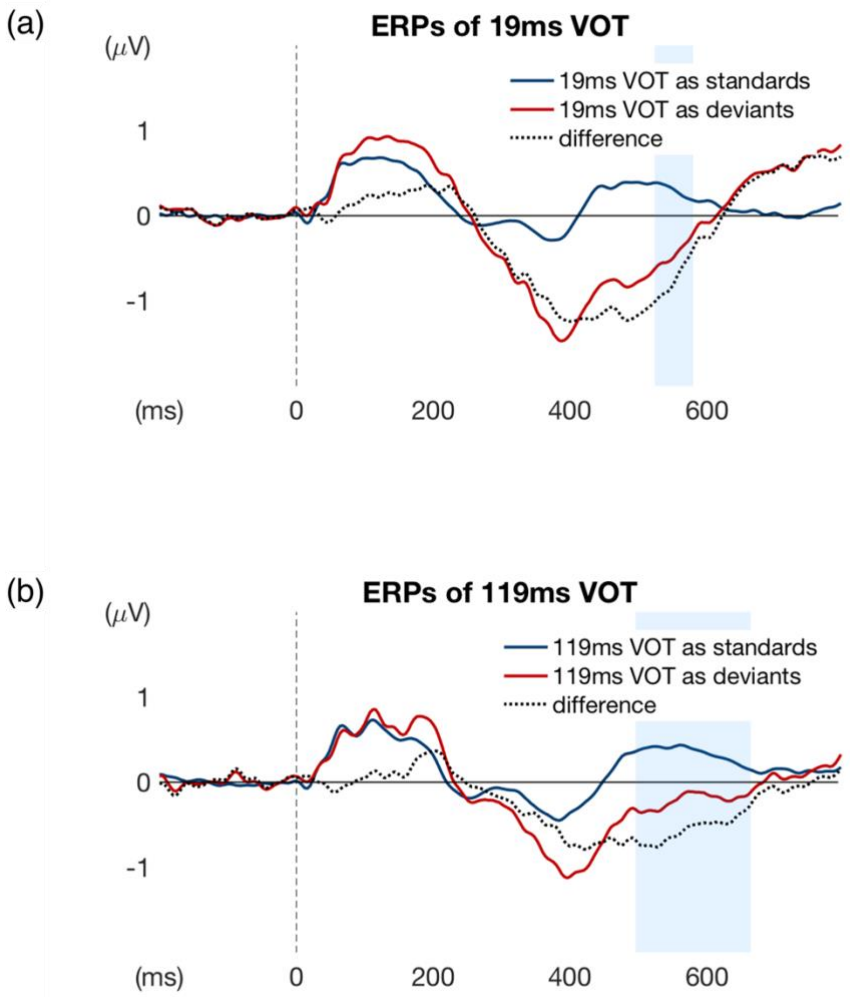


Figure 43: ERP waveforms averaged over subjects and the delimited channels. The blue shaded area indicates the time window for analysis (524-588ms for the 19ms VOT; 496-664ms for the 119ms VOT).

The figure below presents a violin plot showing the ERP averaged over the selected time window and channels for each subject.

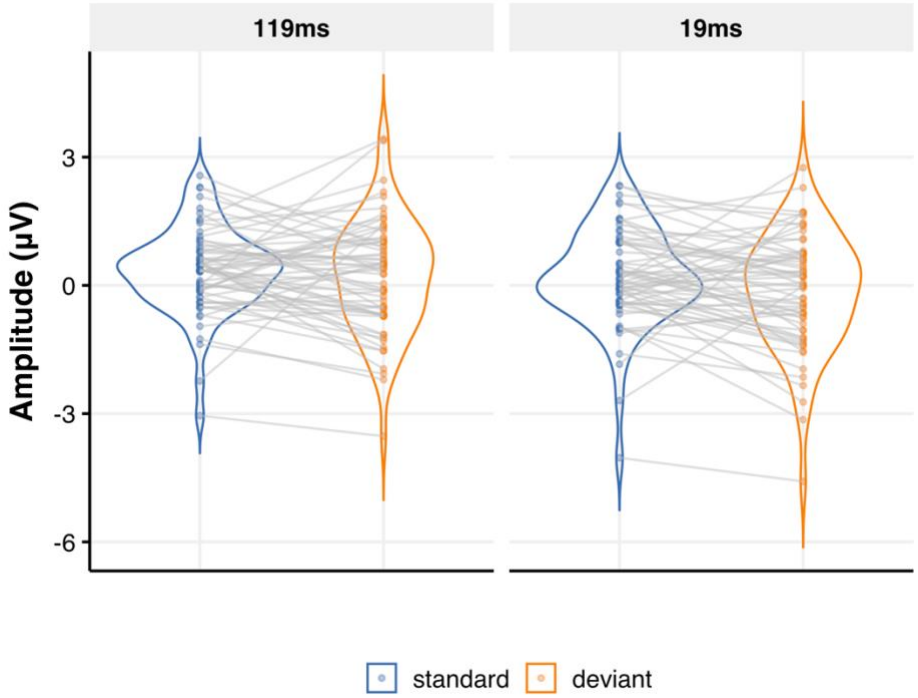


Figure 44: Individual ERPs averaged over the selected time window and the selected channels as a function of VOT and stimulus type. Each dot represents one subject's data for a given condition. The grey line connects two data points from the same subject, indicating the amplitude change between standards and deviants. The shape of the violin plots indicates the data distribution.

For the statistical analysis, I used the averaged ERP amplitude as the dependent measure. The mixed-effects model with two fixed factors and their interaction. The two factors were Stimulus (standard vs. deviant) and VOT (19ms vs. 119ms). The model also included Subject as a random intercept. I report below the fixed factors' coefficients, the corresponding standard errors (SE), 95% confidence intervals (CI), t values, and p values.

Table 13: Model summary

Fixed factors	**Coefficient**	**SE**	**95% CI**	**t(242)**	**p**
(Intercept)	0.15	0.13	[-0.10, 0.40]	1.17	0.243
Stimulus	-0.22	0.10	[-0.42, -0.02]	-2.15	0.032*
VOT	-0.27	0.10	[-0.47, -0.07]	-2.62	0.009**
Stimulus × Group	-0.18	0.20	[-0.59, 0.22]	-0.90	0.369

There is a main effect of Stimulus [t(242) = -2.15, p = .032]. The partial eta squared (η_p^2) associated with the effect is 0.02, indicating a small effect size. To further examine the difference between the standard ERP and the deviant ERP for each VOT stimulus, I ran one-tailed t-tests on the effect of Stimulus within each level of VOT. There was a small effect of Stimulus for the 19ms VOT [t(183) = 2.157, p = 0.0161, Cohen's d = 0.16], while the effect of Stimulus was not significant for the 119ms VOT [t(183) = 0.885, p = 0.1887]. The figure below shows the bar plot of the averaged ERP for each condition.

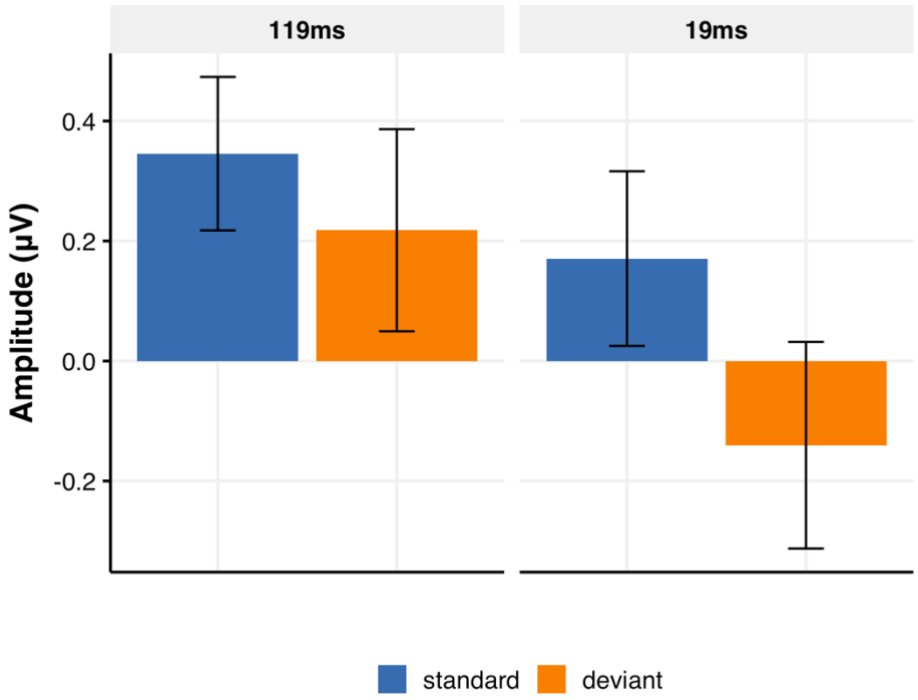


Figure 45: ERP amplitude averaged over subjects, selected time points, and channels. The error bar indicates standard error.

Appendix B

IRB/HUMAN SUBJECTS APPROVAL

DATE:	January 28, 2021
TO:	Arild Hestvik
FROM:	University of Delaware IRB
STUDY TITLE:	[1691848-1] Evoked Category Representations
SUBMISSION TYPE:	Funding/Grant
ACTION:	APPROVED
EFFECTIVE DATE:	January 26, 2021
NEXT REPORT DUE:	January 20, 2022
REVIEW TYPE:	Expedited Review
REVIEW CATEGORY:	Expedited review category # *(4)*

Thank you for your Funding/Grant submission to the University of Delaware Institutional Review Board (UD IRB). The UD IRB has reviewed and APPROVED the proposed research and submitted documents via Expedited Review in compliance with the pertinent federal regulations.

As the Principal Investigator for this study, you are responsible for, and agree that:

> All research must be conducted in accordance with the protocol and all other study forms as approved in this submission. Any revisions to the approved study procedures or documents must be reviewed and approved by the IRB prior to their implementation. Please use the UD amendment form to request the review of any changes to approved study procedures or documents.
>
> Informed consent is a process that must allow prospective participants sufficient opportunity to discuss and consider whether to participate. IRB-approved and stamped consent documents must be used when enrolling participants and a written copy shall be given to the person signing the informed consent form.
>
> Unanticipated problems, serious adverse events involving risk to participants, and all non-compliance issues must be reported to this office in a timely fashion according with the UD requirements for reportable events. All sponsor reporting requirements must also be followed.

The UD IRB REQUIRES the submission of a PROGRESS REPORT DUE ON January 20, 2022. A continuing review/progress report form must be submitted to the UD IRB at least 45 days prior to the due date to allow for the review of that report.

If you have any questions, please contact the UD IRB Office at (302) 831-2137 or via email at hsrb-research@udel.edu. Please include the study title and reference number in all correspondence with this office.

INSTITUTIONAL REVIEW BOARD