Detecting integration of top-down information using the mismatch negativity: Preliminary evidence from phoneme restoration

Charles Redmon[a,1], Yuyu Zeng[a], Juhi Kidwai[b], Xiao Yang[a], Delaney Wilson[a], and Robert Fiorentino[a]

[a]Department of Linguistics, [b]Department of Speech-Language-Hearing
University of Kansas

The current study utilizes mismatch negativity in the phenomenon of phoneme restoration to investigate the critical debate regarding the integration of top down (lexical) and bottom up (acoustic) processing in spoken word recognition. Phoneme restoration, which occurs when phonemes missing from a speech signal are restored by the brain and may appear to be heard, was examined in a multi-standard oddball paradigm. Participants heard stimuli while watching a quiet animated film. Stimuli were divided into word and non-word conditions, with noise added to some stimuli to make them ambiguous. The many-to-one ratio of standards to deviants for generation of mismatch negativity (MMN) was achieved only if the brain could recover the missing phoneme in the ambiguous, noise-spliced stimuli. Both word and nonword conditions were compared to verify that an elicited MMN among words was contingent on involvement of the lexicon in the grouping of standards, and not some more general cognitive grouping procedure. Results from seven participants show preliminary support for the predicted effect: i.e., mismatch negativity for words but not for nonwords. This effect is contingent on phoneme restoration, and thus is consistent with recent literature suggesting that MMN is sensitive to higher information structures such as the mental lexicon.
Keywords: phoneme restoration, MMN, lexical access, top-down information

1. Introduction

1.1. Background. At present there is a growing consensus in the psycholinguistic literature on spoken word recognition that both bottom-up (acoustic) and top-down (lexical) sources of information are utilized as the speech signal unfolds (e.g., Sohoglu, Peelle, Carlyon, & Davis, 2012). Incoming acoustic information is utilized by the auditory system immediately from its reception is well agreed upon, but what remains a matter of debate and a growing focus of research in this area is the question of when top-down information is applied in the process (Norris et al., 2000; Samuel & Pitt, 2003; Magnuson et al., 2003; McQueen, 2003; McQueen et al., 2009). One phenomenon that has shed light on this interplay between bottom-up and top-down information is that of *phoneme restoration*, which broadly is defined as the maintenance of a constant lexical percept despite the replacement of phonemic information by noise in the signal (Samuel, 1996). In a 2001 study, Samuel used a selective adaptation paradigm to demonstrate that the influences of top-down lexical information on the processing of degraded acoustic information do not occur at a post-lexical decision stage, but rather are active earlier at a 'sub-lexical' stage. Briefly, Samuel spliced a fricative ambiguous between [s] and [ʃ] into the consonant interval at the end of words such as 'bronchitis' and 'demolish', where it is expected to be perceived as /s/ and /ʃ/, respectively, based on top-down information. Listeners were presented with either /s/-biasing items or /ʃ/-biasing items in an *adaptation*

---

phase, and then in the test phase were given an /ɪs/–/ɪʃ/ continuum to identify as /s/ or /ʃ/. Following a /s/-biasing adaptation phase, listeners perceived a greater portion of the continuum as /s/, and vice versa for the /ʃ/-biasing adaptation phase. Further, the same pattern of results was obtained when white noise was cross-spliced onto stimuli instead of ambiguous fricatives. This result, according to Samuel (2001), reflects the influence of lexical information at a subconscious level, which is not driven by attention but rather is automatic in spoken word recognition. However, this conclusion relies on assumptions about the role of attention in selective adaptation experiments, which is difficult to test. An alternative index of auditory processing which is also sensitive to linguistic structure, but nevertheless is argued to be pre-attentive in scope is the mismatch negativity (MMN) response commonly used in neurolinguistic research (Shtyrov & Pulvermüller, 2007).[2]

MMN, and its MEG counterpart, the mismatch field (MMF), is a brain response which has been shown to reflect auditory responses to a range of phenomena, from changes in pure tones (e.g., Sams et al., 1985) to linguistically relevant phonemic categories (e.g., Phillips et al., 2000) to higher-order units such as words (e.g., Pulvermüller et al., 2001). In an oddball paradigm, a long sequence of stimuli (typically between 4 and 7) comprising the same unit (the 'standard') are followed by a stimulus (the 'deviant') that differs from the standard. MMN is a negative-going wave for the deviant relative to the standard, peaking between 100 and 250 ms after the onset of the deviant (Näätänen et al., 2007).

The range of possible processing stages that the MMN is sensitive to has been the source of much debate among researchers (Näätänen et al., 2007; Pettigrew et al., 2004). For example, research using MMN to study speech perception must rule out the potential confound of pure (non-linguistic) acoustic processing (e.g., Phillips et al., 2000). As research further investigates higher-level cognitive processes (by using, for instance, MMN to investigate morpho-syntactic processing), it becomes increasingly critical to distinguish between lower-level and higher-level cognitive processes. Though many attempts have been made to utilize the MMN response to investigate morpho-syntactic processing, to date, little work has been done using MMN to study the involvement of top-down information in parsing the auditory signal (for a review of the literature, see Beres, 2017).

One study which utilized EEG to examine the role of top-down information on speech processing was Hannemann et al. (2007). Hannemann and colleagues presented their participants with degraded speech and found an enhancement in the induced gamma band activity (GBA) at left temporal electrode sites around 350 ms, which they argue reflects access from the speech signal to long-term memory representations. Importantly, this enhancement was present only for words correctly identified. Hannemann and colleagues interpreted this enhancement as reflecting a matching process of top-down lexical memory traces with degraded sensory input to form a coherent speech percept. Hannemann (2008) extended Hannemann et al. (2007) by studying degraded German nouns (using moving average degradation; i.e., a window of samples in the waveform was replaced by its average value) and replicated the enhanced GBA effect. In the 2008 study, Hannemann used a roving-oddball design (i.e., standards

---

[2] The MMN response, as elicited in the standard passive listening paradigm, is argued to reflect automatic, pre-attentive processing because the participant is not given any explicit task and is asked to attend away from the stimuli, usually by watching a silent movie (Näätänen et al., 2007).

varied in acoustic characteristics and sequence lengths, with the final standard preceding a deviant used as the comparison against which the relative negativity of the deviant (MMN) is assessed; Haenschel et al., 2005), but the elicited MMNs were found to reflect simply acoustic differences between speech and nonspeech (noise) portions of the stimuli. Nevertheless, Hannemann's (2007) finding, based on induced GBA, does support the general availability of top-down information in the perception of acoustically degraded speech.

Integration of top-down information in auditory perception has also been shown in sentence processing, where words of low cloze probability show significant negative-going event-related potentials (ERPs) relative to high cloze-probability words in the 350-550 ms window consistent with the N400 component, but such effects disappear when the signal is distorted (Boulenger et al., 2011). However, no interaction between cloze probability and ERP amplitude in the earlier window consistent with mismatch negativity was found. Further, these results are complicated by a range of syntactic, semantic, and lexical prediction effects on brain responses during sentence processing, many of which remain the subject of considerable debate (Van Petten and Luka, 2012; Yan et al., 2017; Nieuwland, 2019; Nieuwland et al., in press), which makes attribution of a given ERP component to a single process of top-down resolution of acoustic ambiguity less certain.

Thus, given the ongoing debate about the time-course of the integration of top-down information in speech processing (e.g., Mattys et al., 2012; Sohoglu et al., 2012), as well as the role of higher-order structures in MMN responses (e.g., Pulvermüller et al., 2001), this gap regarding the availability of higher-order information in early auditory perception deserves further attention. The aims of the present study are two-fold. Theoretically, it informs one of the fundamental debates in psycholinguistics; i.e., the nature of top-down and bottom-up mechanisms in speech perception. Methodologically, we will further test the range of factors that MMN is sensitive to and how they impact MMN effects.

1.2. Current Study. The present study seeks to address these related questions by presenting listeners with a sequence of tokens of a minimal pair which randomly alternate between noise-spliced and acoustically intact stimuli. To this end, the minimal pair 'football'–'footfall' was chosen, where 'football' was substantially more frequent than 'footfall' (10.5 as compared with 4.8 in COCA log-frequency; Davies, 2008). In our noise-spliced stimuli, the critical segment *b/f* was replaced by white noise (thus the term, *noise-spliced stimuli*). The noise-spliced stimuli were interspersed among acoustically intact tokens of "football", thus forming a sequence of standards if the noise-spliced stimuli were restored as "football". We hypothesize that if top-down processing influences early auditory processing, the noise-spliced stimuli are more likely to be restored as 'football'. This way, when a deviant stimulus, an acoustically intact token of "footfall" occurs, an MMN should be elicited. In other words, if phoneme restoration occurs early in the recognition process, there will be a many-to-one ratio between standard "football" and deviant "footfall", and an MMN should emerge. In contrast, if phoneme restoration occurs later in the recognition process or does not happen pre-attentively and automatically (as MMN is assumed to reflect), then no valid many-to-one ratio exists, and thus, no MMN should emerge.

In addition to the word condition, a nonword condition ('dutbav' vs. 'dutfav'; constructed to mirror the syllabic, prosodic, and critical segment structure of the 'football'–'footfall' word pair) was included to verify that any effects seen in the word condition are indeed lexical. Note that the standard, 'dutbav', by definition, will occur more frequently in the experiment and thereby could bias resolution of the intervening noise even in the absence of lexical information. This condition allowed us to examine whether the MMN elicited in the current study is due to reconstruction of deleted segments via lexical access, which should occur only in the word condition, or due to ad-hoc grouping of noise-spliced stimuli with the more common of the two unaltered stimuli, a mechanism that would apply equally to words and nonwords. Thus, if an MMN emerges in our word condition and is absent in our nonword condition, we can attribute our results to the involvement of top-down lexical information.

Finally, control conditions were added alongside both word and nonword conditions, where several word/nonword stimuli were presented along with the deviant (*footfall* in the word condition, *dutfav* in the nonword condition) such that no many-to-one ratio could form prior to the listener hearing the deviant. This condition allowed for direct comparison between deviants with and without a standard precursor (see Section 2.2 for further details).

We formulate three predictions which are used to guide the interpretation of the preliminary results in the Discussion. First, if phoneme restoration is not only *sub-lexical*, but is *pre-attentive* as well, suggesting that it is not an artifact of post-lexical processing but rather originates from top-down influences, then listeners should show an MMN to word deviants, but not to non-word deviants. Note that in our design, both word and nonword conditions have no stimuli which occur in a many-to-one ratio at a pure acoustic level, thanks to the random interspersion of noise-spliced stimuli among acoustically intact stimuli, ruling out a purely acoustic origin for any MMN effects observed.

A second possibility is that participants show mismatch responses to deviants in both word and nonword conditions, not as a result of phoneme restoration, but rather to the ability of the MMN to reflect ad-hoc grouping (Paavilainen, 2013). Given that nonword standards appear more often than nonword deviants, listeners may develop a bias toward interpreting the noise-spliced stimulus as an instance of the nonword standard and thus may develop a many-to-one ratio between the nonword standards and the nonword deviant as a result of this grouping, even though its source is non-lexical. Notably, under this prediction, we should also expect an MMN in the word condition, as we do not expect lexical information to inhibit ad-hoc grouping. In other words, it is possible that MMN may emerge in both the word and the nonword conditions, because of two different underlying mechanisms. Several open questions would remain to be explored under such a scenario. These questions include whether the added lexical information in support of grouping in the word condition should elicit a larger MMN relative to the nonword condition, even in the presence of a nonword MMN (e.g., Bakker, MacGregor, Pulvermüller, & Shtyrov, 2013), and whether there would be a difference between the two MMNs regarding timing, amplitude, or topological distribution (Deouell, 2007). Finally, if lexical information is not accessible at a pre-attentive stage of processing and if listeners are not able to form ad-hoc groups as described for the nonword condition above, then no MMN should be elicited for either the word condition or the nonword condition.

2. Methods

2.1. Participants. Seven native English speakers were recruited from the University of Kansas student population to participate in the study, and were given course credit or received payment as compensation for their time. All participants reported no known speech or hearing impairments, and 6 out of 7 were right-handed.

2.2. Materials and procedure. Two sets of stimuli were developed for this experiment: a real word set and a nonword set. Both consisted entirely of two-syllable sequences of similar segmental and prosodic composition. In the real word set, the standard–deviant minimal pair was 'football' /fʊtbɑl/ and 'footfall' /fʊtfɑl/. Five additional items were included in the word control condition. In the control condition, the deviant stimulus "footfall" was preceded by a mixture of words (*wholesale, cosmos, tantrum, carefree,* and *random*). Hereafter, we will refer to the deviant in the control condition simply as the '*control*'. By comparing the *deviant* to the *control*, we are able to restrict comparisons to the same word—i.e., 'footfall'—while varying only the context in which that item occurs: one containing only standards and thus sufficient to yield a mismatch negativity, and one where such conditions are absent.

For the nonword set, the minimal pair /dʌtbæv/–/dʌtfæv/ was used as the standard–deviant nonword analogue of 'football'–'footfall'. These two-syllable nonwords were chosen so that each was phonotactically legal, neither syllable was a real word, and the syllable structure was prosodically consistent with the real word contrast (including both syllables receiving stress in a manner typical of compound words). Additional nonwords in the control condition were chosen by altering the phonological composition of the initial and final syllables of their word counterparts, while keeping the medial cluster intact. These nonwords were /krolstʌt/, /rɑzmɑf/, /læntrʌp/, /plerfrɛm/, and /mɪndɑp/.

All 14 stimuli were recorded by the first author, a phonetically trained native speaker of Midwestern American English, in multiple repetitions with a fixed cardioid dynamic microphone in an anechoic chamber at the University of Kansas. All sound files were analyzed in Praat (Boersma & Weenink, 2016) and cross-spliced such that no item occurred unaltered, and standard-deviant pairs contained the exact same acoustic information (cross-spliced from another item) except for the critical *b/f* segment. That segment, which was approximately 50 ms in duration, was replaced with 50 ms of uniform white noise of amplitude equal to the mean amplitude of the critical segments in the standard and deviant items. This procedure generally follows that of Samuel (1996). In this manner, depending on whether phoneme restoration is needed, all stimuli can be divided into two groups: the acoustically intact (but edited) items, and the noise-spliced items. Table 1 below summarizes the experimental conditions, where # indicates noise. Note that for additional items in the control condition (both word and nonword), noise was added to one of the segments of the medial cluster, but the position of the noise in the cluster (i.e., whether C_1 or C_2 in a C_1C_2 cluster was replaced) was not held constant, as the goal was to disrupt formation of standards. In this sense the reason for adding noise was simply to keep the general exposure of the participants to disrupted and non-disrupted stimuli consistent across conditions.

		STANDARDS	DEVIANT
WORD CONDITION	INTACT	*football*	*footfall*
	MASKED	*foot#all*	
NONWORD CONDITION	INTACT	*dʌtbæv*	*dʌtfæv*
	MASKED	*dʌt#æv*	
WORD CONTROL CONDITION	INTACT	*football, wholesale, cosmos, tantrum, carefree, random*	*footfall*
	MASKED	*foot#all, who#sale, cos#os, ta#trum, ca#free, ra#dom*	
NONWORD CONTROL CONDITION	INTACT	*dʌtbæv, læntrʌp, krolstʌt, mɪndap, plerfrɛm, razmaf*	*dʌtfæv*
	MASKED	*dʌt#æv, læ#trʌp, kro#stʌt, mɪ#dap, ple#frɛm, raz#af*	

Table 1. Stimuli for each condition. *Intact* indicates that the stimuli are not masked with noise. *Masked* means some portion of the stimulus is masked with white noise.

In total, 2400 stimuli were presented in a pseudo-randomized order in 6 blocks of 400 stimuli (separated by a 500 ms inter-stimulus interval, ISI). All blocks were further divided into two subblocks of either words or nonwords (i.e., there were no all-word or all-nonword blocks). Subblock order was counterbalanced across participants. This design was implemented to ensure that participants received both word and nonword exposure between breaks.

For the word and nonword conditions, standards stood in a 7:1 ratio to deviants and deviants never occurred after fewer than 5 or more than 10 standards. For the word and nonword control conditions, randomization was constrained so that no more than 3 words ever occurred in the same sequence, thus ensuring that deviant items in these conditions were proper controls as no stimuli occurred in a many-to-one ratio with the deviant. Not including electrode cap preparation time, the experiment took approximately 50 minutes to complete.

2.3. EEG Recording. An electroencephalogram (EEG) was recorded in the Neurolinguistics and Language Processing Laboratory at the University of Kansas for each participant while they passively listened to the stimuli and watched a silent movie simultaneously to maintain alertness. EEG signals were acquired from a 70-channel Neuroscan Synamps2 system (Compumedics Neuroscan, Inc.) using a 32-channel Ag/AgCl electrode cap (Electro-Cap International, Inc.) fitted to the subject's head. Bipolar electrodes were placed above and below each eye, and at the outer canthi, to monitor for blinks and eye-movements. Electrode impedances were reduced to below 5 kΩ by applying an electrolyte gel to the subject's scalp. EEG signals were referenced online to the left mastoid, low-pass filtered at 200 Hz, and high-pass filtered at 0.1 Hz. The sampling rate for the recording was set at 1000 Hz.

2.4. Data Processing. Triggers were placed at the onset of the second syllable, prior to the critical sound differentiating standard, deviant, and noise-spliced standard stimuli (i.e., in between the /t/ and /b/ for 'football', /t/ and /f/ for 'footfall', and /t/ and # for 'foot#all'). Continuous EEG data were re-referenced offline to the mean of the left and the right mastoids using Neuroscan Edit (Compumedics Neuroscan, Inc.). Subsequent analyses were carried out using EEGLAB (Delorme & Makeig, 2004) and MATLAB (The MathWorks, Inc.). Bad channels were interpolated; no more than one such channel was found per participant. Continuous data were epoched into -500 ms to 500 ms intervals relative to all the triggers and

de-meaned using the mean of the whole epoch (as recommended by Groppe, Makeig, & Kutas, 2009). Epochs were then decomposed into independent components using Independent Component Analysis with the *runica* function in EEGLAB (Makeig, Bell, Jung, & Sejnowski, 1996). For each participant, 1-4 independent components that are typical of eye movements, blinks, and muscular activity were identified by visual inspection and pruned from the data.

Data were then divided into smaller epochs spanning from 100 ms prior to the onset of the stimulus to 400 ms post-critical sound, separately for the standards and deviants within the word, word control, nonword, and nonword control conditions. Trials were baseline-corrected from a 100 ms interval prior to the onset of the stimulus. For the word conditions, this corresponds to between -386 and -286 ms in the epoch, as the pre-critical sound interval for standard and deviant words (the first syllable) was 286 ms long. For the nonword conditions, the baseline corresponded to between -382 and -282 ms in the epoch, as the pre-critical sound interval was 282 ms long. Data were then passed through a 30Hz low-pass filter and remaining artifacts were automatically deleted for trials with amplitude fluctuations exceeding ±100 µV at any channel. This procedure excluded 9.5% of all trials. The remaining trials were averaged by condition in MATLAB.

2.5. Analysis. Analysis was restricted to frontal electrode Fz for the purposes of this preliminary report; electrode Fz serves as a useful representative electrode, as responses across participants generally showed a fronto-central scalp distribution that is typical of MMN. Mean amplitudes between 150 and 250 ms in each epoch were computed using MATLAB. Finally, given that the size of the present data (seven participants), mean amplitudes were not analyzed statistically, but rather are presented descriptively in Section 3, as data collection is ongoing and statistical analysis will be withheld until the sample reaches sufficient power.

3. Results

3.1. Data averaged across participants. Averaged ERP waveforms for standard and deviant stimuli in the Word and Nonword conditions at electrode Fz are shown in Figure 1 below. Between 150 and 250 ms after the onset of the second syllable (the onset of the critical segment to be restored in the noise-spliced stimuli), the Word Condition shows a mismatch negativity (–5.37 µV for deviant, –3.94 µV for standard; SD = 2.36, 1.83, respectively), whereas the Nonword Condition shows no sizeable difference between the two waveforms (deviant = –4.61 µV, standard = –4.25 µV; SD = 1.87, 2.54, respectively).
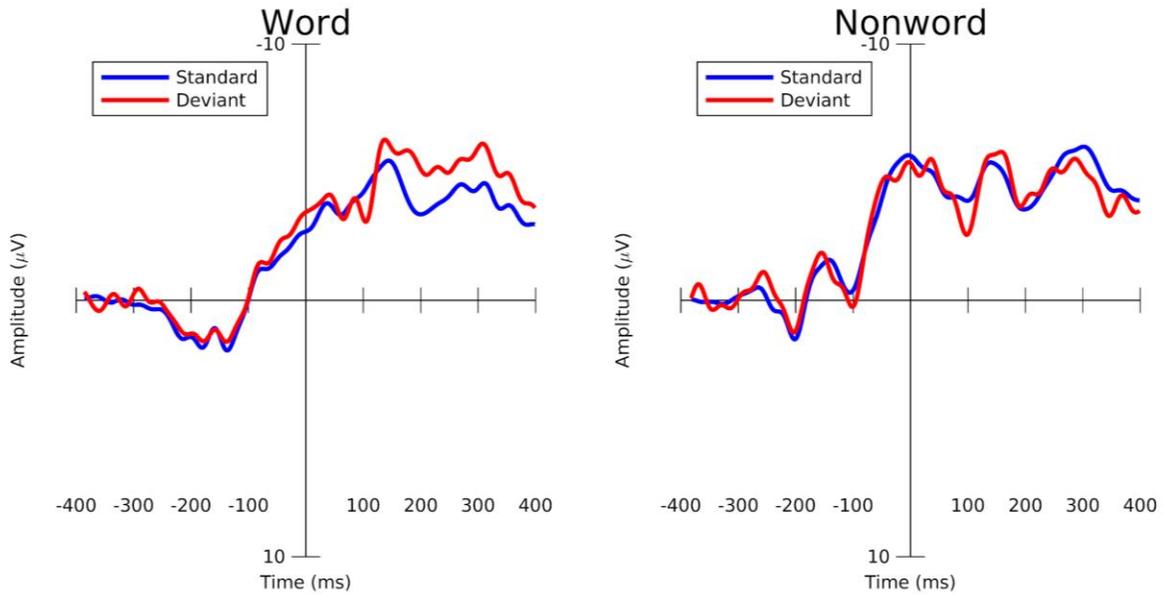
Figure 1. Mean ERP waveforms at electrode FZ for standard and deviant stimuli in Word and Nonword conditions. The *y*-axis (time = 0) is the onset of the critical segment (i.e., /b/, /f/, #).

The pattern in Figure 1, where an MMN is notably absent in the Nonword condition, is in line with our first prediction, which hypothesizes that top-down information (as is available for word but not nonword stimuli) is necessary to allow noise-spliced segments to be restored. Due to this restoration effect, the acoustically distinct noise-spliced and non-noise-spliced, standard tokens can be grouped together, yielding the many-to-one ratio required for MMN elicitation.
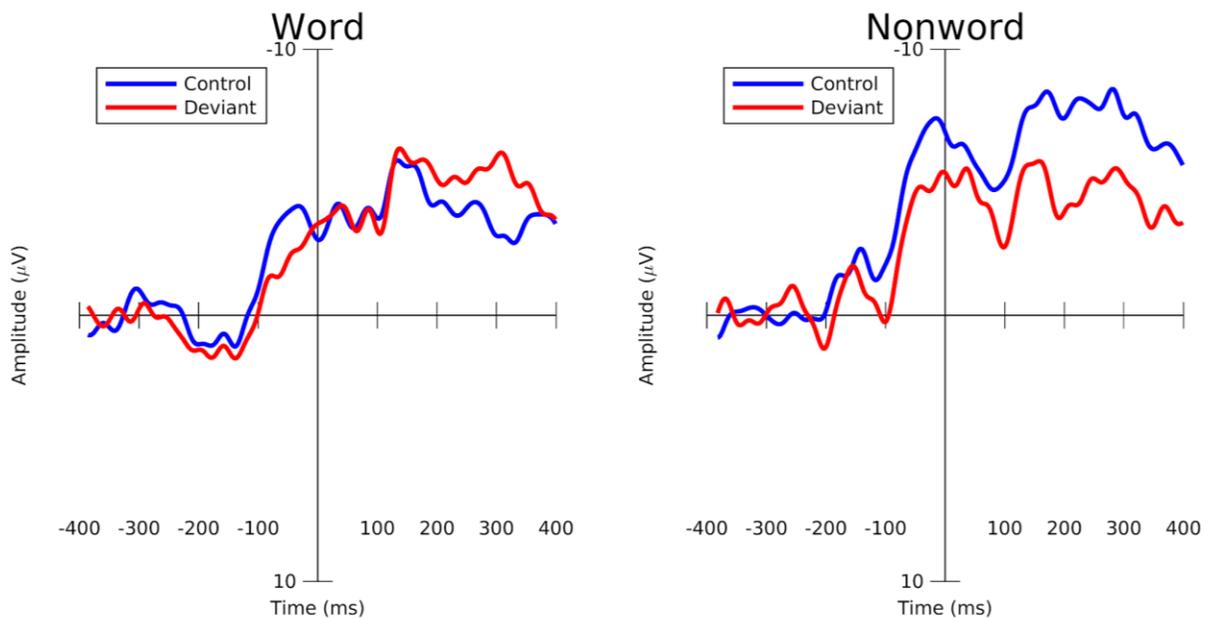

Figure 2. Mean ERP waveforms at electrode FZ for control and deviant stimuli in Word and Nonword conditions. The *y*-axis (time = 0) is the onset of the critical segment.

Figure 2 shows a similar difference in waveforms between deviant and control stimuli in the Word condition (i.e., *footfall* when preceded or not preceded by a standard, respectively), as was observed between deviants and standards in Figure 1. By contrast, the Nonword condition shows a sizeable difference in the opposite direction of an MMN (control more negative than deviant). While the result in the Word condition converges with the standard vs. deviant analysis above and is consistent with the prediction that phoneme restoration influences pre-attentive processing as indexed by MMN, the Nonword result for the deviant vs. control comparison was not an expected pattern. Recall that we had predicted that there should be no robust MMN for nonwords due to phoneme restoration, but that nonwords could possibly show MMN due to ad-hoc grouping of stimuli into standards and deviants; our findings for the nonwords vs. controls conform to neither of these predicted outcomes.

It is unclear what the greater negativity in the controls than the deviants derives from, but it could be an artifact of the stimuli used in the control condition. Compared with the experimental conditions, standards in the two control conditions are acoustically diverse in the interval preceding the critical segment, which could be responsible for the difference in the two waveforms before the onset of the critical region (before 0 ms in Fig. 2). This difference disappears prior to the MMN analysis window in the Word condition, but extends throughout the remainder of the epoch in the Nonword condition. It remains to be seen whether this trend will persist after collecting more data from participants.

3.2. Individual data. Table 2 presents individual participants' mean amplitudes in the analysis window (150–250 ms after the onset of the critical region) at electrode Fz. Negative values are bolded to help illustrate the relative consistency of MMN responses across participants. Of the 7 participants tested, in the deviant vs. standard comparison, 6 show reliable negativities in the Word condition which suggests the presence of MMN. By contrast, only 3/7 participants show such differences in the Nonword condition. This might be due to only some participants forming ad-hoc groups for our nonword stimuli. Examination of a larger sample of participants will help inform how robust this pattern for nonwords is across participants. Similarly, 5/7 participants show evidence of an MMN between deviants and controls in the Word control condition, whereas the majority of participants (5/7) show large differences between deviant and control in the opposite direction (as reflected in Figure 2).

	DEVIANT − STANDARD		DEVIANT − CONTROL	
PARTICIPANT	WORD	NONWORD	WORD	NONWORD
P01	**−1.463**	1.945	1.011	4.906
P02	**−1.178**	0.196	**−2.847**	1.661
P03	**−2.848**	**−1.514**	**−1.301**	0.265
P04	**−1.521**	0.486	**−3.077**	4.177
P05	1.493	2.820	**−1.412**	11.01
P06	**−3.026**	**−4.193**	2.918	2.137
P07	**−1.456**	**−2.304**	**−1.579**	**−0.718**

Table 2. Mean differences (in μV) by participant in ERP amplitudes (deviant − standard, deviant − control) in the 150-250 ms window following second syllable onset. Negative differences, consistent with the MMN response (i.e., deviant more negative-going than standard or control), are bolded.

4. Discussion

This experiment was designed to test whether top-down information, as reflected in the phenomenon of phoneme restoration, is applied online in early, pre-attentive stages of speech processing. The preliminary data here suggest that the MMN component is sensitive to lexical information, as a mismatch response between the deviant and standard was elicited in the Word condition but not in the Nonword condition, where crucially the many-to-one ratio necessary for the MMN was only present if listeners treated the clean and noise-spliced standards as the same items after applying phoneme restoration. Further analysis comparing deviants following standards, vs. a control condition when the deviant stimulus was not in a many-to-one ratio with standards, yielded a similar pattern. Examination of the individual participant data demonstrated that a greater negative amplitude in the deviant stimuli relative to the standards for the Word condition is present for all but one speaker (P05), with the deviant vs. control condition results similarly consistent with five participants showing the effect. For nonwords, in the deviant vs. standard comparison, amplitude differences are more evenly distributed between positive and negative deflections from the standard (3 negative, 4 positive), again consistent with predictions that phoneme restoration (a lexically driven phenomenon) is required to elicit an MMN under the present design. However, the nonword deviant vs. control condition shows overall positive differences inconsistent with any of our predictions, and remains unexplained, though we suspect it is an artifact of the variability in onset syllable across the control items. Ultimately, while our preliminary data suggest the MMN can reflect integration of top-down information, a larger participant sample is required in order to lend further support to this finding.

Our predictions and our interpretation of the present results rely on several critical assumptions about the MMN response, and phoneme restoration more broadly, chief among them being the role of attention. We stated at the outset that the mismatch negativity is largely held to reflect pre-attentive processes, though it may be accompanied by other attention-related components which may affect the timing and spatial distribution of the response (Näätänen et al., 2007). Future modifications of the study could test

for the role of attention in regulating response patterns to word and nonword conditions by randomly assigning half of the participants to a condition where they are explicitly told to pay attention to the stimuli, while the other half will continue with the present design of listening passively while watching a silent movie.

One interesting consequence of the implications of phoneme restoration for speech processing is that such top-down information, which is assumed to be regularly utilized by listeners in resolving the auditory signal, especially in conditions of acoustic degradation, is potentially much more fine-grained than we have tested in the present word-nonword design. Lexical frequency has been shown to be a strong determinant of the likelihood and robustness of phoneme restoration (Samuel, 1996), and indeed frequency was a primary factor considered in the development of our 'football'-'footfall' standard-deviant word pair. Thus, a critical next step in the project would be to test asymmetric (frequent-infrequent) word pairs against more symmetrically distributed pairs where there is less of a bias toward resolving ambiguity in favor of a more frequent word, therefore the relative likelihood of like standards being grouped in a many-to-one ratio should be reduced as well. In this design, we would predict the greater balance in frequency for items in the second pair to result in less-informative top-down information and therefore less robust of an MMN (or no mismatch response at all).

From an information-theoretic standpoint this expectation has a mathematically precise foundation, as it reflects the relationship between the relative likelihood of two potential messages and the conditional entropy of the message source when an ambiguous signal, such as one spliced with noise, is encountered (Shannon, 1948). Under equal frequencies, entropy is at a maximum; therefore, such word pairs should increase uncertainty in the listener relative to pairs where frequencies are unbalanced. Thus, with this project we not only have a means of testing the limits of a generally robust phenomenon in psycholinguistic research – that of phoneme restoration – but we also have a context in which we can test more broadly the manner by which the brain accommodates uncertainty in the speech signal in the process of maintaining efficient communication.

References

Bakker, I., MacGregor, L. J., Pulvermüller, F., & Shtyrov, Y. (2013). Past tense in the brain's time: Neurophysiological evidence for dual-route processing of past-tense verbs. NeuroImage, 71, 187–195.

Beres, A. M. (2017). Time is of the Essence: A Review of Electroencephalography (EEG) and Event-Related Brain Potentials (ERPs) in Language Research. *Applied Psychophysiology and Biofeedback, 42*(4), 247–255.

Boersma, P., & Weenink, D. (2016). Praat: Doing phonetics by computer [Computer software]. http://www.praat.org/.

Boulenger, V., Hoen, M., Jacquier, C., & Meunier, F. (2011). Interplay between acoustic/phonetic and semantic processes during spoken sentence comprehension: An ERP study. *Brain & Language, 116*, 51-63.

Davies, M. (2008). The Corpus of Contemporary American English (COCA): 520 million words, 1990-present.

Dehaene-Lambertz, G., Dupoux, E., & Gout, A. (2000). Electrophysiological correlates of phonological processing: A cross-linguistic study. *Journal of Cognitive Neuroscience, 12*(4), 635–647.

Delorme, A. & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134*(1), 9–21.

Deouell, L. Y. (2007). The Frontal Generator of the Mismatch Negativity Revisited. Journal of Psychophysiology, 21(3–4), 188–203.

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.

Hannemann, R., Obleser, J., & Eulitz, C. (2007). Top-down knowledge supports the retrieval of lexical information from degraded speech. *Brain Research*, *1153*, 134–143.

Hannemann, R. (2008). *Deficient speech repaired: Evidence from event related brain responses*. Doctoral dissertation, Universität Konstanz.

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience, 8*, 213.

Magnuson, J. S., McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2003). Lexical effects on compensation for coarticulation: The ghost of Christmash past. *Cognitive Science*, *27*(2), 285-298.

MathWorks (2015). MATLAB version 8.5 (R2015a). MathWorks, Inc., Natick, Massachusetts.

Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes*, *27*(7–8), 953–978.

McQueen, J. M. (2003). The ghost of Christmas future: didn't scrooge learn to be good?: Commentary on Magnuson, McMurray, Tanenhaus, and Aslin (2003). *Cognitive Science*, *27*(5), 795-799.

McQueen, J. M., Jesse, A., & Norris, D. (2009). No lexical–prelexical feedback during speech perception or: Is it time to stop playing those Christmas tapes?. *Journal of Memory and Language*, *61*(1), 1-18.

Miller, C. A., Abbas, P. J., & Robinson, B. K. (2001). Response Properties of the Refractory Auditory Nerve Fiber. *Journal of the Association for Research in Otolaryngology*, *2*(3), 216–232.

Nieuwland, M. S. (2019). Do 'early' brain responses reveal word form prediction during language comprehension? A critical review. *Neuroscience & Biobehavioral Reviews, 96*, 367-400.

Nieuwland, M., Barr, D., Bartolozzi, F., Busch-Moreno, S., Donaldson, D., Ferguson, H. J., Fu, X., Heyselaar, E., Huettig, F., Husband, M., Ito, A., Kazanina, N., Kogan, V., Kohut, Z., Kulakova, E., Meziere, D., Politzer-Ahles, S., Rousselet, G., Rueschmeyer, S.-A., Segaert, K., Tuomainen, J., & Von Grebmer Zu Wolfsthurn, S. (in press). Dissociable effects of prediction and integration during language comprehension: Evidence from a large-scale study using brain potentials. *Proceedings of the Royal Society B: Biological Sciences.*

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences*, *23*(3), 299-325.

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.

Paavilainen, P. (2013). The mismatch-negativity (MMN) component of the auditory event-related potential to violations of abstract regularities: A review. International Journal of Psychophysiology, 88(2), 109–123.

Pettigrew, C., Murdoch, B., Chenery, H., & Kei, J. (2004). Review The relationship between the mismatch negativity (MMN) and psycholinguistic models of spoken word processing. *Aphasiology*, *18*(1), 3–28.

Phillips, C., Pellathy, T., Marantz, A., Yellin, E., Wexler, K., Poeppel, D., McGinnis, M., & Roberts, T. (2000). Auditory cortex accesses phonological categories: An MEG mismatch study. *Journal of Cognitive Neuroscience, 12*(6), 1038–1055.

Pulvermüller, F., Kujala, T., Shtyrov, Y., Simola, J., Tiitinen, H., Alku, P., Alho, K., Martinkauppi, S., Ilmoniemi, R. J., & Näätänen, R. (2001). Memory traces for words as revealed by the mismatch negativity. *Neuroimage, 14*(3), 607–616.

Sams, M., Paavilainen, P., Alho, K., & Näätänen, R. (1985). Auditory frequency discrimination and event-related potentials. *Electroencephalography and Clinical Neurophysiology: Evoked Potentials Section, 62*(6),

437–448.

Samuel, A. G. (1996). Phoneme restoration. *Language and Cognitive Processes, 11*(6), 647–654.

Samuel, A. G. (2001). Knowing a word affects the fundamental perception of the sounds within it. *Psychological Science, 12*(4), 348–351.

Samuel, A. G., & Pitt, M. A. (2003). Lexical activation (and other factors) can mediate compensation for coarticulation. *Journal of Memory and Language*, *48*(2), 416-434.

Shannon, C. (1948). A mathematical theory of communication. *Bell System Technical Journal, 27*, 379–423.

Shtyrov, Y., & Pulvermüller, F. (2007). Language in the Mismatch Negativity Design. Journal of Psychophysiology, 21(3–4), 176–187.

Sohoglu, E., Peelle, J. E., Carlyon, R. P., & Davis, M. H. (2012). Predictive Top-Down Integration of Prior Knowledge during Speech Perception. *Journal of Neuroscience*, *32*(25), 8443–8453.

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology, 83*(2), 176-190.

Yan, S., Kuperberg, G. R., & Jaeger, T. F. (2017). Prediction (or not) during language processing. A commentary on Nieuwland et al.(2017) and DeLong et al.(2005). *bioRxiv*, 143750.