

# Using MMF to situate perceptual epenthesis in the speech processing pathway: a proposal

sb swae  
[sbss@umich.edu](mailto:sbss@umich.edu)  
University of Michigan  
Department of Linguistics

This proposed study uses an adapted oddball paradigm to explore at what processing stage perceptual epenthesis occurs. I will compare two conditions: the perceptual epenthesis condition (e.g., [lbif] [l!bif] textscstd-dev pair) and a control condition (e.g., [sla] [s!la] textscstd-dev pair). With magnetoencephalography, I will compare the mismatch field response to each token. If the perceptual epenthesis condition shows a smaller MMF response than the controls, I would interpret that as evidence that [lbif] and [l!bif] are phonologically equivalent at that stage (early processing, <300ms). A result of this nature would reinforce psycholinguistic findings and allow the field to adopt a strong hypothesis about when perceptual epenthesis is carried out on a representation.

**Keywords:** MEG, MMF, perceptual epenthesis, speech processing

## Introduction and Background

Two accounts exist that locate perceptual epenthesis in different parts of the speech processing pathway. The passive account posits that perceptual epenthesis is a phonetic failure that occurs early (less than 300ms) after stimulus onset. The active, or strong, hypothesis claims that perceptual epenthesis is carried out at a later stage. In this paper I propose a study to investigate the nature of perceptual epenthesis using a mismatch field paradigm with magnetoencephalography (MEG). The goal is to situate perceptual epenthesis within the current schematic of early speech processing. Specifically, the design proposed here utilizes an oddball effect-elicitation paradigm to collect data that may allow the field to decide between the two possible locations for perceptual epenthesis.

Studies of the mismatch field (MMF) have indicated that it is sensitive to boundaries between phonemes. This makes it an adequate tool to investigate how segments are represented in normal human brains and thus when in the speech-processing pathway perceptual epenthesis occurs

### Perceptual epenthesis: phonetic failure or active repair?

Perceptual epenthesis (PE) is the phenomenon in which a listener perceives or inserts a vowel between two consonants that form a prohibited cluster in their language. PE has been observed in numerous languages (Dupoux et al. (1999); Kabak and Idasrdi (2003)), but the inspiration for this endeavor presented itself in response to English data in Berent (2013). Furthermore, the possible confound of contamination from other languages' phonological systems prescribes

that PE be studied in monolingual subjects.

In three experiments, Coetzee (2010) observed apparent PE in English. The first experiment involved a "same or different?" matching task. Monolingual speakers of English were auditorily presented with pairs of nonce words and asked to indicate whether they thought that the pair was distinct or identical. As an example, subjects would hear first a [sthápi] and [s!thápi] and then use a response box to indicate "same" or "different." In these instances, where one word conforms to English phonotactics and the other is not permitted, subjects emphatically judged the two tokens as identical, suggesting that, because English phonology only permits stop aspiration in syllable-initial positions, the subjects epenthized a vowel between the poorly-formed onset cluster (i.e., \*CC ! CVC), allowing parsing to proceed.

A syllable-counting task comprised the second experiment. On items that complied with English phonology, subjects performed well on syllable counts. In tokens like [skhan] that did not have a valid English parse, subjects gave fewer correct answers, took longer to respond, and reported an extra syllable. This final behavioral change appears to be explained by PE: subjects epenthized a vowel in the \*CC clusters, which accounts for the extra syllable.

In the third experiment, the same materials were used, but subjects were now instructed to choose an orthographic rendering that corresponded with what they heard. To illustrate, when subjects heard [skhan], they were asked to choose between <sk...> or <sek...>. Subjects strongly chose the correct <sk> option, contrary to the PE suggestions of the previous two experiments. These instead suggest that subjects were creating a faithful representation of the acoustic stimuli

in early audition/speech processing, which accounts for the correct spelling choices. Coetzee (2010) interpreted this as evidence for the existence of multiple processing stages: an early stage integrating acoustic information into a representation which is then passed on to the phonological grammar used to parse and make judgments about the percept's well-formedness. The claim for PE is that somewhere along this computation, the brain coerces an acceptable form out of the stimulus.

By claiming that there exist 2 (or more) distinct processing stages necessarily presents the problem of situating PE within this mechanism. Coetzee (2010) provides a schematic of the mechanism that is reproduced in Figure 1 below. The figure also includes a complementary diagram from Berent (2013) that corresponds very neatly to the schematic and points out the two possible locations where PE can enter the process. She then adopts the strong or active hypothesis that situates PE in the latter section.

### Mismatch field: the brain's "change detector"

The mismatch field (MMF) is a task-independent, automatic response to significant changes in a stimulus. (The meaning of significant changes will be treated in following paragraphs.) In response to auditory linguistic stimuli, the MMF is left-lateralized and its generator is localized to the auditory cortex. The MMF exhibits a latency of 150-250ms after the onset of a significant change in the stimulus

Habituated to a consistent or repetitive auditory stimulus, a listener forms a faithful representation of the standard token. When a deviant (textscdev) is introduced among the many standards (textscstd), the MMF presents. The MMF is a response to the DEV, but there is a rapid sensitization effect; if the DEV is played more frequently, MMF amplitude decreases for subsequent tokens. Moreover, a high textscdev:std ratio also results in decreased MMF amplitude overall (Näätänen, 2007)

The MMF has been used to explore access of phonological categories in neural structures. Phillips, et al. (2000) presented subjects with a "train" of dental stops that were digitally manipulated to vary in voice onset time (VOT); the textscstd was voiced ([da]), the textscdev was voiceless ([ta]). In a general form, the train might be represented as:

STD STD STD STD STD STD DEV STD STD STD

MEG collected data time-locked to the train. Auditory MMF was found to be sensitive to phonological—not acoustic—differences: to control for phonetic cues their subjects may have been using online during the task, Phillips et al. manipulated the VOT of the tokens. Only when the VOT values crossed a "phonemic boundary" was an MMF observed. In subsequent blocks the relative VOT differences were manipulated equally (an equally applied

X-millisecond shift). The phonemic boundary appeared to shift by the same amount, because no additional MMF was observed. The results suggest that the auditory cortex is exquisitely sensitive to significant phonemic features in speech stimuli, but this sensitivity is not reducible to simple acoustic variation

These features of the MMF mean that it is well-suited to be used to test against Berent's hypothesis of active phonological repair in a later stage of speech processing. The remainder of this proposal rests on this crucial understanding of the sensitivity of the MMF.

To summarize up to this point, we have assumed that perceptual epenthesis exists because it allows us to account for behavioral data like those found in Coetzee (2010). By using the MMF to explore access of phonological categories, it is possible to explore when PE occurs. In the descriptions of the experiments that follow, I hypothesize that the absence of an MMF can be interpreted as evidence for Berent's adoption of the strong hypothesis of later, active repair. Thus, the question that has been planted is: Will the MMF will be generated in response to stimuli that are constructed to be both phonetically and phonologically different, or do speakers treat them as phonologically identical? If the MMF presents, it indicates that phonological categories or representations are being accessed early on; if it absent, that function must be performed at a later stage in the stream/computation.

### Methods

The purpose of this experimental design is to construct a paradigm yielding data that allow us to explore the neural correlates of PE. This should allow us answer the question just introduced, and, by extension, will help determine if the strong hypothesis obtains.

I propose a series of basic science experiments, the order of which is still undetermined. First is a necessary normalization task to determine if the stimuli are available to PE. That is, if we reproduce the effect found in Coetzee's first experiment, we can proceed under the assumption that some vocalic epenthetic process permits a parse and then a match; subjects must perceive our well-formed STDs and the illicit DEVS as identical. The second experiment will inspect whether or not MMF generation is correlated with instances of PE. A potential third experiment is the most auxiliary of the bunch; it aims to replicate the spelling effect found in Coetzee's third procedure. This has the dual purpose of replicating his effect and will serve to bolster the claim of distinct stages of speech processing.

### Subjects

At least 30 (50% male) healthy, right-handed undergraduates will be recruited to participate as subjects in the series of experiments. All participants must be monolingual English speakers; the motivation for this is to avoid any confounds in



the form of interference from other languages (Menning et al. (2002)). (This may present some difficulty, owing to the University of Michigan's four-term language requirement for undergraduates, though there are some academic units that exempt their students. That subjects be monolingual is indispensable to the design of this investigation, however: English PE is instantiated as a response to violations of English phonology—L2 may alter this reactivity.)

## Experiment 1

### Hypothesis

When presented with STD CVC and DEV \*CC onset clusters in a “same or different?” matching task, subjects will report perceiving them as the same item.

### Materials

We will record a native speaker of standard newscaster English (Northern Cities Shift, NCS) reading pairs of nonce words. Control pairs are defined as legal CC and CVC onsets. Testing pairs are defined as legal textscstd CVC and prohibited DEV \*CC onset clusters, where the only difference between the two tokens is the textscstds V—which allows for a legal phonotactic parse. These tokens will then be manipulated in Praat (Boersma & Weenik, 2013) to vary the V occurring in the STD, resulting in numerous “flavors” of a single base pair. This exercise serves the purpose of determining whether the product of subjects' PE is exclusively schwa or if other vowels can be epenthesized. This manipulation has value because it allows us to deduce the product of a subject's PE: if a subject matches [lbif] and [!lbif] but reports that [lbif] and [lbif] are different, then we can be reasonably certain that the subject only perceives schwa as the result of PE (Blue and Sherwood, 2013). The STD tokens will be disyllabic, with the stress falling on the second syllable. To avoid the low sonority peaks (like [+stop][+stop] clusters), Blue and Sherwood indicate that their tokens will be exclusively composed of [+fricative][C] clusters. However, since Berent's object of inquiry that set us off on this path was /lbif/, I would also include laterals in the recordings. A few examples of potential stimuli follow below in Table 1.

Table 1  
*Examples of potential stimuli.*

AAA	BBB	CCC
1.0	2.0	3.0
1.0	2.0	3.0

### Task

While this experiment does not use MEG, subjects will still be tested in the same location as the experiment that

does contain the MEG for the sake of consistency. Subjects will wear noise-cancelling headphones. A Praat script will present STD-DEV pairs randomly within and without each pair. A 500ms period of silence will occur between the first and second member of a pair. At the end of each pair's presentation, a screen will prompt subjects to use a response box to make a same/different judgment. Response times (RTs) will be logged from the end of each pair.

### Expected Results

It is expected that subjects will tend to collapse the STD-DEV pairs into a single exemplar by reporting “same” for these pairs. It is not expected that the pairs containing a STD token with a vowel other than a schwa will see the same response, but these pairs are included to confirm whether or not schwa is the universal product of PE.

There exists the possibility that subjects will report “same” for non-schwa STD-DEV pairs. Variation between subjects is also a possibility. Should either of eventualities arise, we cannot claim that subjects exclusively epenthesize schwa and must consider two things. First, the potential need to redesign the stimuli for the second MEG trial. Second, and most worrying, would be the need to reconsider the very nature of PE altogether.

On the outside chance that subject's correctly judge STD-DEV pairs to be different, the RTs for these cases are expected to be significantly longer than the well-formed control pairs. This follows from Coetzee's (2010) results which posits that there is a stage of processing in which PE occurs—this processing will be taxed by time.

## Experiment 2

### Hypothesis

When subjects are presented with STD-DEV pairs, no MMF response should be generated or the MMF amplitude should be significantly less than if subjects were presented with the well-formed control pairs.

### Subjects

The same subjects will be used for this experiment. However, if subjects from the first experiment do appear to be PE'ing, their results may need to be excluded. By necessity, there will be a time gap (days, weeks) between the first and second experiment to avoid priming, allow for analysis, and potentially redesign stimuli.

### Materials

One pair each of the well-formed controls and STD-DEV pairs (excluding the manipulated-vowel flavors) from the previous experiment will be used in this trial. Using a Praat (Boersma and Weenik, 2013) script, the pairs



will be presented in random (within and without pairs) sequences in a 9:1 STD:DEV ratio, with stacking of DEV pairs being prohibited. An interval of 800-1000ms, randomly generated by the script, will occur between each word. A potential train, lasting at most 20 seconds, may look like this:

STD STD STD STD STD STD **DEV** STD STD STD

Each sequence of trains will contain 450 STD tokens and 50 DEV tokens. Each sequence will be repeated with the STD:DEV ratio inverted.

### Task

While the MEG collects time locked data (the locus of data collection being the DEV token of each train) to record whether or not an MMF is generated, subjects will be instructed to listen to the train sequences through noise-cancelling headphones. Following from the claims made above, the MEG data will indicate to us whether or not the DEV crosses a phonological boundary.

### Expected Results

There are several possible outcomes to this particular experiment, though they are not all equally probable.

**Expected Result 1: hypothesis confirmed.** The MMF amplitude is larger for the control pairs than the test pairs. This set of results would indicate that there is no phonemic difference between a PE schwa and an acoustic schwa at the level of representation that the MMF coincides with. If PE occurs at a later processing stage (cf. Figure 1) than the creation of the initial acoustic representation, then this possible outcome indicates (as expected) that the MMF is a correlate of phonemic, not acoustic, representations. (The presence of an MMF indicates that the subject is performing some algebraic phonological operation on the stimulus, and implies

that PE/repair occur early on.)

**Expected Result 2: back to the drawing board.** Observed MMFs are equally strong for both control and test pairs. This possible outcome has two interpretations:

1. There exists a phonemic difference between PE schwa and acoustic schwa.
2. Crossing of phonemic boundaries is not responsible for MMF generation.
3. “experiment fails!” The MMF is stronger/larger for the test condition than the control condition. This outcome is highly improbable and indicates a poor experimental design or an error in execution of the trial.
4. “things are not always what they seem” The MMF occurs with differing latencies for each condition. For example, consider the situation if the control condition results in an MMF at 180ms after stimulus onset and the test condition results in a generated MMF at 240ms. This may indicate that PE occurs at a different processing stage than that which detects the difference between [CC] and [C!C].

### Discussion

This proposed investigation should begin to establish the neurological basis of PE. Whether the test condition results in generation of an MMF will shed light onto where in the speech processing computation PE occurs. This paradigm, if the hypothesis is confirmed, should also serve to provide justification for Berent’s (2013) strong hypothesis that repair occurs later in speech processing, at the phonological grammar that conducts parsing

bib