Neurobiological evidence for voicing underspecification in English

Arild Hestvik[1] and Karthik Durvasula[2]

[1]Department of Linguistics and Cognitive Science, University of Delaware

[2]Department of Linguistics and Languages, Michigan State University

Abstract

In long-term memory, the phoneme units that make up words are coded for the distinctive features and feature values that are necessary to distinguish between words in the mental lexicon. Underspecification theory says that the phonemes that have unmarked feature values are even more abstract in that the feature is omitted from the representation altogether. This makes phoneme representations in words more sparse than the fully specified phonetic representations of the same words. Eulitz and Lahiri (2004) demonstrated that this theory predicts certain asymmetries in the Mismatch Negativity (MMN) response to phoneme contrasts. We expand on this research by demonstrating underspecification-driven asymmetry in the brain response to laryngeal feature contrasts in English (i.e. what makes /d/ and /t/ different). We add a new test by showing that the asymmetry disappears if the MMN paradigm is modified to encourage the formation of phonetic memory traces instead of phonemic memory traces. This result adds further neurobiological evidence that long-term phonological representations are more sparsely represented than phonetic representations.

1. INTRODUCTION

1.1 Background: Phonological underspecification

Long-term phonological representations are more abstract than acoustic/articulatory (phonetic) representations. However, theories disagree about how abstract long-term phonological representations can be. According to classical Generative Phonology [1], long-term memory representations of words consist of phonemes, which in turn are bundles of distinctive features, the minimal subset of phonetic features required for distinguishing between the words in the lexicon [1], [2]. The full phonetic information required to produce or perceive a word is the result of predictable feature changes and feature add-ons (allophonic rules) in the mapping from long-term storage to pronunciation. From this view, a major dichotomy between long-term memory (phonemic) and phonetic representations (temporary bounded events in real time) arises: whereas phonemic word representations are sparse and minimal and only contain unpredictable and non-redundant information, phonetic representations are rich and redundant and contain predictable information, including information not required for making distinctions between words (e.g. dialect information, speaker identification, etc).[1]

The theory of Phonological Underspecification takes this one step further, by incorporating an additional abstraction into the phonemic representations, namely the idea that one value of a binary feature may be the “default”, with the other value being “marked” [3]–[8]. Default rules will then fill in the missing (and predictable) values to produce a full phonetic representation. (This view of phonological representations is a notational equivalent of a view incorporating unary features [9], [10].) While this hypothesis has been criticized [11]–[13], arguments have been put forward that it provides a better analysis of phonological patterns in

---

[1] Note that the argument based on discrimination does not entail that meaning-redundant phonetic detail may not also be present separately in long-term memory representations.

many languages [3], [6], [14]–[16]. Furthermore, this view of abstract phonological representations not only allows for an efficient coding of long-term representations [1], but it also allows for efficient recognition of incoming acoustic signal in the face of substantial acoustic variability [17], [18].

1.2 Cognitive neuroscience evidence for underspecification

Whereas arguments for underspecification in linguistics have been based on the analysis of phonological patterns, evidence has recently also come from techniques that access the brain's representations of speech sounds. Eulitz and Lahiri [17] used the Mismatch Negativity (MMN) paradigm to provide neurobiological evidence for underspecified phonemic representations[2]. The MMN response is a change detection response from the brain, measured as part of the auditory event related potential (AEP) [19]. The MMN response has been shown to be pre-attentive [20], and does not require conscious awareness of the stimulus (e.g. it has been observed during sleep [21], as well as in comatose patients, as an indicator of awakening from coma [22]), and has been shown to be sensitive to language-specific knowledge in individuals [23].

The MMN experimental paradigm works as follows: sequences or trains of "standard" sounds are presented to subjects. Each sound presentation elicits an Auditory Evoked Potential (AEP), which is a temporal sequence of negative and positive waveforms derived from averaging EEG signals. Of most relevance is the N1-P2 complex (N1=the first negative dip in the waveform, and P2=the first positive shift after N1), typically occurring within 200ms after stimulus onset. After a series of standards, the stimulus train is interrupted by a different stimulus

---

[2] For another paradigm that can be interpreted as evidence for underspecification, though not discussed in such terms, we refer the reader to Magnetoencephalography (MEG) work done by Flagg and colleagues looking at the latency of M50 responses to cross-spliced nasal and oral vowels with following nasal or oral consonants in American English [76].

(the “oddball” or “deviant” stimulus). The oddball leads to an attenuation of the N1, and/or P2, as well as later parts of the waveform. The subtraction of the oddball wave from the standard wave results in a difference waveform, which describes the brain’s response to the stimulus change. This is the Mismatch Negativity (MMN) waveform [24]–[26].

The MMN reflects discrimination between any auditory stimuli at the sensory level, and it has been extensively used to demonstrate discrimination between speech sounds, specifically phonetic distinctions within a language [26]–[28]. It has also been used to show that subjects are sensitive to more abstract phonemic contrasts [23], [29]–[33]. In an experiment that we partially replicate in the current study, Phillips et al. [34] presented subjects with multiple tokens of [d] and [t] in an MMN paradigm, by randomly varying the Voice Onset Time (VOT) of the within-category exemplars. By varying the standards within category, a single memory trace representing the standard sequence can be constructed by recruiting the subject’s knowledge of the phonemic representation of the standard. Indeed, Phillips et al. observed a mismatch response to a deviant [d] stimulus presented after a series of varying standard [t] stimuli, even though the VOT distance between [t] and some [d] was the same as that between several different [t] stimuli. Phillips et al. also tested a second “acoustic” condition, where the VOT values for all the stimuli were increased by 20 msec, so that half the stimuli for the lower VOT group were below the perceptual boundary, and the other half were above the boundary (as were all the stimuli from the higher VOT group). Thus, although the proportion of frequently occurring stimuli (standards) and rare occurring stimuli (deviants) as a function of VOT range was the same as in the “phonological” condition, the frequent/rare distinction did not coincide with phonemic categories. In this case, no mismatch was observed, showing that a distribution of rare/frequent VOT values had to coincide with a phoneme category to produce an MMN. This provided evidence that by varying the standards within category, the generator of the MMN is encouraged to recruit phoneme representations of the memory trace of the standards.

Eulitz and Lahiri [17] employed this “varying standards” MMN paradigm to test whether phonemes are underspecified. In accordance with their theory of phonemic and phonetic representations, front vowels are specified for the feature [CORONAL] at the phonetic level, but are not specified for the feature at the phonemic level, because [CORONAL] is unmarked in the grammar. However, back vowels are specified for the feature [DORSAL] both phonetically and phonemically. Finally, while rounded vowels are both phonetically and phonemically specified for [LABIAL], unrounded vowels are specified neither phonemically or phonetically for any rounding features. Therefore, in a language like German, the three vowels [e], [o], and [ø] would have differing phonetic and phonemic representations. Phonemically, the three vowels would be specified as follows: [e] – [$_{\text{fully underspecified}}$]; [o] – [LABIAL]; [ø] - [LABIAL, DORSAL]. And phonetically, the three vowels would be specified as follow: [e] – [CORONAL]; [o] – [LABIAL, CORONAL]; [ø] - [LABIAL, DORSAL]. As can be observed, the proposed phonetic and phonemic representations are asymmetrically specified for different features.

The set of representational asymmetries above, Eulitz and Lahiri claim, leads to MMN asymmetries in the “varying standards” MMN paradigm. For example, if the standard is the vowel [o], and the deviant is [ø], there is a feature incompatibility between the standard *phonemic* memory trace [DORSAL, LABIAL], and the deviant *phonetic* representation [CORONAL, LABIAL]. This feature incompatibility leads to a mismatch. However, if the standard is the vowel [ø], and the deviant is [o], there is no feature incompatibility between the standard *phonemic* memory trace [LABIAL] and the deviant *phonetic* representation [DORSAL, LABIAL]. Therefore, this latter situation does not lead to a mismatch. The logic of the argument and the results were substantiated for German vowels [17], [35], [36], and the method has also been applied to the study of contrasts involving the place of articulation features distinguishing coronals from labials [37], [38].

1.3 The current study

While previous underspecification MMN studies showed that underspecification predicted observed MMN asymmetries, they did not control for the possibility that the asymmetries could be due to intrinsic phonetic differences between the categories to be compared. To preview our findings, we applied the Eulitz/Lahiri logic to test a predicted underspecification asymmetry in the voicing contrast for stop consonants in English, and we show that the asymmetry obtains only when the experimental conditions force the MMN generator to access phonological representations. Specifically, when the conditions forcing a phonological memory trace are removed from the experiment, by encouraging the MMN generator to access phonetic representations, the asymmetry disappears, and we observe a mismatch for both phonemes. This provides additional evidence that the original asymmetry is not due to confounding intrinsic phonetic differences between /d/ and /t/, but rather due to an asymmetry in abstract phonological feature contrasts.

As a basis for these predictions, we utilize a specific phonological theory that says that English voiceless stops are phonologically specified for the feature [+spread], representing the articulatory target of a spread glottis, while the voiced stops are underspecified for any voicing or laryngeal features [39]–[42] (but see Hwang et al. [43] for an opposing view). According to this theory, /d,g,b/ are underspecified, whereas /t, k, p/ are fully specified for the feature [spread] in the lexicon.

In order to test this underspecification hypothesis using the Eulitz/Lahiri experimental logic, we implemented an exact replication of Philips et al.'s [34][3] Experiment 1 , which tests the effect of presenting a stream of [t] sounds with [d] as the deviant, as well as the reverse condition with a stream of [d] sounds with [t] as the deviant. As in Phillips et al. [34], we ensured the formation of a phoneme-based memory representation of the standard stimuli by varying the VOT values of the stimuli within each category's boundaries.

As schematically presented in Figure 1, according to the logic of Eulitz and Lahiri [17], when the standard consists of varying stimuli from the /t/ category (e.g., varying in VOT in 5ms increments between 50ms-65ms), the only possible single memory trace for the standards would be the phonemic representation of /t/. In other words, a copy of the long-term memory representation of the phoneme is moved into auditory working memory. Furthermore, because /t/ is phonemically specified for the feature [+spread], the memory representation will contain the feature specification [+spread]. Subsequently, when a deviant from the /d/ category is presented (e.g., a stimulus with a VOT of 15ms, which we established behaviorally was below the threshold between /d/ and /t/ of about 40ms VOT, see below), the token stimulus [d] is phonetically represented as [-spread] and is compared to the memory representation of the phoneme /t/ specified as [+spread][4]. The direct contradiction of feature values (solid line in Figure 1) is predicted to generate an MMN response.

---

[3] In their study which employed MEG to demonstrate a phoneme-specific MMN generator, Phillips et al. [34] mention in passing that they observed an unexpected asymmetry between /t/ and /d/ in the MMN response; however, they attempt to explain away the asymmetry as a result of general auditory asymmetries involved in the perception of voiced and voiceless stops.

[4] The reviewers ask us how we can know that voiced stops are not also *phonetically* underspecified for the feature [spread]. Indeed, it has been argued that there is underspecification in phonetic representations as well [77], [78]. However, if the voiced stops in English were also phonetically underspecified for the feature [spread], then there would be no specific phonetic target for such sounds with respect to glottal width, i.e., their articulations would vary freely in degrees of spread glottis (and thereby aspiration). This is not observed, as voiced stops in English are invariably pronounced without a spread glottis configuration (or any aspiration). Therefore, voiced stops must be specified with respect to the spread glottis configuration, i.e., phonetically specified as [-spread].

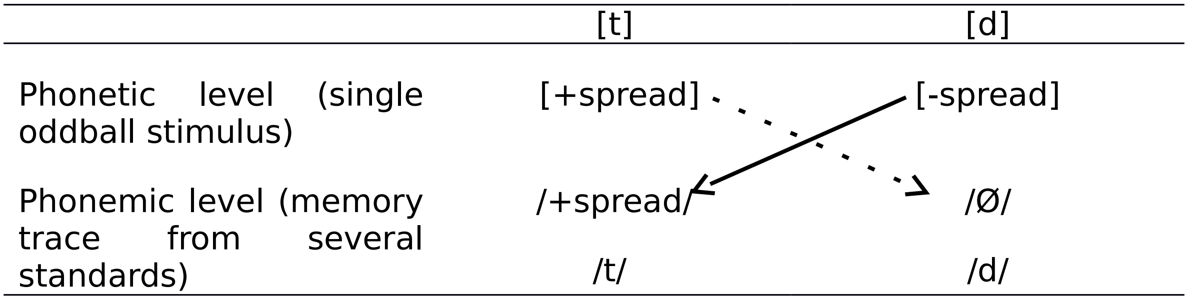


Figure 1: MMNs as per FUL. Ø = unspecified. Arrows indicate which pairs are compared by the perceptual mechanisms. Dotted arrow represents the “no mismatch” comparison.

On the other hand, when the standard consists of varying stimuli from the /d/ category (e.g., varying in VOT between 15ms-30ms), the only possible single memory trace for the standards would be the phonemic representation of /d/. Furthermore, because /d/ is not phonemically specified for any laryngeal feature, the memory representation will not include a specification for voicing or aspiration. Subsequently, when a deviant from the /t/ category (e.g., a stimulus with a VOT of 60ms) is presented, a token stimulus [t] is phonetically represented as [+spread] and is compared to the memory representation of the phoneme /d/ that is not specified for any laryngeal features. Under this condition, the deviant token of [t] does not lead to a direct feature conflict with the phonemic representation of /d/, and no MMN is predicted on the basis of the memory representation. (This logic could be formalized by modeling the memory trace comparison as feature unification, which is basically set union: {[+F]}   {[-F]} = {[+F], [-F]} (a contradiction), but {[+F]}   ={[+F]}.

A standard control in MMN studies is to statistically compare each phoneme to itself as standard vs. deviant, in order to control for intrinsic differences in the AEP for each separate category, in order to not confound the comparison. The data, however, are generated in real time by a deviant from one category being compared to the standard of the other, as illustrated in Figure 1.

Critically, if the standards are not varied, and the same phonetic token is presented as the standard stimulus every time, then we assume that the memory trace is formed using a phonetic category representation. In this case, a fully specified phonetic token in the memory representation of the standards is always compared to a fully specified oddball stimulus.

A reviewer asks how we can know that the memory representation of the standards in Experiment 3 *must* be a phonetic representation. Why cannot it also be a phonemic representation? The answer goes back to the force of the argument in Phillips et al. study [34] which is precisely that you cannot observe the phonemic representation with MMN unless you vary the standards. Therefore, in Experiment 3, where we do not vary the standards, there is nothing in the experiment that is forcing the brain to move towards a more abstract representation of the memory trace.

Furthermore, as both the standard and the deviant are fully specified in Experiment 3, there will always be a direct feature conflict, irrespective of whether [t] or [d] is the oddball, as we predict a fully symmetrical, parallel MMN for both sound categories. This provides a test of the possible “intrinsic phonetic difference” confound: If the asymmetry that was observed in the phonemic condition were in fact due to a phonetic difference between the stimuli, then the same asymmetry should be observed also when a single token is used as the standard.

We tested these predictions in three experiments. Experiment 1 tested the basic prediction of a greater amplitude MMN for /d/ than for /t/, using a task that directed subjects’ attention to the stimulus stream. Experiment 2 repeated Experiment 1 but without the attention task, in order to assess whether the MMN modulation was independent of attention and automatic. Finally, in

Experiment 3 we removed the condition that forced the MMN generator to rely only on phonemic information, by not varying the standards within category. This predicts that there should be no asymmetry, because phonetic information is used for comparing standards to deviants both for /d/ and /t/. f

2. EXPERIMENT 1

2.1 Methods

2.1.1 Subjects

A total of 35 University of Delaware students were recruited as subjects and provided written consent in Experiment 1. None of the subjects reported a history of hearing loss or speech/language impairments, and all reported having English as their first and only language. Each subject was paid $10 for participation.

2.1.2 Stimuli and design

A sequence of “da” and “ta” syllabic utterances was synthetically generated and used in all the experiments. The sequence was created by constructing a VOT continuum going from /d/ to /t/, varying in 5ms increments from 0ms VOT to 100ms VOT. The stimuli were created to exactly match the stimulus specifications used in Phillips et al. [34] in all parametric choices except for the actual number of steps in the VOT continuum, and were synthesized with an online version of the low-level Klatt synthesizer [44], [45]. Each stimulus syllable lasted 290ms. In addition, we synthesized target stimuli for a tracking task consisting of a male- vs. female-sounding voice saying “ba” .

The statistical design of the experiment is given in Table 1. Both phoneme categories were used as standards and deviants, so that the standard for a given phoneme could be compared to itself as a deviant. This avoids introducing confounds due to intrinsic differences in the ERP response to the sounds. By comparing each phoneme as standard to itself as deviant, we control for this possible confound. In addition, we counterbalanced the order of presentation as illustrated in Table 1, so as to be able to analyze post-hoc whether the MMN response differed in the first half vs. the second half of the experiment (it turned out that they did).

		BLOCK ORDER (between-subject):			
		[d]=first deviant		[t]=first deviant	
PHONEME (within-subject):		*/d/*	*/t/*	*/d/*	*/t/*
CONDITION (within-subject):	*Standard*	Standard-D (Block 2)	Standard-T (Block 1)	Standard-D (Block 1)	Standard-T (Block 2)
	Deviant	Deviant-D (Block 1)	Deviant-T (Block 2)	Deviant-D (Block 2)	Deviant-T (Block 1)

Table 1: Statistical design for all experiments

2.1.3 Procedure

Before EEG acquisition, each subject’s discrimination boundary for the continuum was first identified in order to customize the stimuli to the subject’s threshold (following the procedure in Phillips et al. [34]). This was done in a pre-test with a forced choice behavioral identification task using the stimuli described above. Six trials of each of the 10 VOT value were randomly presented and subjects decided whether they thought they heard [d] or [t]. After this, a set of four /d/ and /t/ tokens were selected straddling that subject’s threshold value for distinguishing the categories, so that there was always 20ms between the longest VOT /d/ and the shortest VOT /t/.

For example, if the VOT threshold for /d/ vs. /t/ was 40ms, we selected four tokens of /d/ with VOT 15, 20, 25, 30ms and four tokens of /t/ with VOT 50, 55, 60, 65ms to be used in the experiment. Each subject thus received stimuli customized to their observed threshold values from this set. (To preview the results, the mean and mode of the threshold boundary for all subjects in the study was 40ms (SD=3.6ms). Each subject was presented with a total of 1700 trials, divided into two blocks: One block with /d/ stimuli as standard and /t/-stimuli as deviants; and one block with /t/-stimuli as standards and /d/-stimuli as deviants. The order of blocks was counterbalanced with about half the subjects in each order.

Subjects were randomly assigned to two different groups, each of which received the stimuli in different block orders. In the first block of 850 trials, one of the two phonological categories was used as stimuli in each of two levels of the factor ODDBALL “standard” vs. “deviant”, e.g. “standard-D” and “deviant-T”. In the second block, this relationship was reversed, such that the phoneme that was deviant in the first block was now standard in the second block. About half the subjects were put in the group where /d/ was the deviant in the first block and about half the subjects were put in the group where /t/ was the first deviant. Each stimulus was randomly sampled at each occasion from the four different VOT values for each category, selected for that subject. The stimuli were delivered continuously, with a random number (between 2 and 7) of standards between each deviant. The interstimulus interval (ISI) randomly varied between 700ms and 890ms in increments of 20ms.

The 850 trials in each block consisted of 100 deviants (12.5%) and 700 standards (87.5%), along with 50 target stimuli, which were either a male or a female voice saying “ba”. The task for the subject was to decide whether the voice saying “ba” was a male or a female. Four different “ba” syllables were used, varying in fundamental frequency, to make it sound either male or female. The subject pressed a response box button to each target to indicate his/her decision, and received accuracy feedback. With each behavioral response, the screen would gradually fill up with different emoticons, so that by the end of the experiment, the whole screen would be filled up. This provided some entertainment and reward for the subjects as well as an indication of the progress of the experiment. The entire sequence of trials was presented without pauses, but the experimenters were able to pause the experiment at any time if necessary. The entire recording session lasted about 45 minutes.

2.1.4 Apparatus, data acquisition, and post-processing

The experiment was programmed with E-Prime Professional software v. 2.0.8.90, running on a Dell desktop PC. E-Prime Extensions for Net Station v.2.0 was used for communications with a 128 channels Electrical Geodesics, Inc. 300 system, using Ag/AgCl plated electrodes housed in electrolyte-soaked sponges. Data were acquired and digitized with EGI Net Station software v.4.5. Subjects were comfortably seated inside a single-walled 9x10 feet International Acoustics Company electrically shielded sound booth. Sound stimuli were presented with two free field speakers placed in front of the subjects at comfortable listening volume; visual input was delivered through an LCD display placed on a table in front of the subjects. Behavioral responses were recorded with the PST Serial Response box. Subjects' head was not placed in a headrest and was free to move.

Before data collection, electrode impedances were lowered to below 50 kΩ (appropriate for the high-density EEG system, [46]. The electroencephalogram (EEG) was continuously recorded with a 24-bit digitization at 250Hz. The analog recording passed through a 0.10Hz first-order highpass filter, and each electrode was referenced to Cz during recording. After recording, the continuous EEG was segmented into epochs of 1000ms. Each epoch included a 200ms pre-stimulus period before the stimulus onset (to be used for baseline correction), thus resulting in 800ms of data for each single sound presentation. Each subject's data was then submitted to an automatic artifact detection procedure for identifying bad channels, eyeblinks and eye movements: A channel was marked bad if the difference between maximum and minimum voltage exceeded 200 μV in a moving average of 80ms. Channels marked as bad in over 20% of trials were considered bad in *all* trials. Trials containing more than 10 bad channels, eye blinks or eye movements were marked as bad. Bad channels were then replaced using the spherical spline interpolation. Each trial was then baseline corrected using then mean voltage of the first 200ms.

After this step, the data were submitted to a second automated procedure which performed an independent component analysis [47] and automatically subtracted eyeblink components that correlated at r = 0.9 or greater with an eyeblink template generated from the data via visual inspection. The single trials were then averaged into each of the four cells of the design ("deviant-D", "standard-D", "deviant-T" and "standard-T"). The data were finally referenced to the average voltage, which is the least biased reference method with high-density EEG [48], [49].

2.1.5 ERP analysis strategy

Our analysis strategy was aimed at simplifying the nature of the dependent measures that are used as inputs to statistical analysis. All analyses were conducted on the difference waves obtained by subtracting the standard waveform for a given phoneme from the deviant waveform for the same phoneme. In this way, we abstract away from the obligatory evoked potentials (such as the N1-P2 complex) in the data and focus on the temporal and spatial distribution of the experimental effects, independently of the other major voltage fluctuations related to the evoked auditory potential. The reason for this is that the MMN, although often characterized as an attenuation of the P2, in fact extends beyond the P2 peak and often is observed in the time window just after the P2.

In addition to the classical analysis of ERPs done by averaging the observed voltage fluctuations observed at the scalp over electrode regions and time windows, we also made extensive use of sequential temporo-spatial PCA[50]–[56]. This is a member of an emerging family of “blind source separation” methods for ERP analysis, which semi-automatically decomposes the temporal and spatial dimensions into a smaller set of abstract ERP components or factors (based on covariance patterns among time points or electrode sets). It has the advantage that it teases apart the many underlying contributions to the summed scalp activity and allows the researcher to more sharply focus on where the experimental effects are in high-dimensional data. It also provides a more objective way of picking time windows and electrode regions, and obviates the need for complex “blind” ANOVA schemes for discovering effects. The abstract factors can themselves be directly analyzed as dependent measures via their factor scores. After preprocessing the data with the temporo-spatial PCA in the first step, we then used the factor analysis in a second step to constrain and inform the selection of time/space regions in the untransformed data (to be detailed below) It is beyond the scope of the current paper to provide further extensive justification for temporo-spatial PCA; we direct the reader to the literature on PCA decomposition [47], [52], [54] for further justification and explanation.

We also used difference waves as the input to the PCA, in order to focus on the temporal and spatial fluctuations of the mismatch effect itself, rather than letting the PCA focus on the temporal and spatial distribution of the main effect amplitude changes of the absolute waveforms (as those peak and valleys would be what accounted for the most variance in the data, rather than the mismatch effect itself). Thus, every experiment is analyzed along both dimensions: First, with temporo-spatial PCA, then with PCA-constrained classical analysis of the scalp-recorded voltages. Crucially, the second type of analysis provides the basis for comparison to previous studies that did not use PCA-based analysis.

2.2 Results

After EEG recording and post-processing, one subject had only 21% good trials and was excluded. Two additional subjects were excluded due to experimenter error (no EEG data collected). Finally, we decided to exclude 8 more subjects based on them having outlier VOT threshold values in the behavioral pre-test (30ms, and 50, 55 and 60ms). This exclusion was based on the following reasoning: The mean VOT population threshold for the d-t continuum in our stimuli was 40ms (SD= 5ms), estimated from a larger subject pool of 135 subjects (data came from the behavioral pretest of category boundary for these subjects and subjects from an additional study.) Inspection of the peak latency of the P2 wave of the Auditory Evoked Potential, pooled data from all experiments, showed that a syllable with 40ms VOT resulted in a P2 peak at about 200ms, with each 5 ms difference in VOT moving the peak about ±10ms (on average). Inclusion of outlier VOT subjects would therefore likely smear the mean latency of the P2 wave in the data. We therefore decided to limit the subjects to those having 35, 40 and 45ms VOT thresholds.

After subject exclusions, most of the remaining 24 participants had about 20% loss of trials due to artifacts. The mean proportion of good trials for the remaining 24 subjects was 80% (SD=15%, range 45%-97%). 6 subjects with less that 75% good trials (62%-74%) were visually screened to determine whether they still had obligatory Auditory Evoked Potential (AEP) responses to the standards. They all did, so all 24 subjects were included for analysis. 17 of the 24 subjects were women and 7 were males (this imbalance arises from the fact that the population we sampled from was overrepresented with women). Four subjects were left-handed[5]. The mean age was 23.5 (SD=5.5, range = 18-44; only 3 subjects were older than 26).

The mean accuracy of the target detection task for the 24 subjects was 97% (SD 1.8%); hence, all subjects attended carefully to the stimulus stream. Visual inspection of the grand average topographic voltage map revealed a typical AEP with an N1-P2 waveform complex at central to anterior electrodes, inverting at the mastoids. A mismatch effect was evident in the P2 peak as well as in the later part of the waveform (300-500ms); in addition, a bilateral slow-wave negativity to deviants was observed at inferior anterior electrodes. Difference waveforms (deviant /d/ minus standard /d/, and deviant /t/ minus standard /t/) were computed, and input to a temporal PCA followed by spatial PCA of each temporal factor[6]. In the first step of the PCA, the single subject averages were combined into a matrix with 250 time points as columns, and subjects, cell averages and electrodes as rows, providing the structure for temporal PCA. Using the scree test in combination with the Parallel Test [48], 12 temporal factors were retained in this initial step, accounting for 86% of the total variance. The factors were then rotated to a simple structure using PROMAX rotation (k=3) with Kaiser correction. To further delineate these

---

[5] We decided *a priori* not to exclude left-handed subject. There are several reasons for this. For example, left-handers process language normally, and most left-handers also have left-lateralized language functions. We had no specific prediction that our measures would differ for left-handers; indeed, the MMN has generators in both the left and right hemispheres, as seen in the ERP by the typical inversion of the MMN at both left and right temporal sites. Left-handers are also part of the normal population and should therefore be represented in research, by ethical rules for representativeness of subject sampling from subpopulations. Finally, self-report of handedness without extensive neuropsychological testing is probably unreliable. We also did a post-hoc review of the AEP in the few left-handers (as part of general single-subject review) and did not observe any abnomalies. Even if there were differences unknown to us, it would be part of the overall subject-as-random factor, and would be unlikely to affect the results (beyond reducing the standard error) given the sample sizes reported here.

[6] The *ERP PCA Matlab Tool* in combination with *EEGLAB* was used in all PCA analyses.

effects, the temporal factors were next submitted to spatial decomposition by inverting the matrix so that the electrodes now are the columns. Scree test determined six spatial factors to be retained for each temporal factor and rotated to simple structure using INFOMAX (i.e. ICA, following recommendations in Dien [54]). Note that this yields 12 x 6 = 72 temporo-spatial factors; however, only a small set of these factors correspond to ERP components that aligns with observable experimental effects in the grand average voltage data. Our strategy was to identify those temporo-spatial components that matched observable effects in the grand average voltage data, with the constraint that the temporal factor had to account for at least 5% of the total variance (following the guidelines of [47], [54]). Specifically, we sought to identify the component that corresponded to the MMN during the P2 peak, as well as the component that corresponded to the Late Discriminatory Negativity ERP. Three temporal factors met the criterion of accounting for at least 5% of the variance, and two of these factors clearly corresponded with observable effects in the grand average voltage data.

The first temporal factor TF1 corresponded to a late and broadly distributed anterior negativity to the deviants (peak latency 652ms), and this factor accounted for 59% of the variance in the data. The second temporal factor TF2 (peak latency 360ms, central-anterior distribution) accounted for 7% but did not match up with a clear effect in the data and was therefore discarded. The third temporal factor TF3 (peak latency 216ms) accounted for 6.5% of the variance, and clearly corresponded to an MMN during the P2 peak (based on visual comparison of the temporal and spatial distribution of the factor score wave with the raw grand average voltage. cf. Figure 2). Each temporal factor was then submitted to a special ICA decomposition to further narrow down the major sources of spatial variance. After spatial ICA of these factors, we again analyzed only those spatial sub-components that had a distribution consistent with *a priori* established ERP components in MMN studies (again, following the guidelines in Dien [54]). For TF1, the first spatial factor TF1SF1[7] had an anterior distribution consistent with Late Discriminatory Negativity (400-600ms). The first and second spatial subfactor of TF3 exhibited mismatch effects; the TF3SF1 had a posterior distribution and TF3SF2 had a central-anterior distribution consistent with MMN. As only the latter was clearly consistent with the previous literature on MMN, it was selected for further analysis.

Analysis next proceeded as follows: first, we analyzed the factor scores for the two temporo-spatial ERP components with an ANOVA, with the within-subject condition PHONEME (/d/ vs. /t/, each represented as difference waves) and the between-subject condition BLOCKORDER (/d/ as first-deviant vs. /t/ as first-deviant). Because difference waves are used as dependent measures, a main effect of mismatch translates into a significant intercept in the general linear model for the ANOVA. A main effect of PHONEME is equivalent to a condition x phoneme interaction. A main effect of BLOCKORDER would mean that the MMN was different in the two blocks; finally, an interaction between BLOCKORDER and PHONEME would mean that the ordering effect was not the same for both phonemes. After analyzing the factor scores, we next analyzed the raw voltage data in the same way, with dependent voltage measures constrained by the PCA analysis (see below for details).

2.2.1 MMN (216ms peak latency)

The third temporal factor, spatial subfactor 2 (TF3SF2), with a peak latency of 216ms, and a central-anterior distribution, is illustrated in Figure 2. This represents a classic mismatch modulation of the P2 peak of the Auditory Evoked Potential. The time course of this factor

---

[7] TF1SF1 should be read, “temporal factor 1, spatial factor 1”, i.e., the first spatial subfactor of the first temporal factor.

indicated a peak latency at 216ms, which was consistent with the peak of the P2 in the non-difference waveform raw voltage data. The upper panel of Figure 2 shows the factor back-projected into voltage space, with the time course illustrated in the left figure at electrode FCz (both difference waveforms for /d/ and /t/), and the spatial distribution of the main effect illustrated in the right figure. Note that because the difference between deviants and standards for /t/ is essentially 0, as is evident in the raw voltage plot, the factor score waveform for the difference wave is essentially a straight line around 0. The lower panel shows both the absolute waveforms and the difference waveforms for each phoneme separately (/d/ to the left, and /t/ to the right), as well as a box indicating roughly the time window selected for analysis. Note that the raw voltage waveform shows a classical Auditory Evoked Potential, with a P1 around 80ms, an N1 around 120ms, and P2 peak around 210ms (identified as 216ms by the factor analysis). As is evident, the majority of the mismatch effect is located around the P2 peak. (In this case, the component analysis is almost redundant, but we keep it for consistency with the rest of the analyses.)

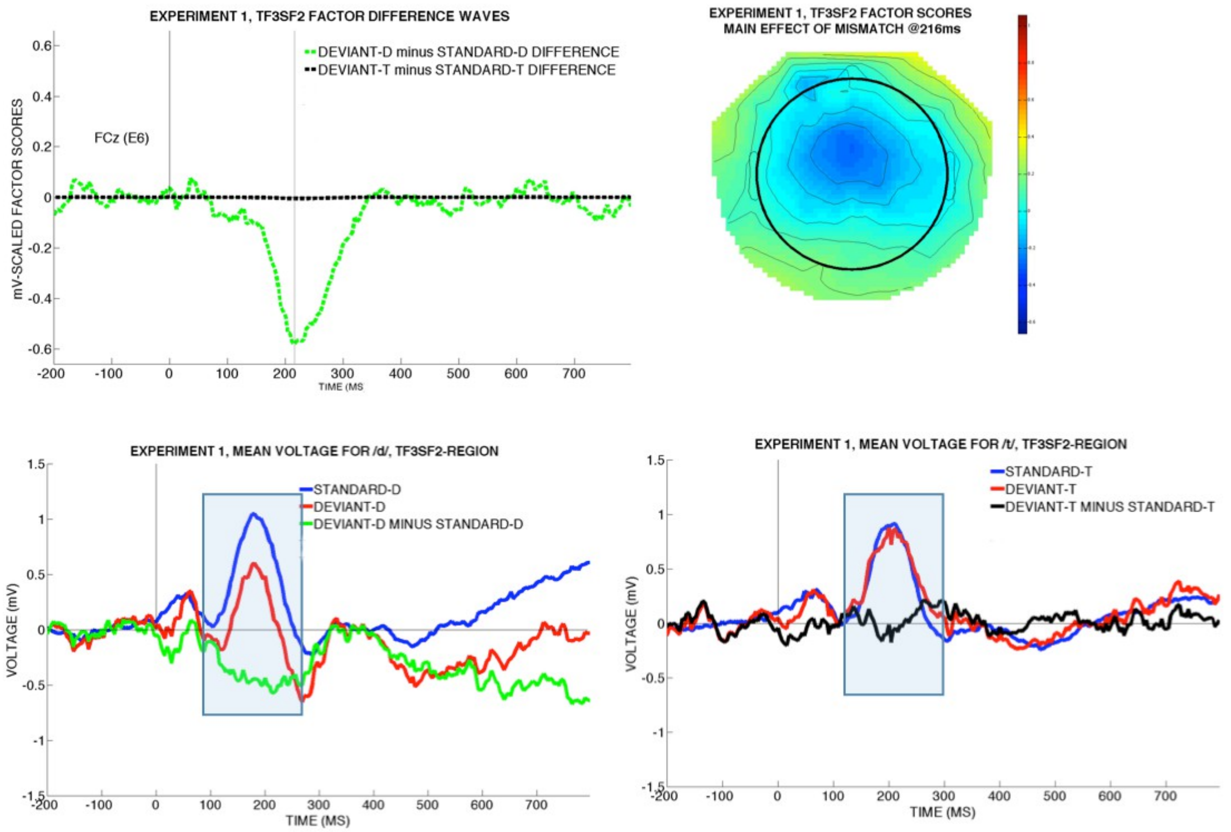


Figure 2: Experiment 1, early MMN effect. Upper panel: Time course (left) and spatial distribution (right) of temporo-spatial factor decomposition; the topoplot shows the mean difference wave at the horizontal line in the waveform plot at 216ms. Lower panel: corresponding absolute voltage waves and difference waves in the raw grand average voltage for /d/ (left panel) and /t/ (right panel).

An ANOVA of the factor scores with within-subject factor PHONEME and between-subject factor BLOCKORDER resulted in a marginal intercept (F(1,22)=4.1, p=.054) and a marginal main effect of PHONEME (F(1,22)=4.2, p=.052). The first effect is interpretable as a main effect of mismatch, and the second effect is interpretable as an interaction between mismatch and phoneme such that there was a mismatch effect for /d/ but not for /t/. As is apparent, this marginal effect in the factor scores is due to a greater mismatch for /d/.

Analysis of the raw data voltage was done by averaging electrodes with TF3SF1-factor loadings greater than 0.6 (roughly the blue box in Figure 2) during the time window defined by temporal factor samples with factor loadings greater than 0.6 (which corresponded to the 188-268ms time window). This resulted in a significant intercept (i.e. a main effect of mismatch, F(1,22)=5.8, p=.02). A marginal effect of BLOCKORDER (F(1,22)=3.8, p=.06) was observed, but this effect is not interpretable by itself vis-à-vis the hypothesis, as it only means that the MMN was overall greater in the first block. The ANOVA also revealed a main effect of phoneme PHONEME (F(1,22)=6.6, p=.01), such that the MMN was bigger for /d/ (-0.44mV) than for /t/ (which was 0). Finally, the interaction PHONEME x BLOCKORDER was significant (F(1,22)=11.2, p<.01); inspection of the interaction plot showed that the interaction was driven by a bigger MMN in the first block than in the second block, and that this difference was greater for /d/ than for /t/. To aid the interpretation of this interaction, consider Figure 3.

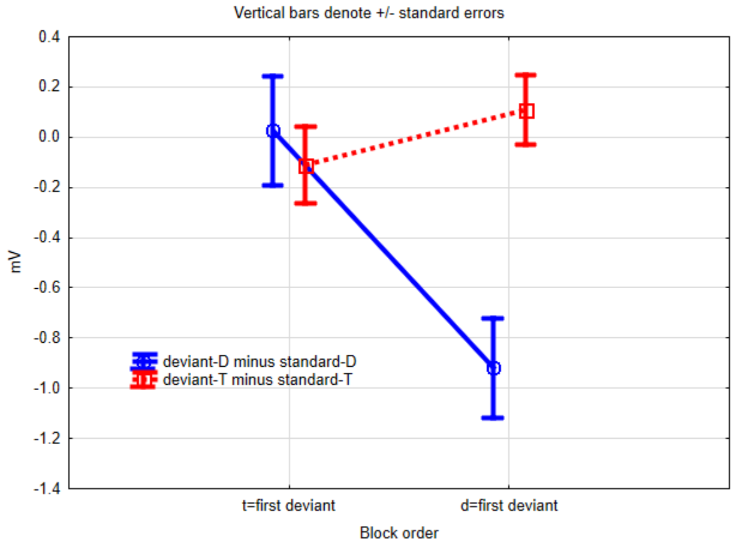


Figure 3: Interaction plot PHONEME (difference wave) x BLOCKORDER. Vertical bars denote +/-standard error.

As is evident, the mismatch for /d/ when it is the first deviant is greater than when /d/ is presented as deviant in the second block (in fact the effect disappears); similarly, the mismatch for /t/ is more negative when /t/ is the deviant in the first block as opposed to when it is the deviant in the second block. However, it does not reach significance. Orthogonal contrast for /d/ when it is the first deviant is highly significant (effect size: -0.44mV, t=-3.03, p<.01), whereas the contrast for /t/ when it is the first deviant is not significantly different from 0 (effect <.001mV, t=-.009, p=.99).

This block order interaction was pervasive in the current study; the same “first-mention advantage” pattern was observed in all three experiments. We report similar interaction statistics for the remaining experiments, but space limitations prevent us from showing the interaction plot for each effect. The explanation is most likely that due to the within-subject block design, once the subject is exposed to the same stimuli as oddball that he/she just heard 700 standard samples of, the surprise effect is attenuated. Such “first-mention” advantage effects have recently been addressed in the MMN literature.[57]

2.2.2 Late Discriminatory Negativity (TF1SF1, 416-800ms)

The largest factor observed in the data was related to the first spatial decomposition of the first temporal factor, TF1SF1. This factor matched the temporospatial location of the Late Discriminatory Negativity [29], [58]–[62], i.e., a late, slow, negativity with a broad anterior inferior distribution, peaking at FCz (EGI channel 6). The effect was driven by a large negativity when /d/ was the deviant, with no such effect for /t/. Using this temporo-spatial PCA factor as a guide, we next constructed an average voltage based on the electrodes with factor loadings > 0.6

(roughly the blue area in the topoplot in Figure 4). The difference waveform for this region-of-interest is shown for /d/ and /t/ in panel 2 in Figure 4, along with the absolute waveforms for the standard and deviant conditions. The time samples with factor loadings greater than 0.6 (416-800ms) is marked with a shade over the grand average voltage waveforms.

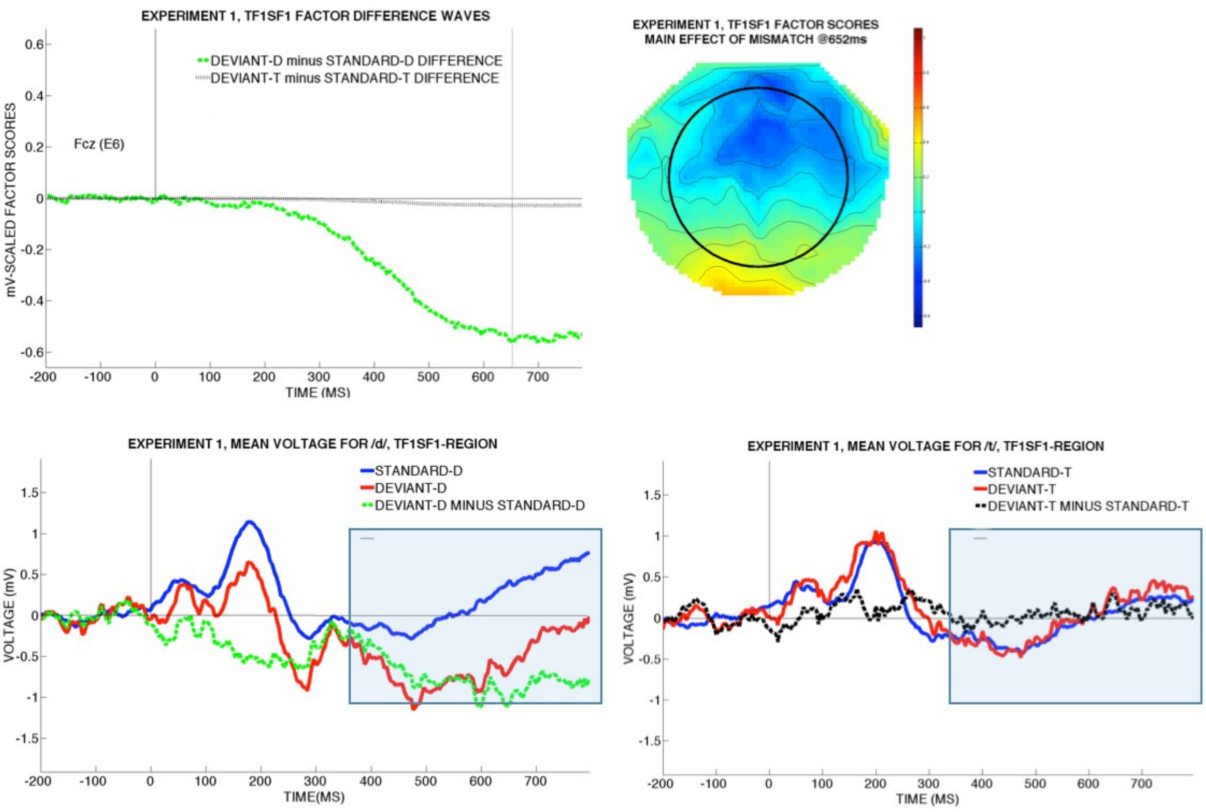


Figure 4: Experiment 1, Late Discriminative Negativity. Upper panel: Time course (left) and spatial distribution (right) of temporo-spatial factor decomposition; the topoplot shows the mean difference wave at the horizontal line in the waveform plot at 652ms. Lower panel: corresponding absolute voltage waves and difference waves in the raw grand average voltage for /d/ (left panel) and /t/ (right panel).

The factor scores were analyzed with an ANOVA with the within-subject factors PHONEME (2 levels) and the between-subject factor BLOCKORDER. The ANOVA did not yield significant effects, but the intercept was marginally significant (F(1,22)=3.02, p=.09), indicative of a main MMN effect. Given that we had a priori predictions, we conducted separate t-tests for the two factor score difference waves. The t-test against 0 for the /d/ difference wave was marginally significant (t(22)=-2.05, p=.05). However, because the experiment predicted a difference in the negative direction (i.e. a Mismatch Negativity), this allows us to interpret this with a one-tailed probability, which makes it significant. The t-test for /t/ was not significant by a one-tailed test (t=-0.13, p=.89).

We next conducted the same ANOVA for the raw voltage difference waves shown in the lower panel of Figure 4. To clarify, the dependent measure is the actual, unweighted, voltage averaged over the TF1SF1-electrodes and the TF1SF1 time samples with factor loadings > 0.6, roughly the time window indicated with a box. This ANOVA resulted in a significant intercept (F(1,22)=5.8, p=.02), which means that there was a main effect MMN; a significant main effect of PHONEME (F(2,22)=5.07, p=.03), which is interpretable as a PHONEME x MMN interaction. This interaction was due to the difference wave for /d/ being less than zero, whereas the difference wave for /t/ was not different from zero. Finally, the ANOVA revealed an interaction PHONEME x BLOCKORDER (F(1,22)=4.53, p=.04). Inspection of the interaction plot revealed that this interaction was due to a greater mismatch for the phoneme that was presented as deviant in the first block, and such that this "first-mention" advantage was significantly greater for /d/ than for /t/.

2.3 Discussion

Experiment 1 showed a clear asymmetry: an MMN was observed for /d/, but not for /t/, as seen in the results for the classic MMN effect in the central-to-anterior during the P2 peak of the

Auditory Evoked Potential. Similarly, a second temporo-spatial component of the ERP response interpretable as the Late Discriminatory Negativity exhibited the same asymmetry: a mismatch for /d/ but not mismatch for /t/. Both these observations bear out the predictions of the theory: Laryngeal features (spread glottis) are underspecified for English stops (and consonants generally), such that the /d/ phoneme does not contain a specification for this feature. Linked to the assumptions about varying standards in MMN experiments, this predicts that when deviant [t] is compared to phonemic /d/, there is no direct feature conflict. In other words, the comparison of the oddball and the memory trace of /d/ should not contribute to the overall MMN effect. In fact, this experiment showed no MMN for /t/ at all in the early time period between 100-300ms comprising the P2 component. This also replicates the finding of a mismatch for /d/ but not for /t/ that was reported in Phillips et al [34].

One possible confound existed in this experiment: We used a target-tracking task, to ensure attention to the auditory stream. The targets that the subjects tracked began with a labial voiced consonants ("ba"). This could conceivably bias the perceptual system to be more sensitive to voiced syllable onsets (this generating a greater MMN for [da] compared to [ta]). To rule this out, we ran the same experiment again but without the target-tracking task, and instead directed subjects' attention away from the auditory stimuli. This also allowed us to test whether the observed asymmetry holds under non-attention conditions, as the MMN is known to be elicited automatically without requiring attention to the auditory stream.

3. EXPERIMENT 2: PASSIVE MMN

3.1 Methods

3.1.1 Subjects

A total of 49 University of Delaware students were recruited as subjects and provided written consent in Experiment 2. Each subject received course credit for participation. Three subjects were excluded from analysis because they had VOT thresholds outside the 35-45ms range (50 and 55ms respectively). One subject was excluded for being bilingual. Two subjects were excluded because of recording errors. One subject was excluded for being heavily medicated and having excessive artifacts. Finally, ten subjects reported a history of hearing loss or having received speech/language therapy in the past and were excluded from analysis. The remaining 32 subjects had on average 92% good trials after artifact correction. 26 of the 32 subjects were women and 6 were males (again this imbalance arises from the fact that the population we sampled from was overrepresented with women). 6 subjects were left-handed, but we did not exclude left-handers, as most left-handed people have left-lateralized language function. The mean age was 19 (SD=1.2, range = 18-23).

3.1.2 Stimuli

The stimuli were identical to those of Experiment 1.

3.1.3 Experimental design and procedure

The design and procedure were identical to that of Experiment 1, except that no tracking task stimuli were presented, and no behavioral responses were required of the subjects. Instead, subjects watched the original black and white movie *The Wizard of Oz*, with the sound track turned off, during the entire data collection stage.

3.1.4 Apparatus, data acquisition, and post-processing

Data recording and data post-processing procedures were the same as in Experiment 1.

3.2 Results

Temporal PCA using the same procedure as in Experiment 1 retained 12 temporal factors. Only the first three factors each accounted for more than 5% of the total variance: TF1 (53%) peaked at 776ms, TF2 (7%) peaked at 208ms, and TF3 (6%) peaked at 332ms. After spatial decomposition, retaining 6 spatial factors per temporal factor, three components had topographies consistent with known ERP components: TF1SF1 was consistent with a LDN component; TF2SF2 was consistent with a classical central MMN during the P2 peak, and TF3SF1 was consistent with a late MMN at central-to-anterior electrodes. Each component exhibited an MMN for /d/ but not for /t/; the amplitude was much smaller, however, than in Experiment 1 where subjects directed their attention to the stimulus stream. For space limitations, we only report on the components corresponding to the classical early MMN and the Late Discriminatory Negativity.

3.2.1 Early MMN (TF2SF2, 208ms)

The classical MMN in Experiment 2 is illustrated in Figure 5; as can be seen in the lower panel, only /d/ showed a typical MMN pattern with attenuation of the deviant wave. In fact, in the /t/ condition, the deviants were more positive than the standards. This is reflected in the opposite polarities of the PCA factor difference scores in the upper left panel of Figure 5.

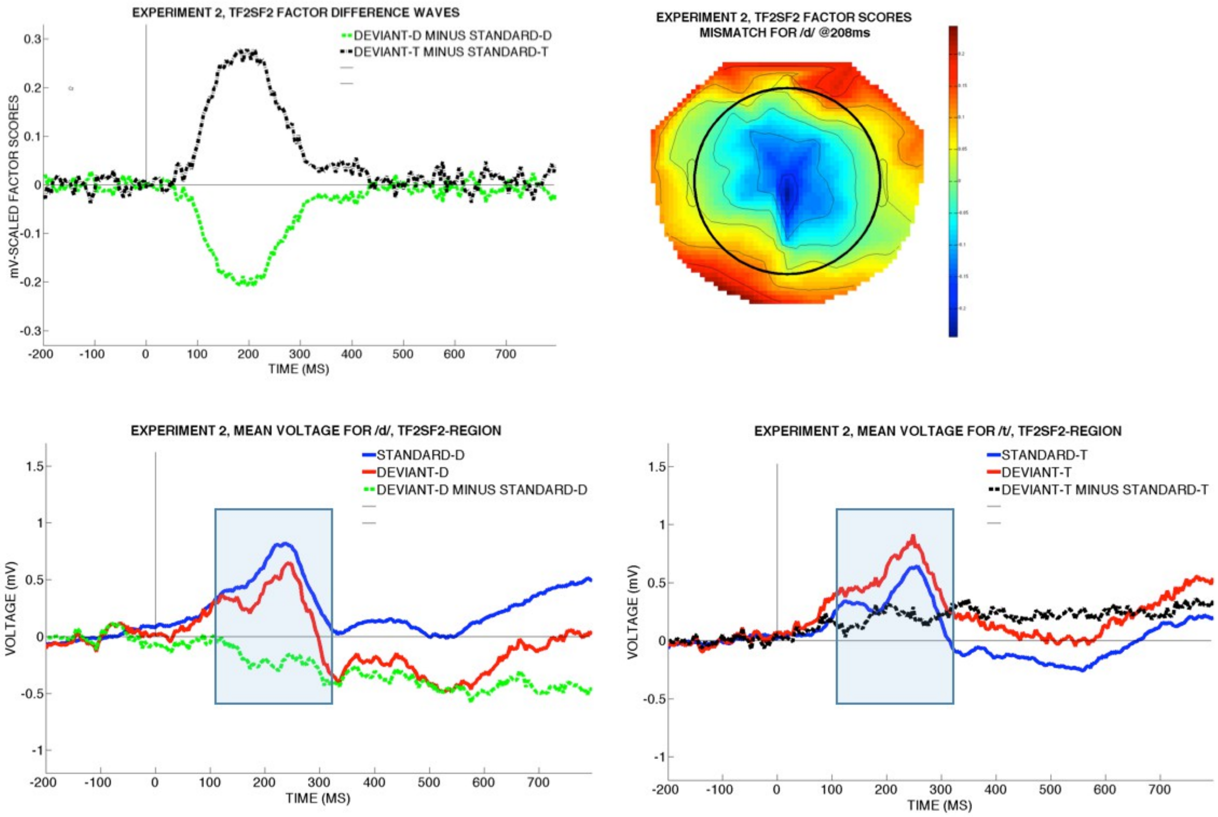


Figure 5: Experiment 2 MMN. Upper panel: Time course (left) and spatial distribution (right) of temporo-spatial factor decomposition; the topoplot shows the mean difference wave at the horizontal line in the waveform plot at 208ms. Lower panel: corresponding absolute voltage waves and difference waves in the raw grand average voltage for /d/ (left panel) and /t/ (right panel).

An ANOVA of the difference wave factor scores with PHONEME as within-subject, and BLOCKORDER as between-subject revealed a PHONEME x BLOCKORDER interaction (F(1,30)=25.5, p<.0001). The source of this interaction was the same “first-mention” advantage observed in Experiment 1, such that the phoneme that was deviant in the first block showed a bigger MMN than the phoneme that was deviant in the 2nd block, and such that this advantage was greater for /d/. Orthogonal contrast analysis restricted to the first block for each phoneme resulted in a significant mismatch for /d/ (-0.05mV, t=-5.1, p<.0001), and no significant contrast for /t/ (-0.02mV, t=-1.06, p=.29). Note that this means that the apparent “reversal” is in fact not significant and therefore has no specific interpretation.

This analysis was replicated with an ANOVA of the raw voltage data, restricted to the time samples with factor loadings greater than 0.6 (136-236ms) and electrodes with factor loadings greater than 0.6, resulting in a PHONEME x BLOCKORDER interaction (F(1,30)=27.1, p<.0001). Inspection of the interaction plot revealed this interaction to be driven by the “first-mention” advantage, such that the MMN is greater for the phoneme that was the deviant in the first block of the experiment. The driver of the interaction was that this first-mention advantage was greater for /d/ than for /t/. For this reason, we again conducted orthogonal contrast analysis of the MMN separately for /d/ and /t/ for the first block only. When /d/ was the first deviant, the difference between deviant /d/ and standard /d/ was 0.41mV; this was significant (t=-5.6, p<.0001). The contrast for /t/ when /t/ was the first deviant was not significant (0.21mV, t=-1.5, p=.14).

3.2.3 Late Discriminative Negativity (LDN): TF1SF1

We finally turn to the Late Discriminatory Negativity component in Experiment 2. This was a slow wave starting at 420ms (based on when TF1’s temporal factor loadings exceeded 0.6). Again, inspection of the corresponding voltage data, presented in Figure 6, showed a mismatch effect for /d/ and not for /t/ (the deviant waveform was again more positive than the standard in the grand average).

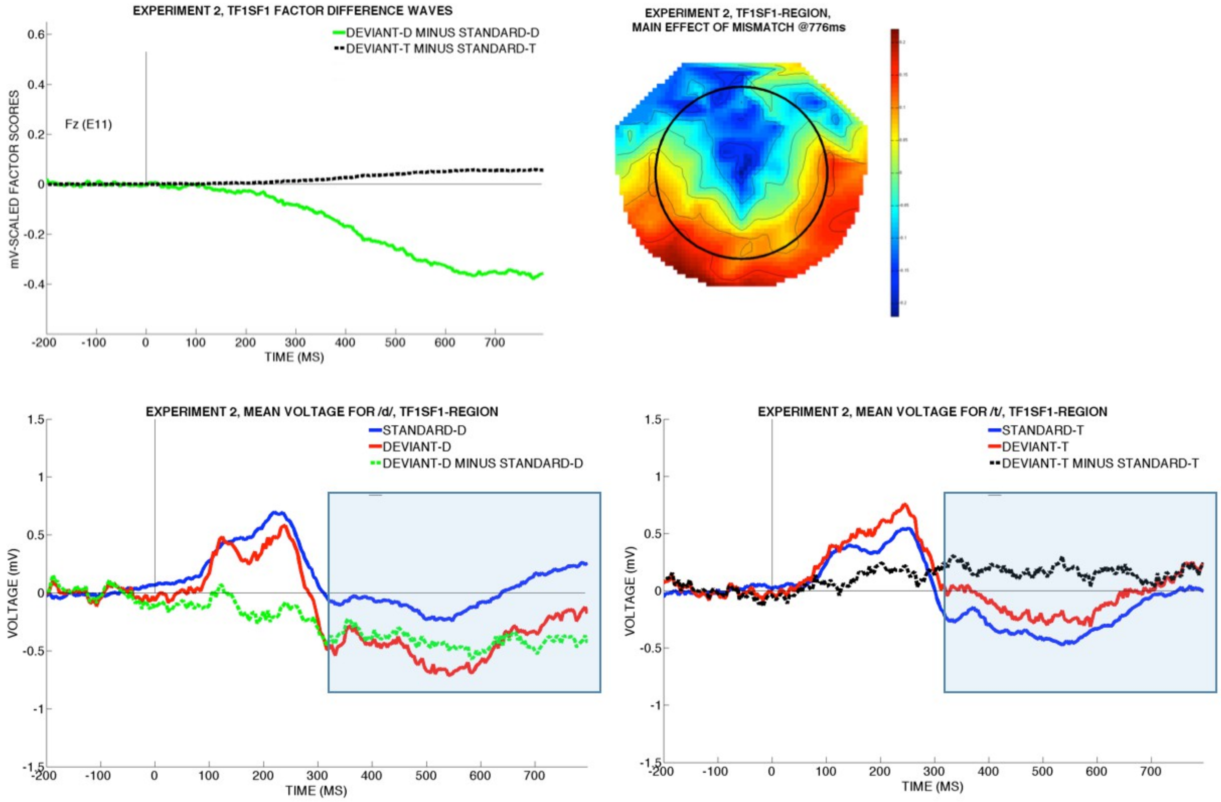


Figure 6: Experiment 2 Late Discriminatory Negativity. Upper panel: Time course (left) and spatial distribution (right) of temporo-spatial factor decomposition; the topoplot shows the mean difference wave at the horizontal line in the waveform plot at 776ms. Lower panel: corresponding absolute voltage waves and difference waves in the raw grand average voltage for /d/ (left panel) and /t/ (right panel).

An ANOVA of the factor scores resulted only in a main effect of mismatch (significant intercept, F(1,30)=6.9, p=.01). However, analysis of voltage restricted to time samples exceeding factor loadings of 0.6 (420-800ms) and electrodes with factor loadings greater than 0.6 resulted in a small set of electrodes so we relaxed the criterion (to 0.4 to sample from a larger area). This revealed a main effect of mismatch (F(1,30)=6.9, p=.01) and a PHONEME x BLOCKORDER interaction (F(1,30)=14.5, p<.001). Again, this interaction was due to greater mismatch for the first-block; orthogonal contrast analysis showed the mismatch for /d/ as first deviant to be highly significant (0.67mV; t=-4.7, p<.0001). Here, the contrast for /t/ as first deviant was significant (0.59mV; t=-2.4, p=.02). Thus, there was a mismatch for both phonemes when that phoneme was presented as deviant in the first block, but the mismatch was significantly greater for /d/ (hence the interaction).

3.3 Discussion

The same asymmetry between voiced and voiceless /d/ vs. /t/ observed in Experiment 1 was also observed in Experiment 2, where subjects had their attention directed away from the stimuli. (Note that the apparent "reversal effect" for /t/ seen in Figure 5 does not contradict this, because the reversal is only true descriptively; statistically there was no difference between standard and deviant.) We conclude that the tracking task that required paying attention to a different set of voiced syllables in the auditory stimuli was not the cause of the bias for /d/ seen in Experiment 1, because the same effects are observed without attention to the tracking stimuli. In addition, this experiment shows that the underspecification asymmetry is reflected in the automatic MMN response (i.e., elicited in the absence of attention).

4. EXPERIMENT 3: PHONETIC MMN

In Experiment 1 and Experiment 2, it was critical that the memory trace was constructed by the auditory cortex's use of a phonemic representation, which we induced by varying the standard tokens within category. If the standard tokens are not varied but kept constant, then the memory representation of the standards can simply be a copy of the recurring phonetic input. In this case, we predict no asymmetry: if an oddball phonetic token is compared to a standard phonetic memory trace, and phonetic representations are fully specified, then the same feature conflict with arise whether [d] or [t] is the oddball. Phonetic [t] compared to phonetic [d], or vice versa, will involve a contrast in the phonetic feature matrix: [d] is represented by [-spread larynx] and [t] is represented as [+spread larynx]. To test this prediction, we repeated Experiment 1 but used a single exemplar for /d/ and single exemplar for /t/.

4.1 Methods

4.1.1 Subjects

36 University of Delaware students were recruited as subjects and provided written consent in Experiment 3, and received course credit for participation. One subject was excluded because of recording error; one subject reported being bilingual and was excluded; one subject was epileptic and was excluded; finally, four subjects were excluded based on speech/language therapy history. The mean age of the remaining 29 subjects was 22.8 (SD=3.6); 16 subjects were female; 13 subjects were male. Four subjects were left-handed. The mean proportion of good trials after artifact removal was 93% (SD=5%). Thirteen subjects were in the “d as first deviant” ordering group and 16 subjects heard “t” as the first deviant.

4.1.2 Stimuli

Only two token stimuli were used in this experiment; one token representing [d] (with VOT=20ms) and one token representing [t] (with VOT=60ms). Note that the VOTs were equidistant from the mean threshold of 40ms, differing with 20ms in each direction.

4.1.3 Experimental design and procedures

No behavioral pre-test was conducted for the subjects in Experiment 3, as they were all exposed to the same two fixed tokens (selected on the basis of an estimate of the mean identification function threshold—40ms--for all subjects up to this point). Subjects were engaged in the same behavioral tracking task as in Experiment 1.

4.1.4 Apparatus, data acquisition, and post-processing

Data recording and data post-processing procedures were the same as in Experiment 1.

4.2 Results

The average target detection accuracy was 95% (SD=4%), so subjects were paying good attention. Visual inspection revealed an early P2 peak (around 150ms) with a mismatch for both /d/ and /t/. In addition, visual inspection of the grand average showed a mismatch for /t/ at central to left electrodes, but also a mismatch for /d/ at more anterior electrodes. Temporal PCA on the difference waves resulted 13 retained temporal factors and subsequent spatial PCA resulted in 5 retained spatial factors for each temporal factor. TF1 (776ms) accounted for 48% of the variance, TF2 (292ms) accounted for 8% of the variance, TF3 (400ms) accounted for 6%, TF4 (248ms) accounted for 5% and TF5 (132ms) accounted for 4%. TF5 falls below our criterion of variance accounted for, but inspection of the temporo-spatial factors clearly indicated that this factor corresponded to the early MMN, so we included it for analysis. As in Experiment 1 and 2, we focused on the temporo-spatial factors that clearly corresponded to the classical early MMN (TF5) during the Auditory Evoked Potential, and the Late Discriminatory Negativity (TF1).

4.2.1 MMN (TF5SF1-132ms)

The temporo-spatial factor corresponding to the peak of the Auditory Evoked potential (TF5SF1), exhibited a clear central-anterior MMN, present for both /d/ and /t/ (Figure 7).

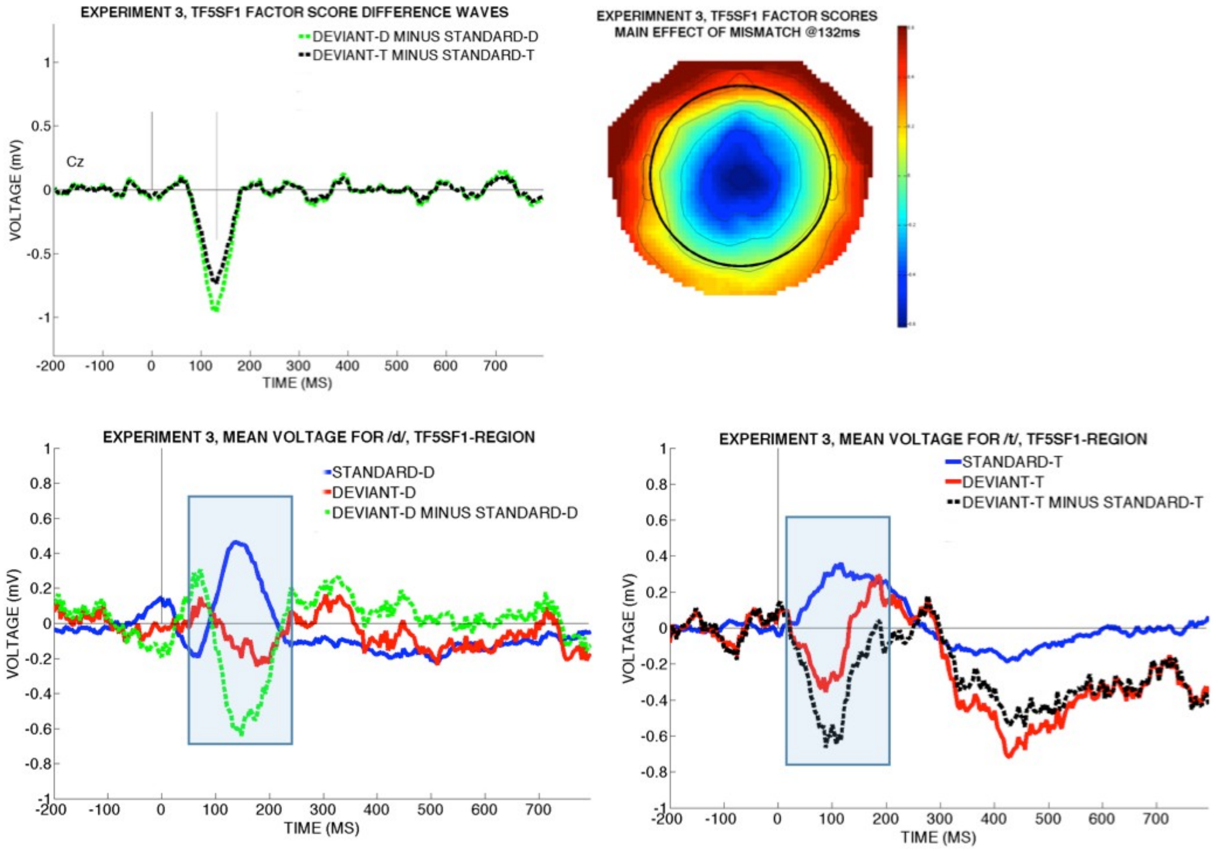


Figure 7: Experiment 3, phonetic MMN. Upper panel: Time course (left) and spatial distribution (right) of temporo-spatial factor decomposition; the topoplot shows the mean difference wave at the horizontal line in the waveform plot at 132ms. Lower panel: corresponding absolute voltage waves and difference waves in the raw grand average voltage for /d/ (left panel) and /t/ (right panel). Both phonemes show MMN difference waves.

An ANOVA of the factor scores resulted in a significant intercept, meaning a main effect of mismatch (F(1,27)=15.9, p<.001); and a PHONEME x BLOCKORDER interaction (F(1,27)=10.8, p<.01). The BLOCKORDER effect was also significant but has no interpretation, as it does not involve a condition effect. Inspection of the interaction plot showed that the interaction was driven by a greater mismatch for /t/ when /t/ was the deviant in the first block, than the mismatch advantage for /d/ when /d/ was the deviant in the first block. We next analyzed the voltage data, by averaging the electrodes with TF5SF1 factor loadings greater than 0.6, and time samples with factor loadings greater than 0.6 (108-152ms). The ANOVA showed the same pattern: a significant intercept (F(1,27)=10.1, p<.001), and a PHONEME x BLOCKORDER interaction (F(1,27)=14.1, p<.001). Orthogonal contrast analysis showed that /d/ as deviant in the first block was significant (-0.7mV, t=-3.5, p=.001), and that /t/ as deviant in the first block exhibited an even greater mismatch (-0.9mV, t=-4.22, p=.0002). In summary, the mismatch effect was symmetrical in this ERP component in the sense that both /d/ and /t/ exhibited MMN. Moreover, the effect was significantly greater for /t/ (which was not predicted by our theory but is not inconsistent with it).

4.2.2 Late Discriminatory Negativity (TF1SF1)

As in the other experiments, the main temporal factor in this experiment was a late, slow negativity, expressed by TF1. There were two spatial subfactors of interest: TF1SF1 which contained a mismatch effect for /t/ with a central distribution, and TF1SF3 which contained a mismatch effect for /d/ with a slightly more anterior distribution. We note that this observation illustrates the advantage of using temporo-spatial PCA: The fact that two different phonemes may have slightly different spatial distribution of their MMN response could easily have been overlooked or missed by an analysis that tries to “squeeze” the both MMNs into a single spatial region. We analyze TF1SF1 first:

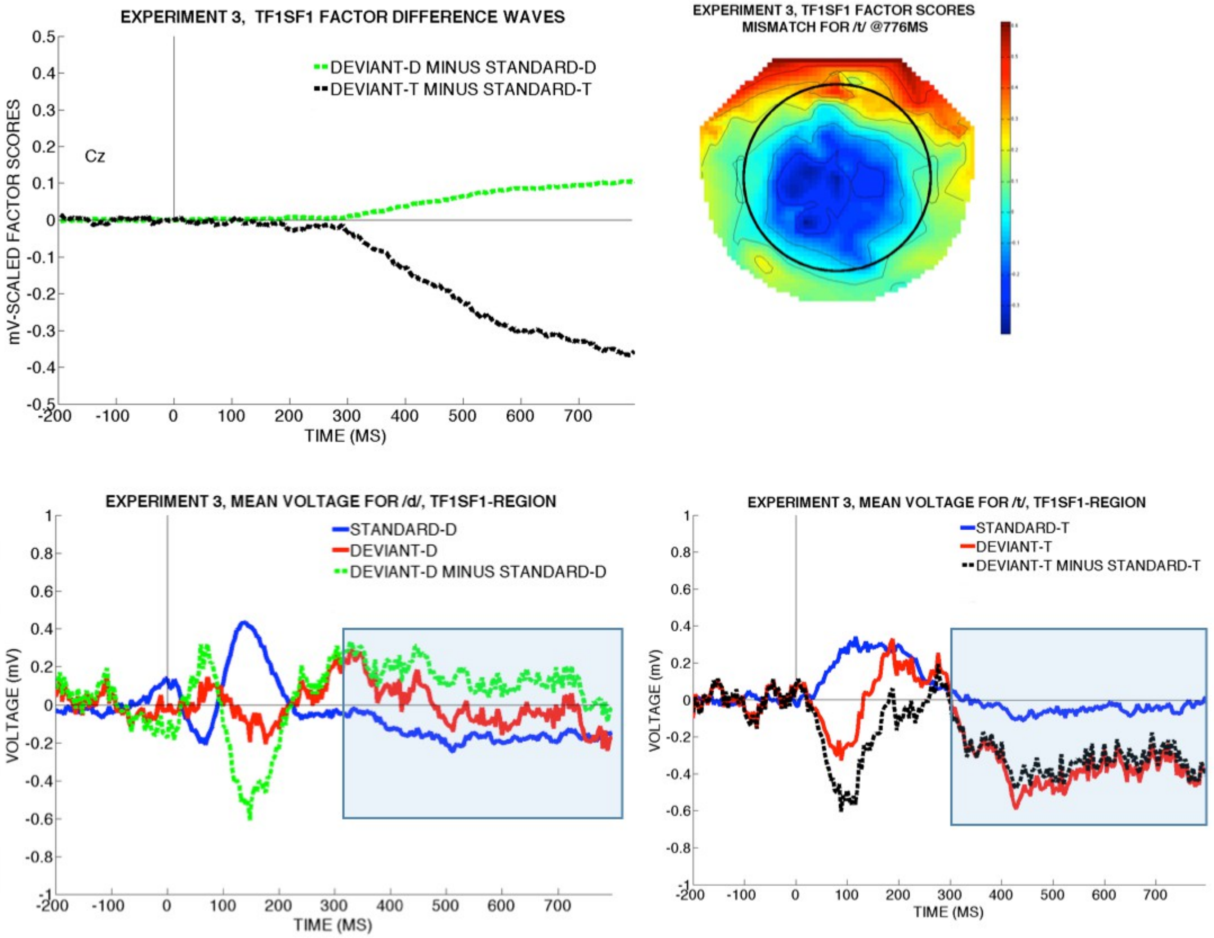


Figure 8: Experiment 3, Late Discriminatory Negativity for /t/. Upper panel: Time course (left) and spatial distribution (right) of temporo-spatial factor decomposition; the topoplot shows the mean difference wave at the horizontal line in the waveform plot at 776ms. Lower panel: corresponding absolute voltage waves and difference waves in the raw grand average voltage for /d/ (left panel) and /t/ (right panel).

An ANOVA of the factor scores resulted in a significant PHONEME x BLOCKORDER interaction (F(1,27)=10.5, p=.003). Inspection of the interaction plot revealed that the interaction was due to the fact that the mismatch effect for /t/ when /t/ was the first deviant was greater than the mismatch effect for /d/ when /d/ was the first deviant. Orthogonal contrast analysis showed that the mismatch was significant for /t/ when it was the deviant in the first block (0.32mV, t=-3.6, p<.01), whereas the mismatch effect for /d/ was not significant when it was the deviant in the first block (0.15mV, t=-1.35, p=.18). We next analyzed the raw voltage data, restricted to the electrodes of TF1SF1 with factor loadings greater than .6 (roughly the blue box of the topoplot in Figure 8), averaged for the time samples with factor loadings greater than 0.6 (476-800ms). Again, this resulted in the same PHONEME x BLOCKORDER interaction (F(1,27)=12.8, p<.01). Orthogonal contrast analysis of /t/ when it was the first deviant was highly significant (-0.79mV, t=-4.21, p<.001); again the contrast for /d/ was not significant (-0.38mV, t=-1.39, p=.17). The factor score and voltage analysis converged sharply.

The second spatial subfactor had a more anterior distribution and is shown in Figure 9:

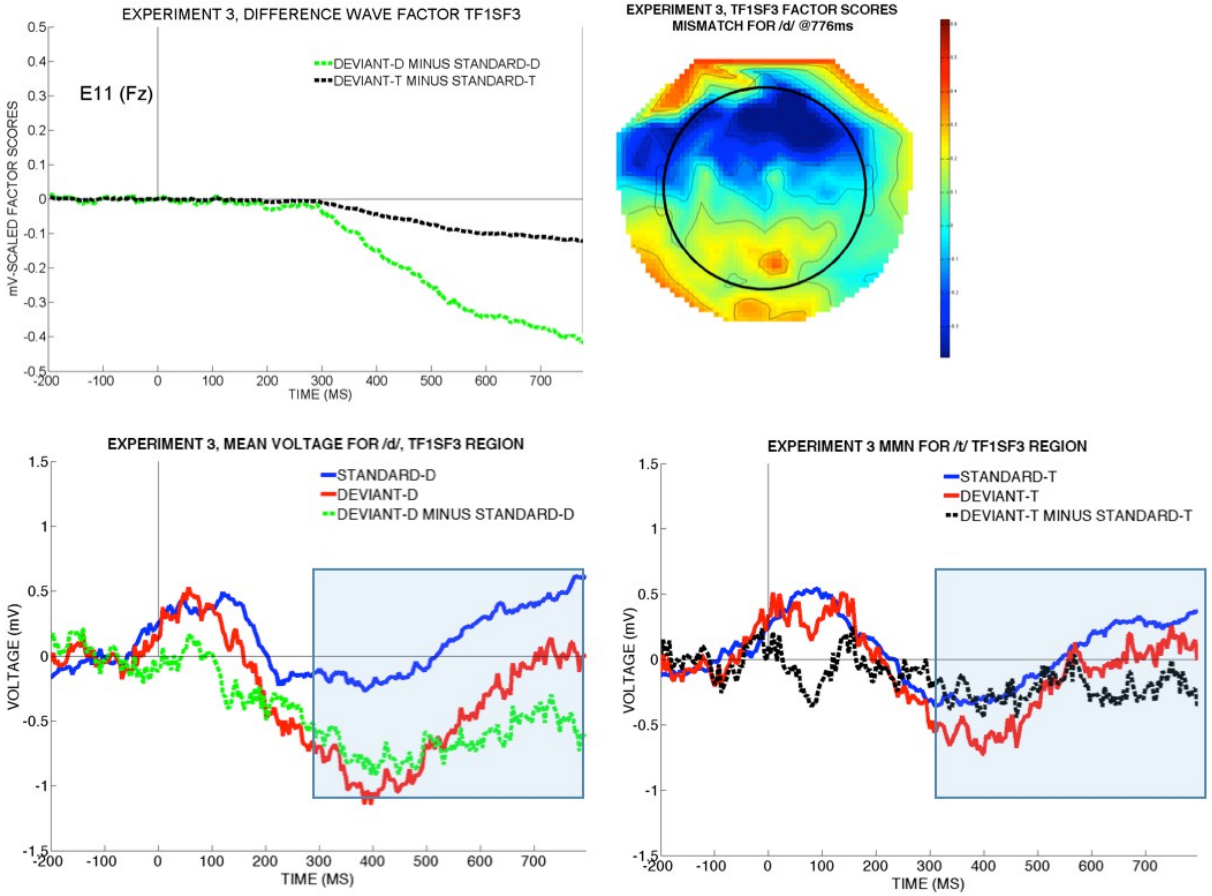


Figure 9: Experiment 3, Late Discriminatory Negativity for /d/. Upper panel: Time course (left) and spatial distribution (right) of temporo-spatial factor decomposition; the topoplot shows the mean difference wave at the horizontal line in the waveform plot at 776ms. Lower panel: corresponding absolute voltage waves and difference waves in the raw grand average voltage for /d/ (left panel) and /t/ (right panel).

An ANOVA of the factor scores resulted in a significant intercept (F(1,27)=11.37, p<.01), meaning there was a main effect of standard vs. deviant (i.e., the main effect difference wave was significantly below zero); as well as a significant PHONEME x BLOCKORDER interaction. Inspection of the interaction plot revealed the difference wave for /d/ was significantly more negative when it was presented as deviant in the first block than the difference wave for /t/ when /t/ was the first deviant. Orthogonal contrast analysis for /d/ as first deviant was highly significant (-0.32mV, t=-3.5, p=.001); the contrast for /t/ was also significant (-0.19mV, t=-2.2, p=.03). We then tested the same contrasts in the raw voltage data, averaging the electrodes in TF1SF3 with factor loadings greater than 0.6 (roughly the blue box in Figure 9), and time samples for TF1 with factor loadings greater than 0.6 (i.e. 476-800ms). An ANOVA of the raw voltage resulted in a highly significant intercept only (F(1,27)=24.5, p<.0001), no other effects were significant. In other words, a mismatch was present for both /d/ and /t/ in this region and time window in the raw voltage data.

Note that whereas the LDN here exhibits the typical “first-mention” interaction observed in this study, it is important to note that in this case, there is still a significant mismatch for /t/, allowing us to conclude that a mismatch was present for both phonemes, both in the MMN (where the effect was symmetrical) and the LDN (where the effect was asymmetrical but with MMN for both phonemes).

4.3 Discussion

Experiment 3 exhibited a number of experimental effects; the most striking of which is the emergence of mismatch effects for /t/ that were absent in Experiments 1-2. First, in the classical early MMN response, a mismatch was present for both phonemes, whereas in the phoneme-sensitive Experiment 1 and 2, only /d/ resulted in a mismatch. Secondly, the Late Discriminatory Negativity exhibited a richer set of results than in Experiment 1 and 2: First, the temporo-spatial PCA revealed that the LDN mismatch for /t/ had a slightly different distribution than the LDN mismatch for /d/. In addition, the spatial component for the LDN effect for /d/ also contained a mismatch effect for /t/. In sum, a mismatch effect for /t/ emerged in this experiment, where the memory trace was likely to have been formed from a phonetic representation. This predicted a symmetrical MMN response across the two phonemes, which is borne out by the data. This result reinforces the inference that the asymmetry in MMN observed in the phoneme-memory trace paradigm is due to abstract differences in the feature matrices, as predicted by underspecification theory.

While we argue that Experiment 3 is tapping into phonetic representations, a reviewer asks how we know it is phonetic rather than general auditory representations. There is abundant evidence that speech sounds are treated differently from nonspeech sounds [63]–[65]. Such results typically show that speech vs. nonspeech engage different cortical regions, where the former are associated with the left-auditory cortex, while the latter do not show a similar left-lateralization. Given this automatic treatment of speech sounds as different from non-speech sounds, it seems most likely to expect that the standard sequence in Experiment 3 results in a language-specific phonetic memory trace, rather than a general auditory representation. Note also that our experimental logic assumes that the oddball stimulus in all three experiments is represented as a phonetic category rather than a set of general auditory features. It is beyond the scope of the current study to submit this assumption to further scrutiny.

The results of Experiment 3 also allow us to address another possible explanation for the asymmetry observed in Experiments 1 and 2: Could an intrinsic phonetic contrast between [d] and [t] give rise to the asymmetry? For example, as shown in work by Polka and colleagues [66], [67], vowels with greater formant frequency convergence are perceptually more salient, which could plausibly lead to MMN asymmetries at the level of auditory encoding, and not related to underspecification. Related to this, a reviewer asks, are there any circumstances which would give rise to an asymmetric MMN response with a non-varying, static standard? We believe the cases studied by Polka and colleagues would likely be such a case.

More relevant to the current study is work by Steinschneider and others [68]–[72] on "single-on" vs. "double-on" response pattern to VOT differences in single cell intracranial recordings of monkeys. Consider first the fact that the shorter VOT of voiced stops results in a sense from a single articulatory event and therefore a more singular acoustic event (i.e. due to the very short VOT, the two discrete acoustic events might amalgate into one from the perspective of sensory neurons. The longer VOT of voiceless stops can be more readily parsed into two acoustic events. Steinschneider's discovery was that this difference indeed corresponded to two different response patterns in individual auditory cortex neurons, recorded in monkeys. Shorter VOTs lead to a single spike response in individual auditory cortex neurons, whereas longer VOTs lead to two temporally separate spike responses, where the latency between the spikes are strongly related to the VOT value in the stimuli. This specific neural sensitivity to differences in temporal information could be a possible non-linguistic source of scalp-recorded MMN asymmetries. For example, Sharma and Dorman [33] argued that two different N1 responses corresponded to these two spike patterns.

We addressed this question by asking whether intrinsic differences in the P2 response, which could be a manifestation of the single-on vs. double-on effect, could plausibly account for the observed differences in our scalp recorded MMN (perhaps because of the temporally more homogeneous neural response to short VOTs). We selected the P2 because it was the most prominent part of our observed AEPs, and because amplitude difference in the P2 region was the central contributor of the MMN; for this reason we did not analyze latency shifts in the N1 (see Sharma and Dorman [33] for parallel N1 observations). Based on pooling data from all subjects who participated in the three experiments, we found that the P2 peak latency from all standards differed significantly as a function of mean VOT. As illustrated in Figure 10, the grand mean difference between the mean VOTs for /d/ vs. /t/ was 20ms (as a function of the experimental design), and the grand mean peak latency of the P2 (measured at Cz) differed by 16ms.

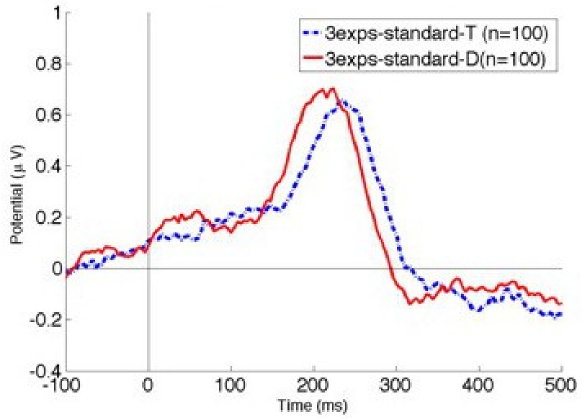


Figure 10: Latency difference (but not significant amplitude difference) between the mean Auditory Evoked Potential to /d/ vs. /t/ (averaged across all tokens and all subjects in all three experiments).

This close correlation has been observed elsewhere in the literature [73]. However, this latency difference is precisely what is controlled for by contrasting each phoneme as a standard with itself as a deviant. The critical question is whether the amplitude differs (which could result in MMN asymmetry), but the small observed difference in peak amplitude between /d/ and /t/ was highly non-significant (F(1,99)=0.05), p=.8). We conclude that a single-on vs. double-on difference does not appear to be involved in our data as a confounding factor in affecting the basic AEP. In addition (as pointed out by a reviewer), there is no homogeneous acoustic/phonetic common factor that unifies the diverse set of underspecification related asymmetries observed in the literature, such as place features ([coronal] vs [labial/dorsal] in vowels[17]; [coronal] vs. [dorsal] stops[74]), laryngeal features[34], and manner of articulation features[75].

3. CONCLUSIONS

Though there is need for abstraction in long-term phonological representations (phonemes) compared to acoustic representations, the exact nature of the abstraction has been unclear. Recent work has revealed evidence in favor of underspecified phonological representations [17], [37], [74], [75]. In this article we extend this line of enquiry to provide evidence for those theories from how the brain represents phonemes.

More particularly, phonological analyses of voiced and voiceless stops in American English argue that voiceless stops are specified for voicing/laryngeal features while voiced stops are phonemically underspecified for them; i.e., the two series of sounds are asymmetrically represented at an abstract phonemic level. In this article, we tested if listeners recruit such knowledge of abstract and underspecified phonological representations during speech perception. More specifically, we showed through a series of three EEG experiments that, consistent with the representational claims of voiced and voiceless stops in American English from phonological analyses, the listeners exhibited asymmetric MMN responses to voiced and voiceless stops. In Experiment 1, we employed a "varying" standard MMN experiment on American English listeners to probe phonological representations, with a distractor task of tracking the presentation the syllable "ba" randomly interspersed with the standards and deviants. The listeners exhibited a larger MMN to deviant voiced stops in the context of voiceless stops as standards than to deviant voiceless stops in the context of voiced stops as standards. In Experiment 2, we examined whether the asymmetry only appeared under attention to the auditory stream, or whether it had the signature of automatic change detection mechanisms. Experiment 2 revealed that even in a passive listening task, there is again an asymmetry with respect to deviant voiced and voiceless stops. As in Experiment 1, listeners exhibited a larger MMN to deviant voiced stops in the context of voiceless stops as standards than to deviant voiceless stops in the context of voiced stops as standards. Finally, in Experiment 3, we showed that the results obtained in Experiments 1 and 2 were unlikely to be due to intrinsic asymmetries in the phonetics of voiced and voiceless stops. Experiment 3 employed a traditional non-varying standards MMN paradigm, which targets phonetic representations. Now, listeners no longer showed the asymmetry observed in Experiments 1 and 2; thereby suggesting that the asymmetries in MMN responses observed in Experiments 1 and 2 are unlikely to be due to asymmetries in phonetic or auditory processing. We conclude that the experiments show that underspecification finds support in the differential MMN responses reported here, under the assumptions of the Eulitz/Lahiri experimental logic, and the assumption that voicing features are underspecified in English “voiced” consonants.

A crucial feature of our argument is contrasting the effect of single tokens in Experiment 3. A reviewer asks if there can be MMN asymmetries with single token standards as well, and is varying standards necessary to get a phonemic MMN. The answer to the latter is “yes”, but the answer to the former is a non-contradictory “no”. As a point of logic (cf. the rules of material implication), Phillips et al.’s conclusion that varying standards is necessary to measure a phonemic MMN does <u>not</u> mean that asymmetric MMNs can<u>not</u> be observed with single token standard experiments in general. In fact, intrinsic physical differences between stimuli typically leads to different MMNs, as illustrated by the simple fact that degree of frequency change corresponds to degree of MMN amplitude [19], Figure 1. For phonetic contrasts, we would similarly predict that differences in frequency convergence in vowels [66] would give rise to asymmetric MMN with single token standards—but this would be orthogonal to the question of underspecification related MMN asymmetries.

Author Note

We acknowledge the generous support of this research by a University of Delaware UNIDEL grant to the first author. We also acknowledge the participation of Catherine Bradley and Evan Bradley for their crucial assistance with data collection and intellectual contributions to this project.

Figure Captions

Figure 1: MMNs as per FUL. Ø = unspecified. Arrows indicate which pairs are compared by the perceptual mechanisms. Dotted arrow represents the “no mismatch” comparison.

Figure 2: Experiment 1, early MMN effect. Upper panel: Time course (left) and spatial distribution (right) of temporo-spatial factor decomposition; the topoplot shows the mean difference wave at the horizontal line in the waveform plot at 216ms. Lower panel: corresponding absolute voltage waves and difference waves in the raw grand average voltage for /d/ (left panel) and /t/ (right panel).

Figure 3: Interaction plot PHONEME (difference wave) x BLOCKORDER. Vertical bars denote +/-standard error.

Figure 4: Experiment 1, Late Discriminatory Negativity. Upper panel: Time course (left) and spatial distribution (right) of temporo-spatial factor decomposition; the topoplot shows the mean difference wave at the horizontal line in the waveform plot at 652ms. Lower panel: corresponding absolute voltage waves and difference waves in the raw grand average voltage for /d/ (left panel) and /t/ (right panel).

Figure 5: Experiment 2 MMN. Upper panel: Time course (left) and spatial distribution (right) of temporo-spatial factor decomposition; the topoplot shows the mean difference wave at the horizontal line in the waveform plot at 208ms. Lower panel: corresponding absolute voltage waves and difference waves in the raw grand average voltage for /d/ (left panel) and /t/ (right

panel).

Figure 6: Experiment 2 Late Discriminatory Negativity. Upper panel: Time course (left) and spatial distribution (right) of temporo-spatial factor decomposition; the topoplot shows the mean difference wave at the horizontal line in the waveform plot at 776ms. Lower panel: corresponding absolute voltage waves and difference waves in the raw grand average voltage for /d/ (left panel) and /t/ (right panel).

Figure 7: Experiment 3, phonetic MMN. Upper panel: Time course (left) and spatial distribution (right) of temporo-spatial factor decomposition; the topoplot shows the mean difference wave at the horizontal line in the waveform plot at 132ms. Lower panel: corresponding absolute voltage waves and difference waves in the raw grand average voltage for /d/ (left panel) and /t/ (right panel). Both phonemes show MMN difference waves.

Figure 8: Experiment 3, Late Discriminatory Negativity for /t/. Upper panel: Time course (left) and spatial distribution (right) of temporo-spatial factor decomposition; the topoplot shows the mean difference wave at the horizontal line in the waveform plot at 776ms. Lower panel: corresponding absolute voltage waves and difference waves in the raw grand average voltage for /d/ (left panel) and /t/ (right panel).

Figure 9: Experiment 3, Late Discriminatory Negativity for /d/. Upper panel: Time course (left) and spatial distribution (right) of temporo-spatial factor decomposition; the topoplot shows the mean difference wave at the horizontal line in the waveform plot at 776ms. Lower panel:

corresponding absolute voltage waves and difference waves in the raw grand average voltage for /d/ (left panel) and /t/ (right panel).

Figure 10: Latency difference (but not significant amplitude difference) between the mean Auditory Evoked Potential to /d/ vs. /t/ (averaged across all tokens and all subjects in all three experiments).

References

[1] N. Chomsky and M. Halle, *The Sound Pattern of English*. New York,: Harper & Row, 1968.

[2] M. Halle, *The Sound Pattern of Russian*. The Hague: Mouton, 1959.

[3] D. C. Hall, *The role and representation of contrast in phonological theory*. Ph. D. Dissertation, University of Toronto, 2007.

[4] D. Steriade, "Redundant values," in *23rd CLS, Part II: Parasession on Autosegmental and Metrical Phonology*, 1987.

[5] D. Archangeli, "Aspects of underspecification theory," *Phonology*, vol. 5, no. 02. p. 183, 1988.

[6] M. Ghini, *Asymmetries in the Phonology of Miogliola*. Berlin: Mouton, 2001.

[7] P. Kiparsky, "Blocking in nonderived environments.," in *Studies in Lexical Phonology*, vol. 4, S. Hargus and E. M. Kaisse, Eds. New York: Academic Press, 1993, pp. 277–314.

[8] P. Kiparsky, "Lexical morphology and phonology," in *Linguistics in the morning calm*, I.-S. Yang, Ed. Seoul: Hanshin, 1982, pp. 3–91.

[9] J. M. Anderson and C. J. Ewen, *Principles of Dependency Phonology*. Cambridge: Cambridge University Press, 1987.

[10] J. Kaye, J. Lowenstamm, and J.-R. Vergnaud, "The Internal Structure of Phonological Elements: A Theory of Charm and Government," *Phonol. Yearb.*, vol. 2, pp. 305–328, 1985.

[11] T. Borowsky, "Antigemination in English Phonology," *Linguist. Inq.*, vol. 18, pp. 671–678, 1987.

[12] J. J. McCarthy and A. Taub, "Review of Paradis & Prunet 1991," *Phonology*, vol. 9.2, pp. 363–370, 1992.

[13] D. Steriade, "Underspecification and markedness," in *The Handbook of Phonological Theory*, J. Goldsmith, Ed. Blackwell, 1995, pp. 114–175.

[14] G. N. Clements, "Representational economy in constraint-based phonology," in *Distinctive Feature Theory*, T. A. Hall, Ed. Berlin: Mouton, 2001, pp. 71–146.

[15] P. Shaw, "Consonant harmony systems: The special status of coronal harmony," in *The special status of coronals*, C. Paradis and J.-F. Prunet, Eds. San Diego: Academic Press,

1991, pp. 127–157.

[16] E. Dresher, *The Contrastive Hierarchy in Phonology*. Cambridge: Cambridge University Press, 2009.

[17] C. Eulitz and A. Lahiri, “Neurobiological Evidence for Abstract Phonological Representations in the Mental Lexicon during Speech Recognition,” *J. Cogn. Neurosci.*, vol. 16, no. 4, pp. 577–583, 2004.

[18] A. Lahiri and H. Reetz, “Underspecified recognition,” in *Laboratory Phonology 7*, 2002, pp. 637–675.

[19] R. Näätänen, P. Paavilainen, T. Rinne, and K. Alho, “The mismatch negativity (MMN) in basic research of central auditory processing: A review,” *Clinical Neurophysiology*, vol. 118, no. 12. pp. 2544–2590, 2007.

[20] R. Näätänen, M. Tervaniemi, E. Sussman, P. Paavilainen, and I. Winkler, “‘Primitive intelligence’ in the auditory cortex,” *Trends in Neurosciences*, vol. 24, no. 5. pp. 283–288, 2001.

[21] M. Sallinen, J. Kaartinen, and H. Lyytinen, “Is the appearance of mismatch negativity during stage 2 sleep related to the elicitation of K-complex?,” *Electroencephalogr Clin Neurophysiol*, vol. 91, pp. 140–148, 1994.

[22] N. M. Kane, S. H. Curry, S. R. Butler, and C. B.H., “Electrophysiological indicator for awakening from coma.,” *Lancet*, vol. 341, p. 688, 1993.

[23] R. Näätänen, A. Lehtokoski, M. Lennes, M. Cheour, M. Huotilainen, A. Iivonen, M. Vainio, P. Alku, R. J. Ilmoniemi, A. Luuk, J. Allik, J. Sinkkonen, and K. Alho, “Language-specific phoneme representations revealed by electric and magnetic brain responses,” *Nature*, vol. 385, no. 6615, pp. 432–434, 1997.

[24] R. Näätänen, “[The mismatch negativity as an index of the perception of speech sounds by the human brain],” *Ross Fiziol Zh Im I M Sechenova*, vol. 86, no. 11, pp. 1481–1501, 2000.

[25] R. Naatanen and K. Alho, “Mismatch negativity--a unique measure of sensory processing in audition,” *Int J Neurosci*, vol. 80, no. 1–4, pp. 317–337, 1995.

[26] R. Näätänen, “The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm),” *Psychophysiology*, vol. 38, no. 1, pp. 1–21, 2001.

[27] R. Naatanen, “The mismatch negativity as an index of the perception of speech sounds by the human brain,” *Ross. Fiziol. zhurnal Im. I.M.Sechenova / Ross. Akad. Nauk*, vol. 86, no. 11, pp. 1481–1501, 2000.

[28] I. Winkler, A. Lehtokoski, P. Alku, M. Vainio, I. Czigler, O. Aaltonen, I. Raimo, K. Alho, H. Lang, A. Livonen, and R. Naatanen, "Pre-attentive detection of vowel contrasts utilizes both phonetic and auditory memory representations," *Cogn. Brain Res.*, vol. 7, pp. 357–369, 1999.

[29] M. Cheour, R. Ceponiene, A. Lehtokoski, A. Luuk, J. Allik, K. Alho, and NaatanenRisto, "Development of language-specific phoneme representations in the infant brain," *Nat. Neurosci.*, vol. 1, no. 5, pp. 351–353, 1998.

[30] H. Datta, A. Hestvik, N. Vidal, M. Hisagi, C. Tessel, and M. Wroblewsky, "Brain indices of speech discrimination in adults," *The XI Congress of the International Association for the Study of Child Language (IASCL 2008)*. Edinburgh, UK, 2008.

[31] A. Shestakova, E. Brattico, M. Huotilainen, V. Galunov, A. Soloviev, M. Sams, R. Ilmoniemi, and R. Naatanen, "Abstract phoneme representations in the left temporal cortex: magnetic mismatch negativity study," *Neuroreport*, vol. 13, no. 14, pp. 1813–1816, 2002.

[32] S. Ylinen, "Cortical Representation for Phonological Quantity," University of Helsinki, Helsinki, 2006.

[33] A. Sharma and M. F. Dorman, "Cortical auditory evoked potential correlates of categorical perception of voice-onset time," *J. Acoust. Soc. Am.*, vol. 106, no. 2, pp. 1078–1083, 1999.

[34] C. Phillips, T. Pellathy, A. Marantz, E. Yellin, K. Wexler, D. Poeppel, M. McGinnis, and T. Roberts, "Auditory Cortex Accesses Phonological Categories: An MEG Mismatch Study," *J. Cogn. Neurosci.*, vol. 12, no. 6, pp. 1038–1055, 2000.

[35] V. Felder and A. Lahiri, "Examining featural underspecification of 'TONGUE HEIGHT' in German mid vowels: An EEG study." Retrieved from internal-pdf://DiplomarbeitFelder2006-3690371073/DiplomarbeitFelder2006.pdf, University of Konstanz, Psychology, Konstanz., 2006.

[36] J. Obleser, A. Lahiri, and C. Eulitz, "Magnetic Brain Response Mirrors Extraction of Phonological Features from Spoken Vowels," *J. Cogn. Neurosci.*, vol. 16, no. 1, pp. 31–39, 2004.

[37] S. A. Cornell, A. Lahiri, and C. Eulitz, "'What you encode is not necessarily what you store': Evidence for sparse feature representations from mismatch negativity," *Brain Res.*, vol. 1394, no. 0, pp. 79–89, 2011.

[38] M. A. Walter and V. Hacquard, "MEG Evidence for Phonological Underspecification," 2004.

[39] P. Avery and W. J. Idsardi, “Laryngeal Dimensions, Completion and Enhancement,” in *Distinctive Feature Theory*, T. A. Hall, Ed. de Gruyter, Berlin, 2001, pp. 41–70.

[40] G. K. Iverson and J. C. Salmons, “Aspiration and laryngeal representation in Germanic,” *Phonology*, vol. 12, pp. 369–396, 1995.

[41] M. Jessen and C. Ringen, “Laryngeal features in German,” *Phonology*, vol. 19, no. 2, pp. 189–218, 2002.

[42] G. K. Iverson and J. C. Salmons, “Laryngeal enhancement in Germanic,” *Phonology*, vol. 20, pp. 43–74.

[43] S.-O. Hwang, P. J. Monahan, and W. J. Idsardi, “Underspecification and asymmetries in voicing perception,” *Phonology*, vol. 27, no. 2, pp. 205–224, 2010.

[44] H. T. Bunnell, “Klatt Synthesis Interface.” Speech Research Lab, A.I. DuPont Hospital for Children and the University of Delaware, Wilmington, 1999.

[45] D. H. Klatt, “Software for a cascade/parallel formant synthesizer,” *J. Acoust. Soc. Am.*, vol. 67, pp. 971–995, 1980.

[46] T. C. Ferree, P. Luu, G. S. Russell, and D. M. Tucker, “Scalp electrode impedance, infection risk, and EEG data quality,” *Clin. Neurophysiol.*, vol. 112, no. 3, pp. 536–544, 2001.

[47] J. Dien, “The ERP PCA Toolkit: An open source program for advanced statistical analysis of event-related potential data,” *J. Neurosci. Methods*, vol. 187, no. 1, pp. 138–145, 2010.

[48] J. Dien, “Addressing misallocation of variance in principal components analysis of event-related potentials,” *Brain Topogr.*, vol. 11, no. 1, pp. 43–55, 1998.

[49] P. L. Nunez and R. Srinivasan, “A theoretical basis for standing and traveling brain waves measured with human EEG with implications for an integrated consciousness,” *Clin. Neurophysiol.*, vol. 117, no. 11, pp. 2424–2435, 2006.

[50] K. M. Spencer, J. Dien, and E. Donchin, “Spatiotemporal analysis of the late ERP responses to deviant stimuli,” *Psychophysiology*, vol. 38, no. 2, pp. 343–358, 2001.

[51] J. Dien, W. Khoe, and G. R. Mangun, “Evaluation of PCA and ICA of simulated ERPs: Promax vs. infomax rotations,” *Hum. Brain Mapp.*, vol. 28, no. 8, pp. 742–763, 2007.

[52] J. Dien and G. A. Frishkoff, “Principal components analysis of ERP data,” in *Event-Related Potentials: A Methods Handbook*, T. Handy, Ed. Cambridge: MIT Press, 2005.

[53] K. M. Spencer, J. Dien, and E. Donchin, “A componential analysis of the {ERP} elicited by novel events using a dense electrode array,” *Psychophysiology*, vol. 36, no. 3, pp. 409–414, 1999.

[54] J. Dien, “Applying Principal Components Analysis to Event-Related Potentials: A Tutorial,” *Dev. Neuropsychol.*, vol. 37, no. 6, pp. 497–517, 2012.

[55] J. Dien, “Evaluating two-step PCA of ERP data with Geomin, Infomax, Oblimin, Promax, and Varimax rotations.,” *Psychophysiology*, vol. 47, no. 1, pp. 170–83, Jan. 2010.

[56] A. Delorme and S. Makeig, “EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis.,” *J. Neurosci. Methods*, vol. 134, no. 1, pp. 9–21, Mar. 2004.

[57] D. Mullens, I. Winkler, A. Heathcote, L. Whitson, A. Provost, and J. Todd, “Can order change modulation of response to standard and deviant tones?,” in *The 7th Mismatch Negativity Conference*, 2015.

[58] R. Čeponien, M. Cheour, and R. Näätänen, “Interstimulus interval and auditory event-related potentials in children: evidence for multiple generators,” *Electroencephalogr. Clin. Neurophysiol. Potentials Sect.*, vol. 108, no. 4, pp. 345–354, 1998.

[59] H. Datta, V. L. Shafer, M. L. Morr, D. Kurtzberg, and R. G. Schwartz, “Electrophysiological indices of discrimination of long-duration, phonetically similar vowels in children with typical and atypical language development.(Report),” *J. Speech, Lang. Hear. Res.*, vol. 53, no. 3, p. 757(21).

[60] P. Korpilahti, H. Lang, and O. Aaltonen, “Is there a late-latency mismatch negativity (MMN) component?,” *Electroencephalography and Clinical Neurophysiology*, vol. 95, no. 4. pp. P96–P96, 1995.

[61] V. L. Shafer, M. L. Morr, H. Datta, D. Kurtzberg, and R. G. Schwartz, “Neurophysiological Indexes of Speech Processing Deficits in Children with Specific Language Impairment,” *J. Cogn. Neurosci.*, vol. 17, no. 7, pp. 1168–1180, 2005.

[62] A. Shestakova, M. Huotilainen, R. Ceponiene, and M. Cheour, “Event-related potentials associated with second language learning in children,” *Clin. Neurophysiol.*, vol. 114, no. 8, pp. 1507–1512, 2003.

[63] A. Vouloumanos, K. A. Kiehl, J. F. Werker, and P. F. Liddle, “Detection of sounds in the auditory stream: event-related fMRI evidence for differential activation to speech and nonspeech,” *J. Cogn. Neurosci.*, vol. 13, pp. 994–1005, 2001.

[64] G. Dehaene-Lambertz, C. Pallier, W. Serniclaes, L. Sprenger-Charolles, A. Jobert, and S. Dehaene, “Neural correlates of switching from auditory to speech perception,” *Neuroimage*, vol. 24, pp. 21–33, 2005.

[65] R. Möttönen, G. A. Calvert, I. P. Jääskeläinen, P. M. Matthews, T. Thesen, J. Tuomainen, and M. Sams, “Perceiving identical sounds as speech or non-speech modulates activity in

the left posterior superior temporal sulcus," *Neuroimage*, vol. 30, pp. 563–569, 2006.

[66] M. Masapollo and L. Polka, "Asymmetries in vowel perception: Do they arise from focalization, perceptual magnets, or both?," *J. Acoust. Soc. Am.*, vol. 135, no. 4, pp. 2224–2224, Apr. 2014.

[67] L. Polka and O. S. Bohn, "Natural Referent Vowel (NRV) framework: An emerging view of early phonetic development," *J. Phon.*, vol. 39, no. 4, pp. 467–478, 2011.

[68] M. Steinschneider, C. E. Schroeder, J. C. Arezzo, and H. G. Vaughan, "Physiologic Correlates of the Voice Onset Time Boundary in Primary Auditory Cortex (A1) of the Awake Monkey: Temporal Response Patterns," *Brain Lang.*, vol. 48, no. 3, pp. 326–340, 1995.

[69] M. Steinschneider, K. V Nourski, and Y. I. Fishman, "Representation of speech in human auditory cortex: is it special?," *Hear. Res.*, vol. 305, pp. 57–73, Nov. 2013.

[70] M. Steinschneider, K. V Nourski, H. Kawasaki, H. Oya, J. F. Brugge, and M. A. Howard, "Intracranial study of speech-elicited activity on the human posterolateral superior temporal gyrus.," *Cereb. Cortex*, vol. 21, no. 10, pp. 2332–47, Oct. 2011.

[71] M. Steinschneider, Y. I. Fishman, and J. C. Arezzo, "Representation of the voice onset time (VOT) speech parameter in population responses within primary auditory cortex of the awake monkey.," *J. Acoust. Soc. Am.*, vol. 114, no. 1, pp. 307–21, Jul. 2003.

[72] M. Steinschneider, I. O. Volkov, Y. I. Fishman, H. Oya, J. C. Arezzo, and M. A. Howard, "Intracortical responses in human and monkey primary auditory cortex support a temporal processing mechanism for encoding of the voice onset time phonetic parameter.," *Cereb. Cortex*, vol. 15, no. 2, pp. 170–86, Feb. 2005.

[73] K. A. King, J. Campbell, A. Sharma, K. Martin, M. Dorman, and J. Langran, "The representation of voice onset time in the cortical auditory evoked potentials of young children.," *Clin. Neurophysiol.*, vol. 119, no. 12, pp. 2855–61, Dec. 2008.

[74] M. Scharinger, A. Bendixen, N. J. Trujillo-Barreto, and J. Obleser, "A Sparse Neural Code for Some Speech Sounds but Not for Others," *PLoS One*, vol. 7, no. 7, p. e40953, Jul. 2012.

[75] S. a Cornell, A. Lahiri, and C. Eulitz, "Inequality across consonantal contrasts in speech perception: evidence from mismatch negativity.," *J. Exp. Psychol. Hum. Percept. Perform.*, vol. 39, no. 3, pp. 757–72, 2013.

[76] E. J. Flagg, J. E. Oram Cardy, and T. P. L. Roberts, "MEG detects neural consequences of anomalous nasalization in vowel-consonant pairs," *Neurosci. Lett.*, vol. 397, no. 3, pp. 263–268, 2006.

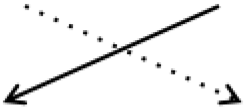
[77] P. A. Keating, “Underspecification in Phonetics,” *Phonology*, vol. 5, no. 02. p. 275, 1988.

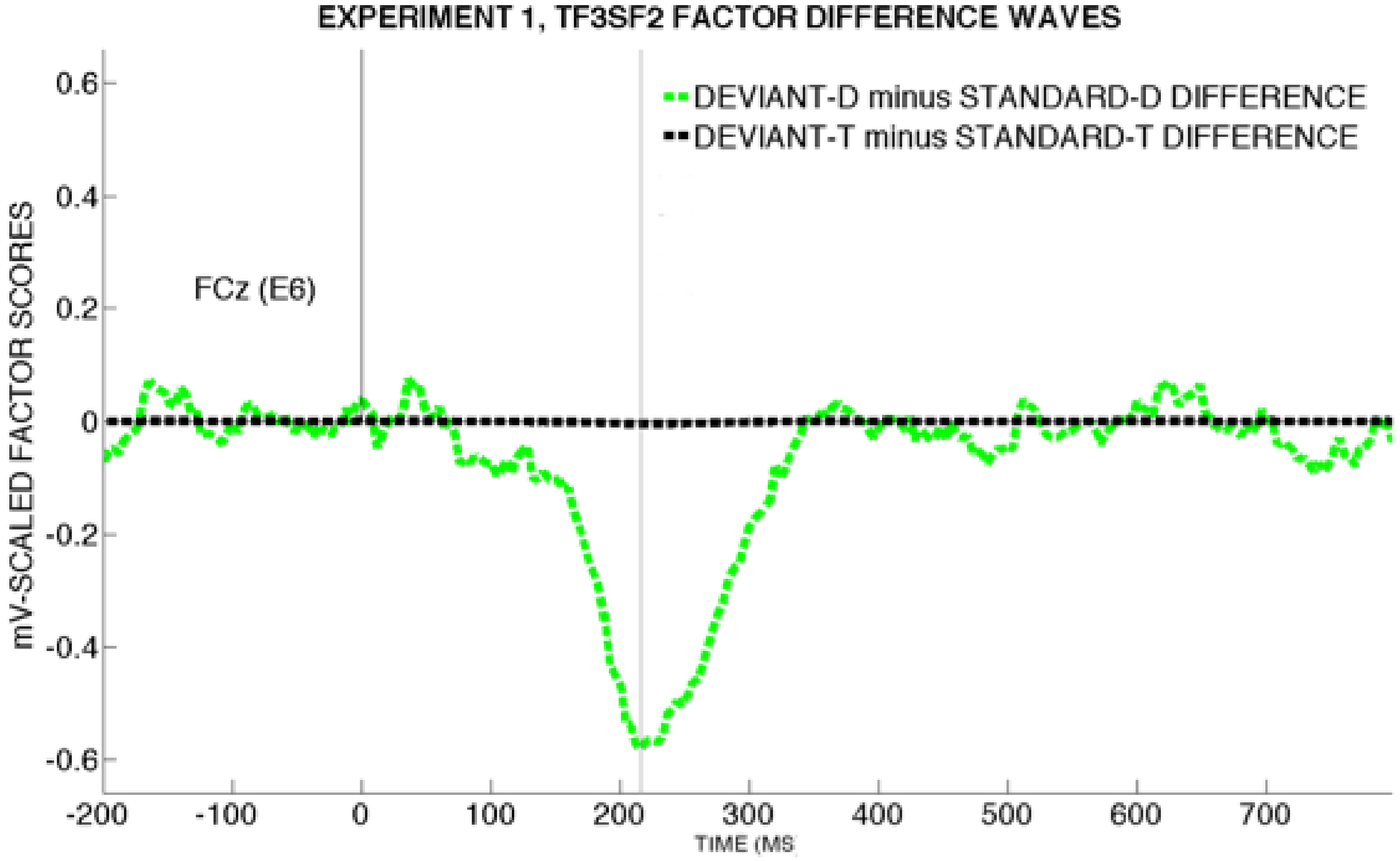
[78] A. C. Cohn, “Phonetic and Phonological Rules of Nasalization,” *Univ. Calif. Work. Pap. Phonetics*, vol. 76, pp. 1–224, 1990.

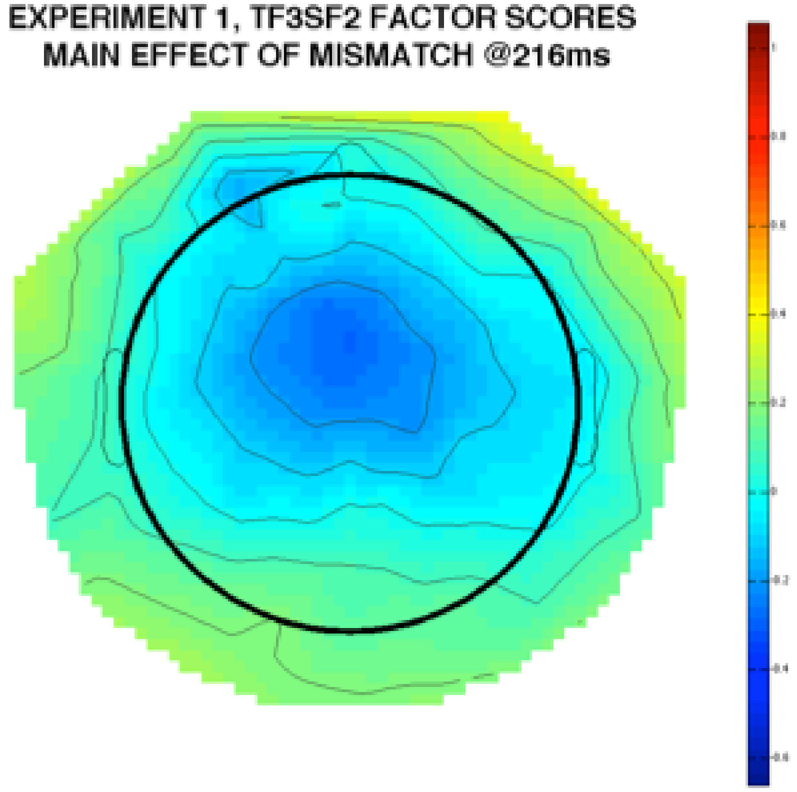
		BLOCK ORDER (between-subject):			
		[d]=first deviant		[t]=first deviant	
PHONEME (within-subject):		*/d/*	*/t/*	*/d/*	*/t/*
CONDITION (within-subject):	*Standard*	Standard-D (Block 2)	Standard-T (Block 1)	Standard-D (Block 1)	Standard-T (Block 2)
	Deviant	Deviant-D (Block 1)	Deviant-T (Block 2)	Deviant-D (Block 2)	Deviant-T (Block 1)

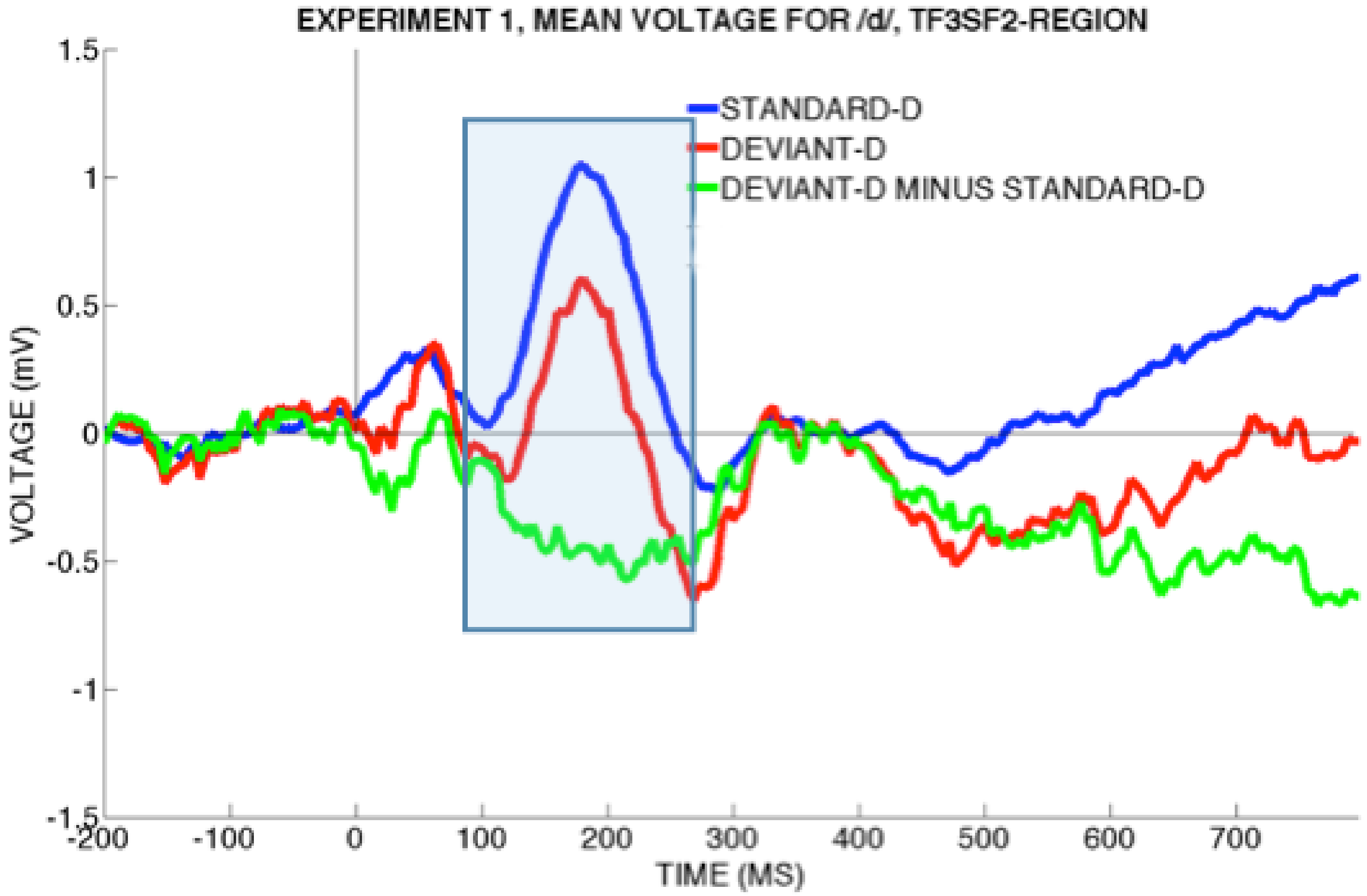
Table 1: Statistical design for all experiments.

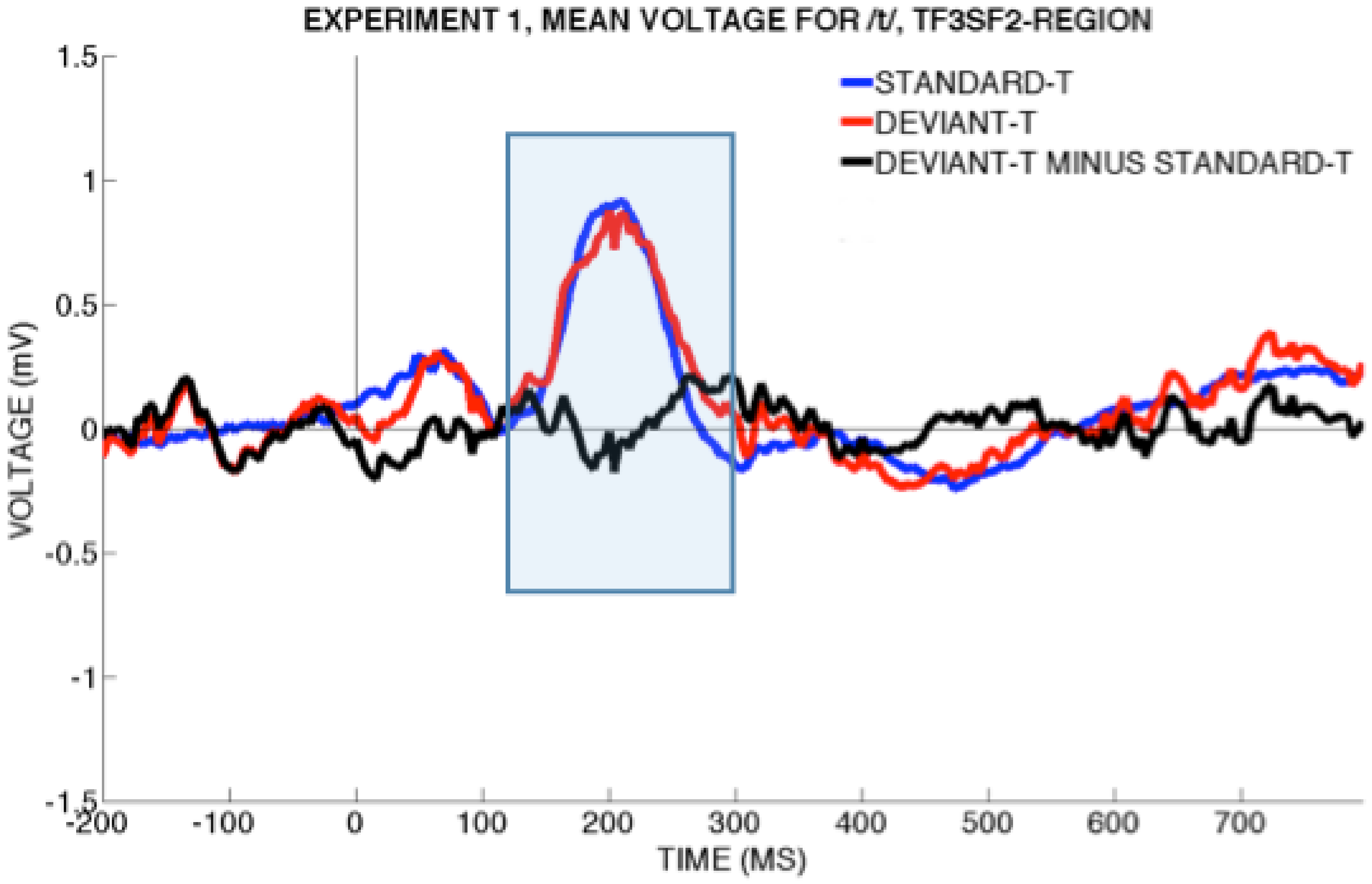
	[t]	[d]
Phonetic level (single oddball stimulus)	[+spread]	[-spread]
Phonemic level (memory trace from several standards)	/+spread/	/Ø/
	/t/	/d/

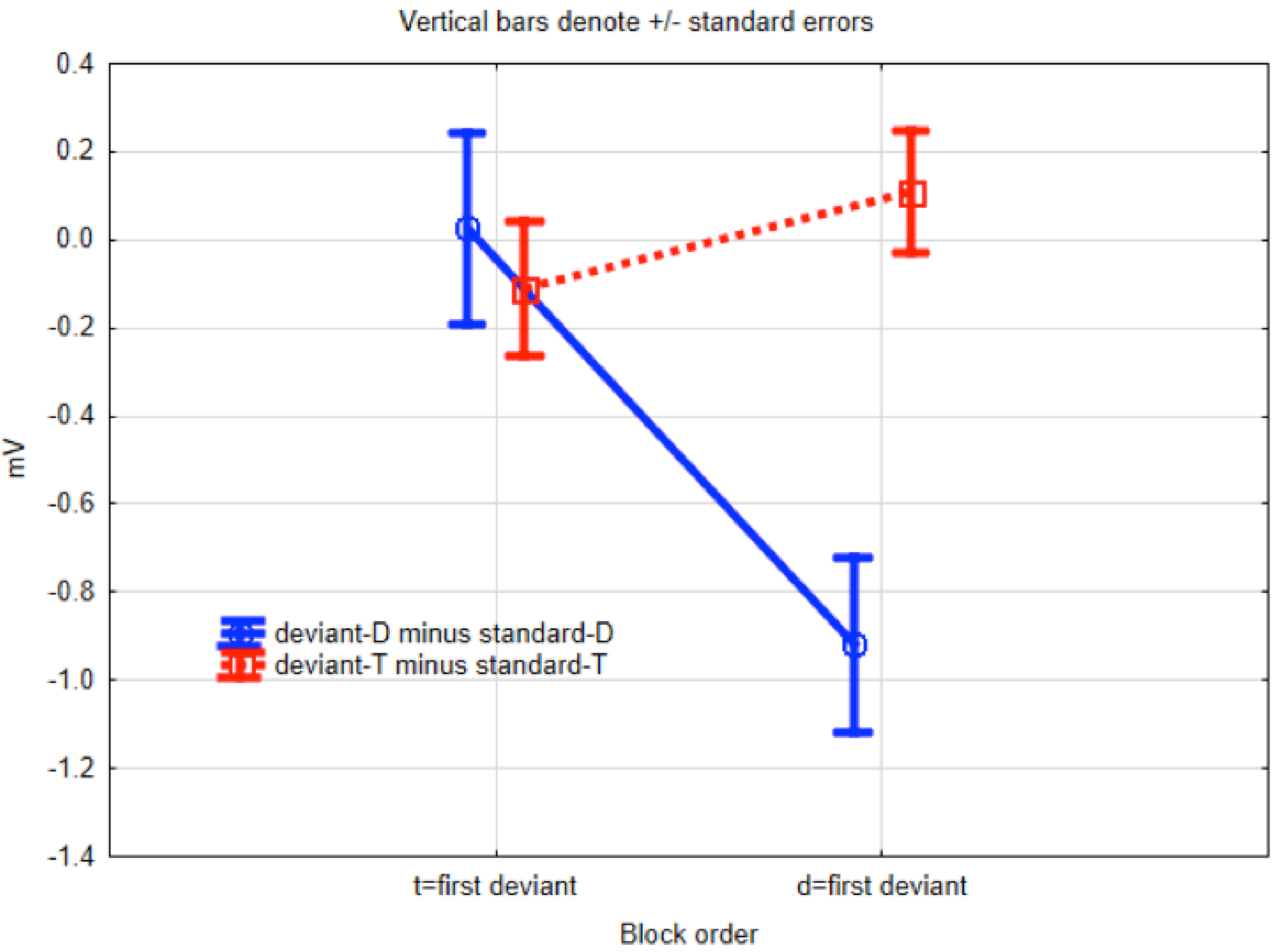


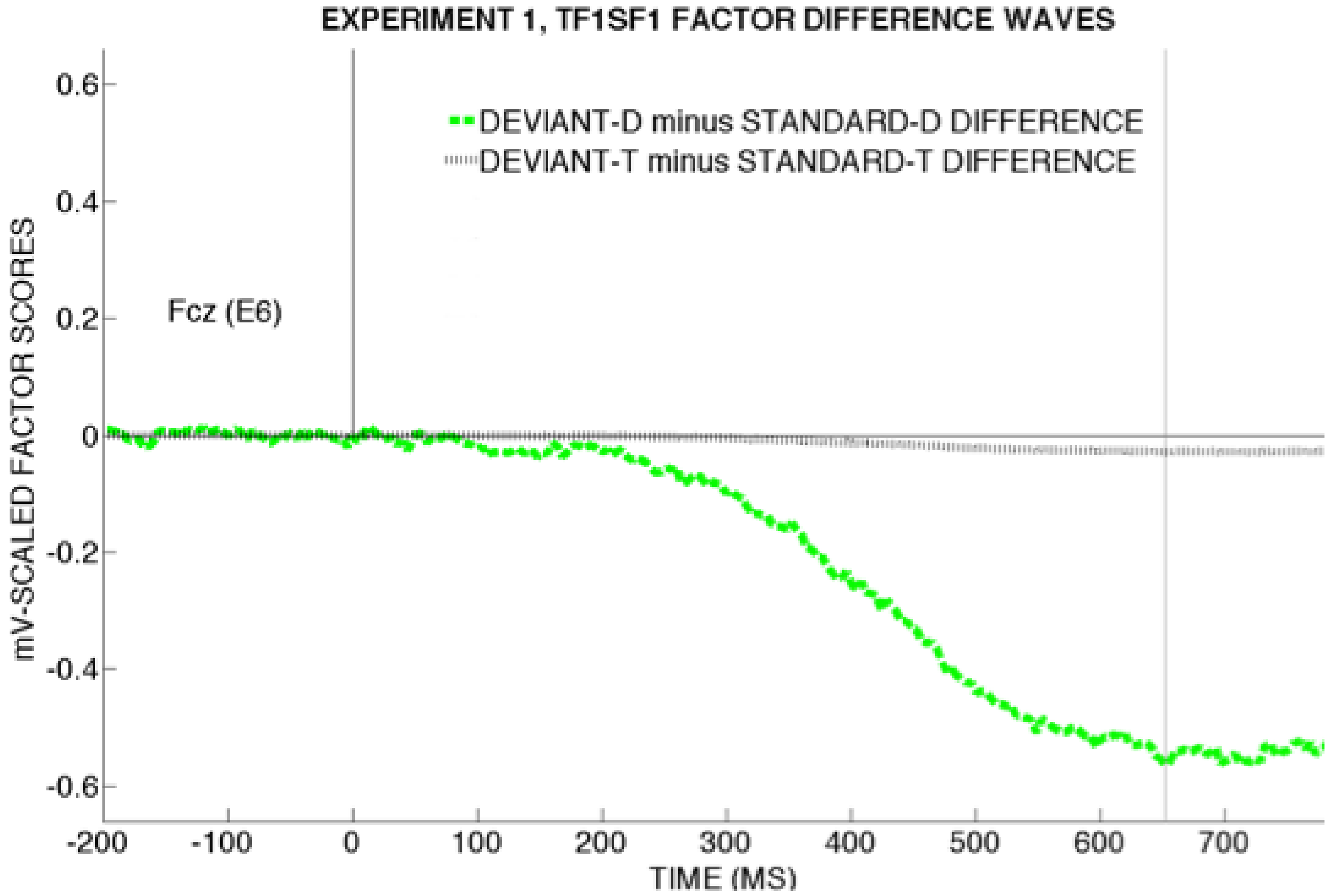

EXPERIMENT 1, TF3SF2 FACTOR DIFFERENCE WAVES

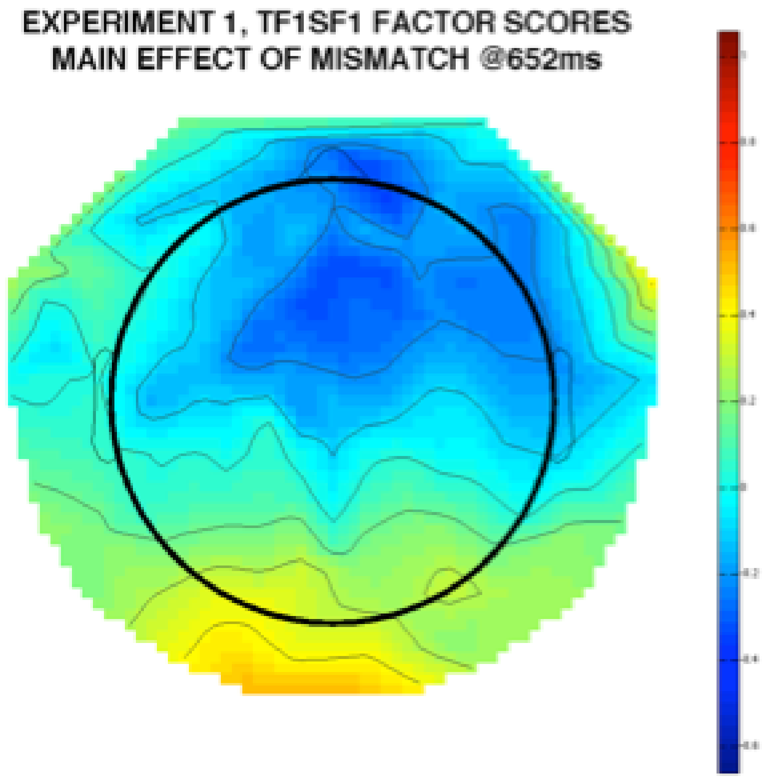

EXPERIMENT 1, TF3SF2 FACTOR SCORES
MAIN EFFECT OF MISMATCH @216ms

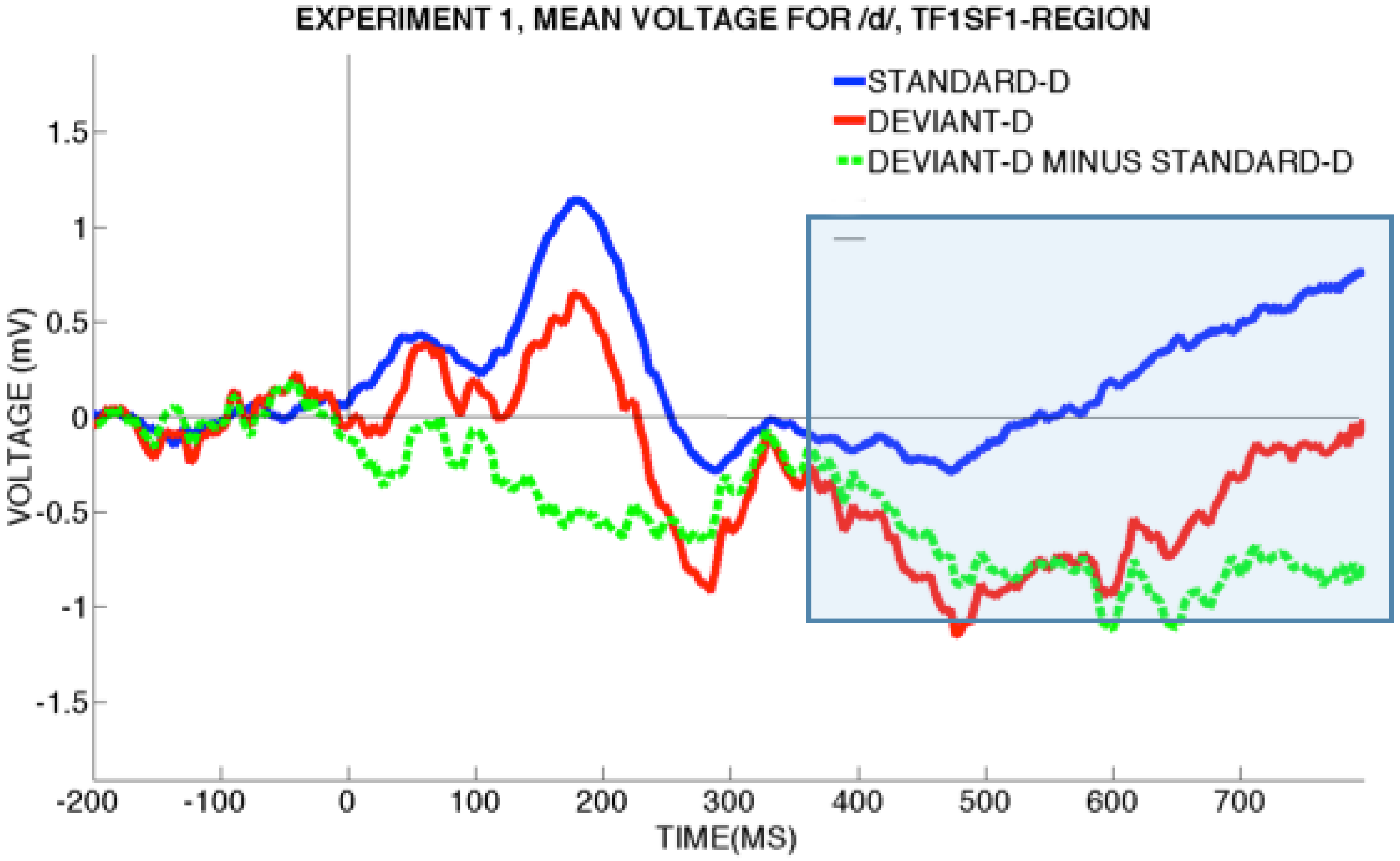

EXPERIMENT 1, MEAN VOLTAGE FOR /d/, TF3SF2-REGION

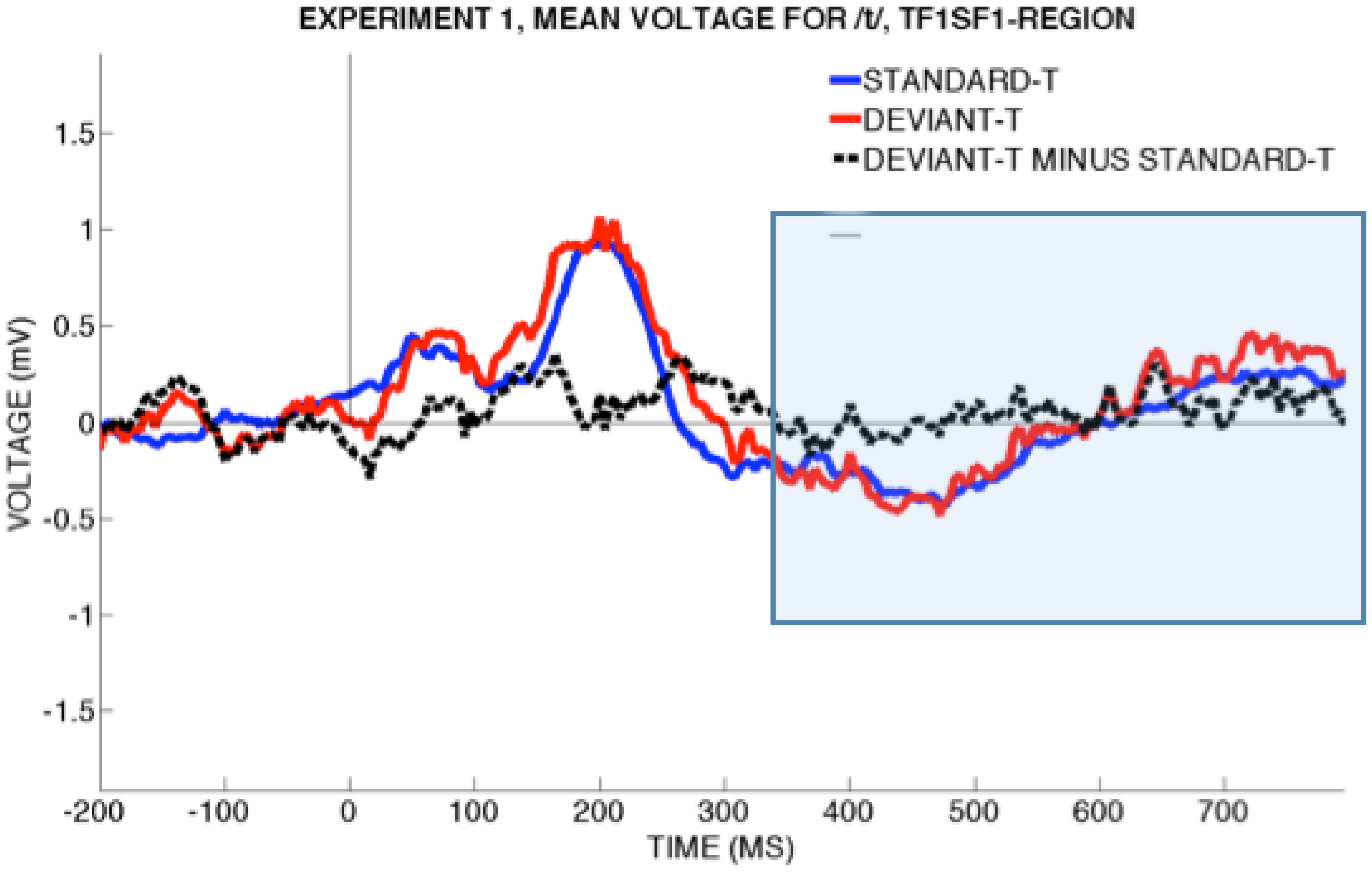

EXPERIMENT 1, MEAN VOLTAGE FOR /t/, TF3SF2-REGION


Vertical bars denote +/- standard errors
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.2
-1.4
mV
deviant-D minus standard-D
deviant-T minus standard-T
t=first deviant
d=first deviant
Block order

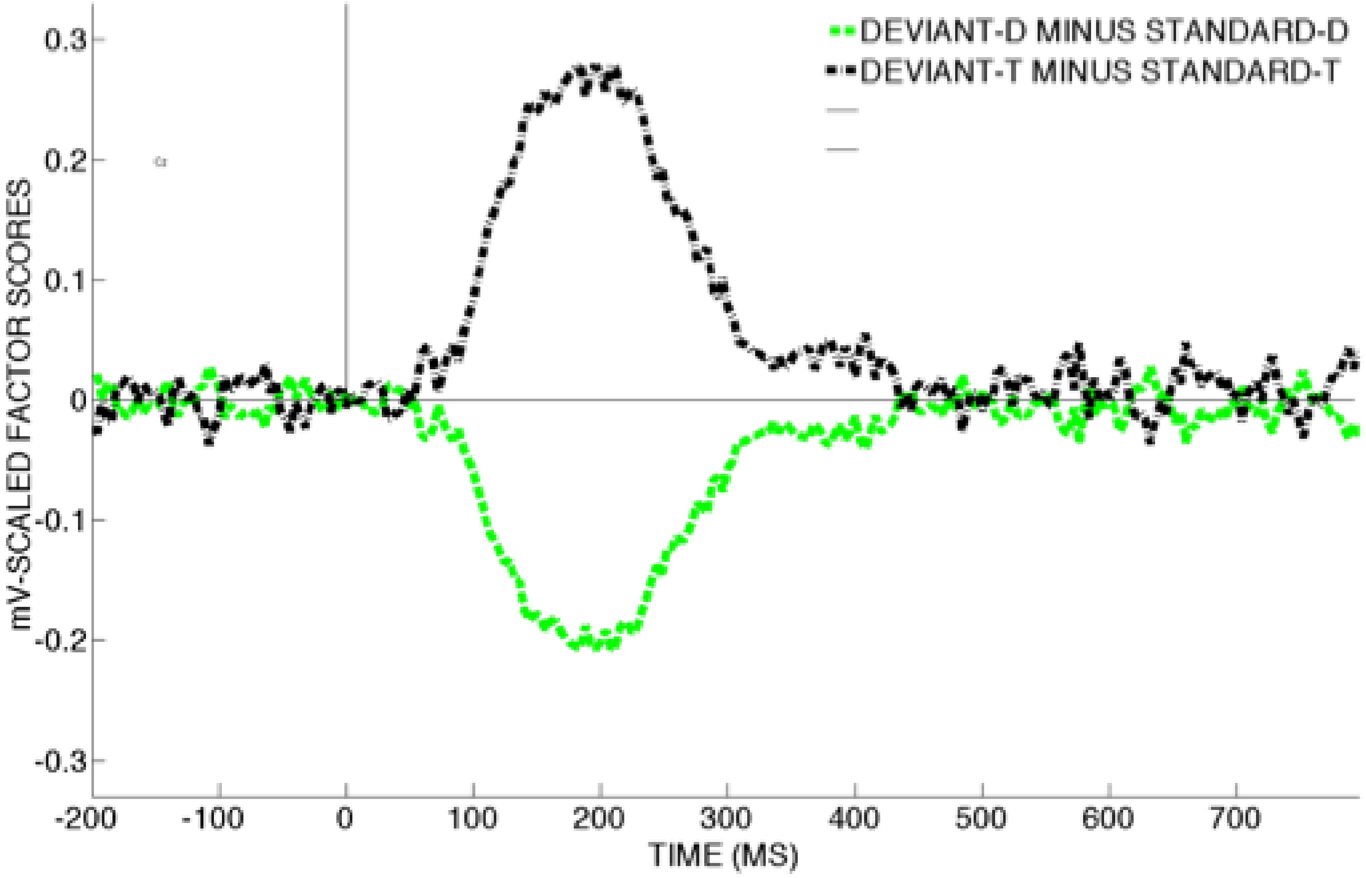

EXPERIMENT 1, TF1SF1 FACTOR DIFFERENCE WAVES


EXPERIMENT 1, TF1SF1 FACTOR SCORES
MAIN EFFECT OF MISMATCH @652ms

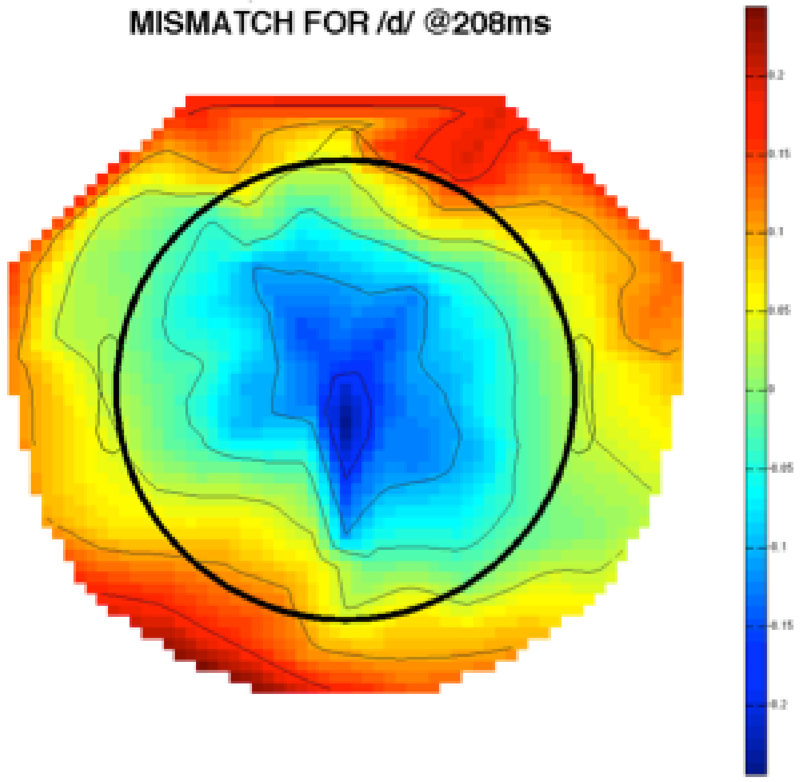

EXPERIMENT 1, MEAN VOLTAGE FOR /d/, TF1SF1-REGION


EXPERIMENT 1, MEAN VOLTAGE FOR /t/, TF1SF1-REGION

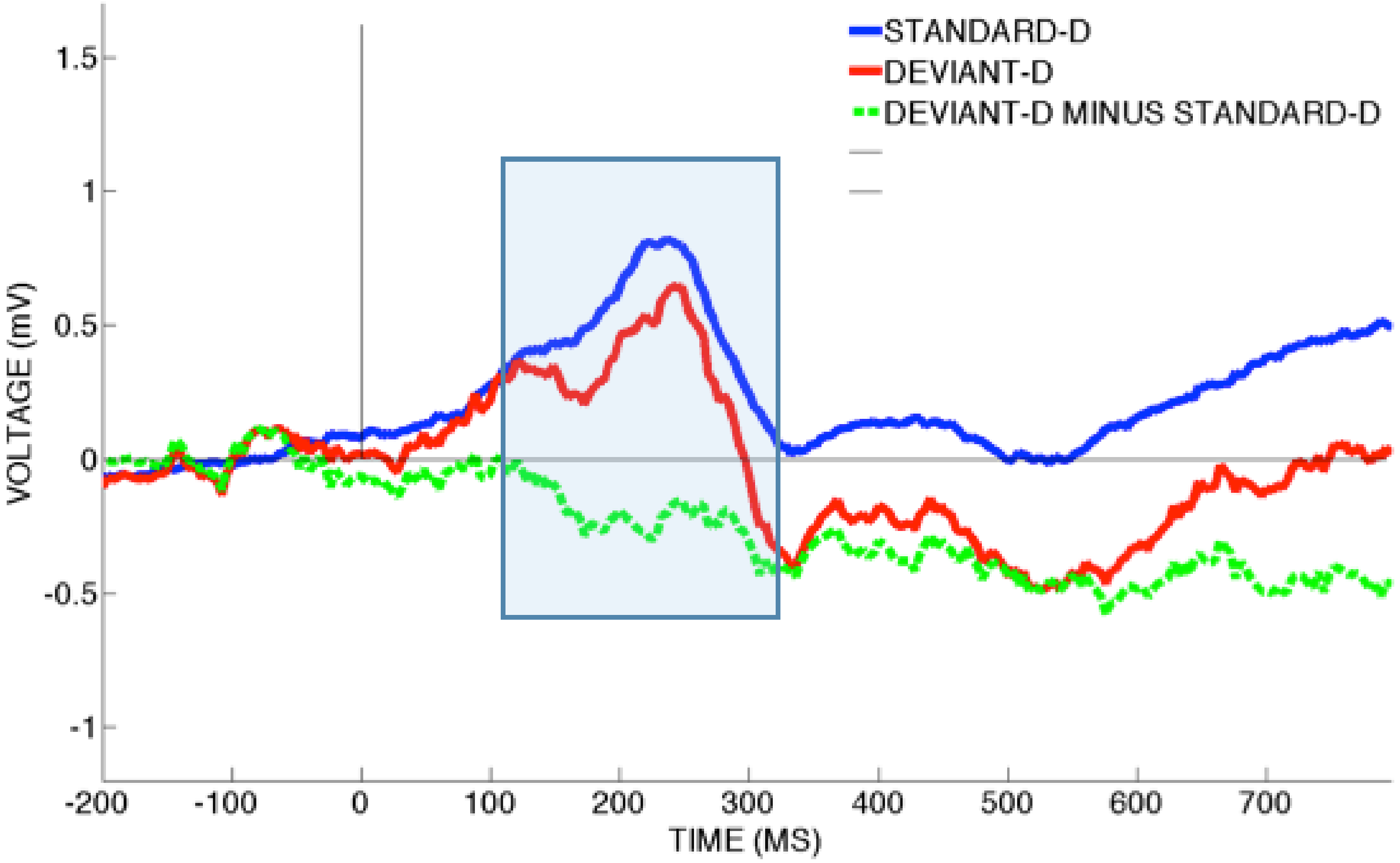
EXPERIMENT 2, TF2SF2 FACTOR DIFFERENCE WAVES



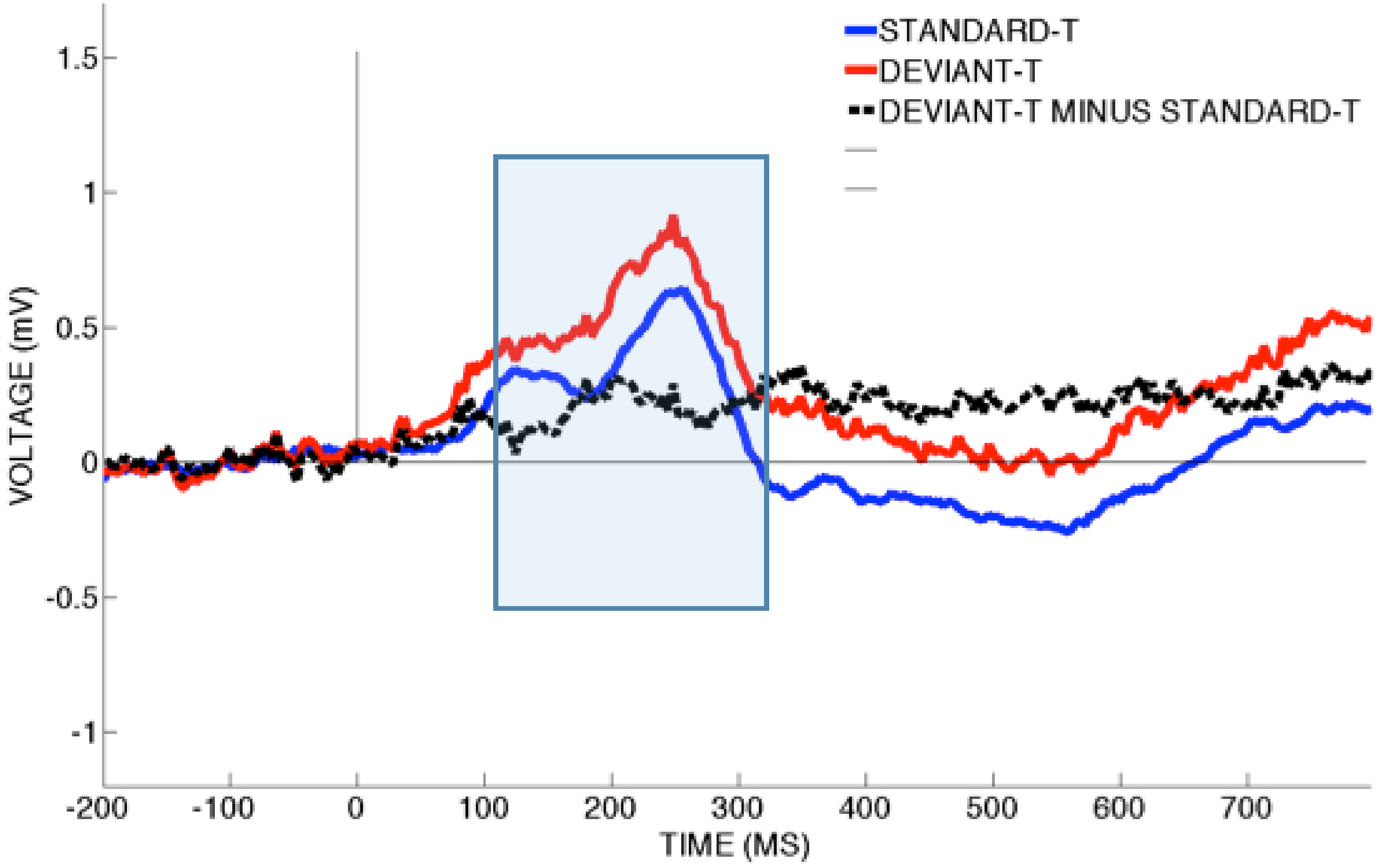
EXPERIMENT 2, TF2SF2 FACTOR SCORES
MISMATCH FOR /d/ @208ms

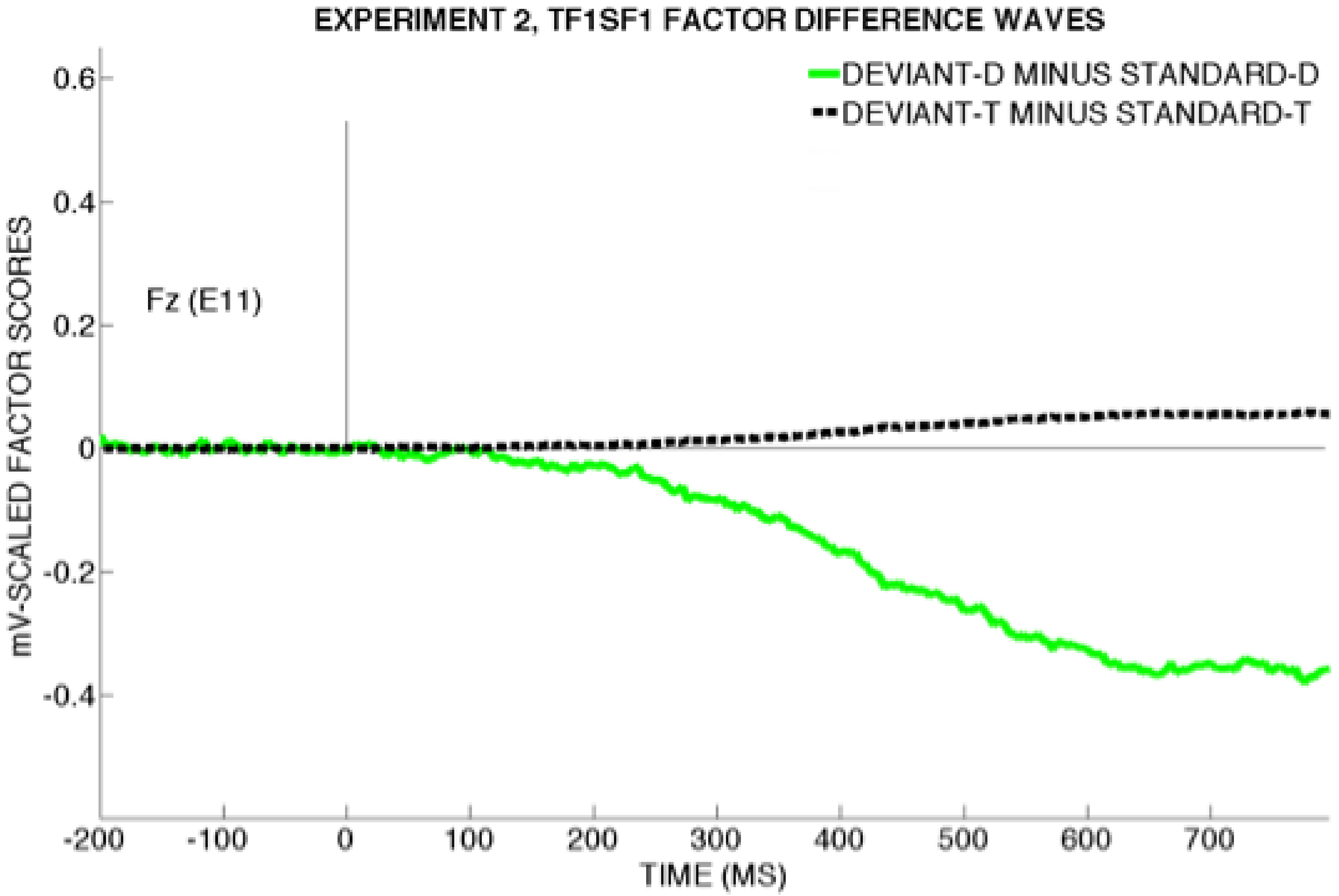


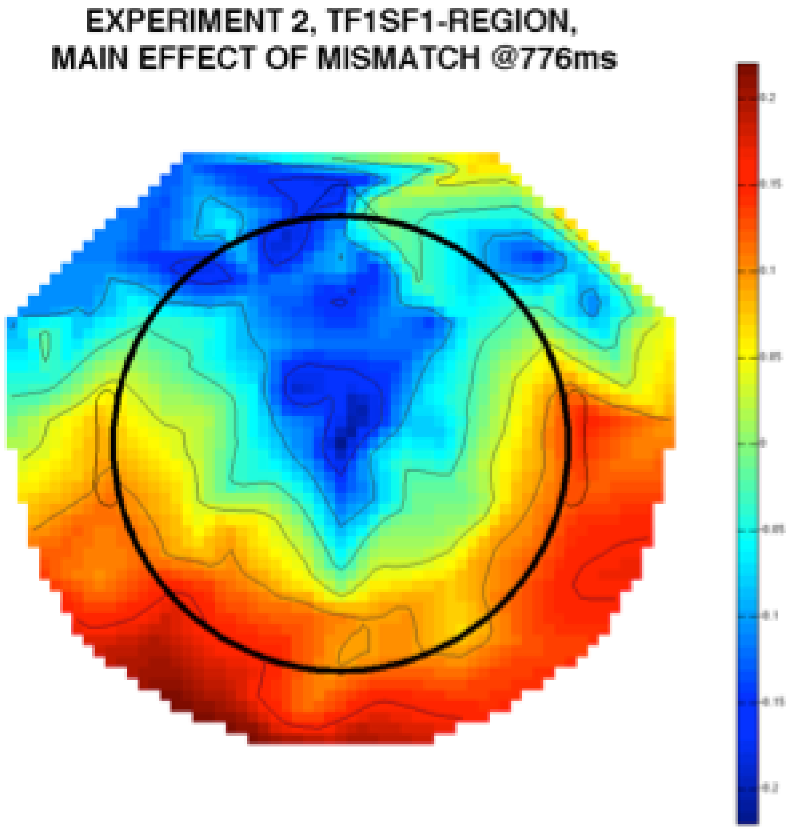
EXPERIMENT 2, MEAN VOLTAGE FOR /d/, TF2SF2-REGION

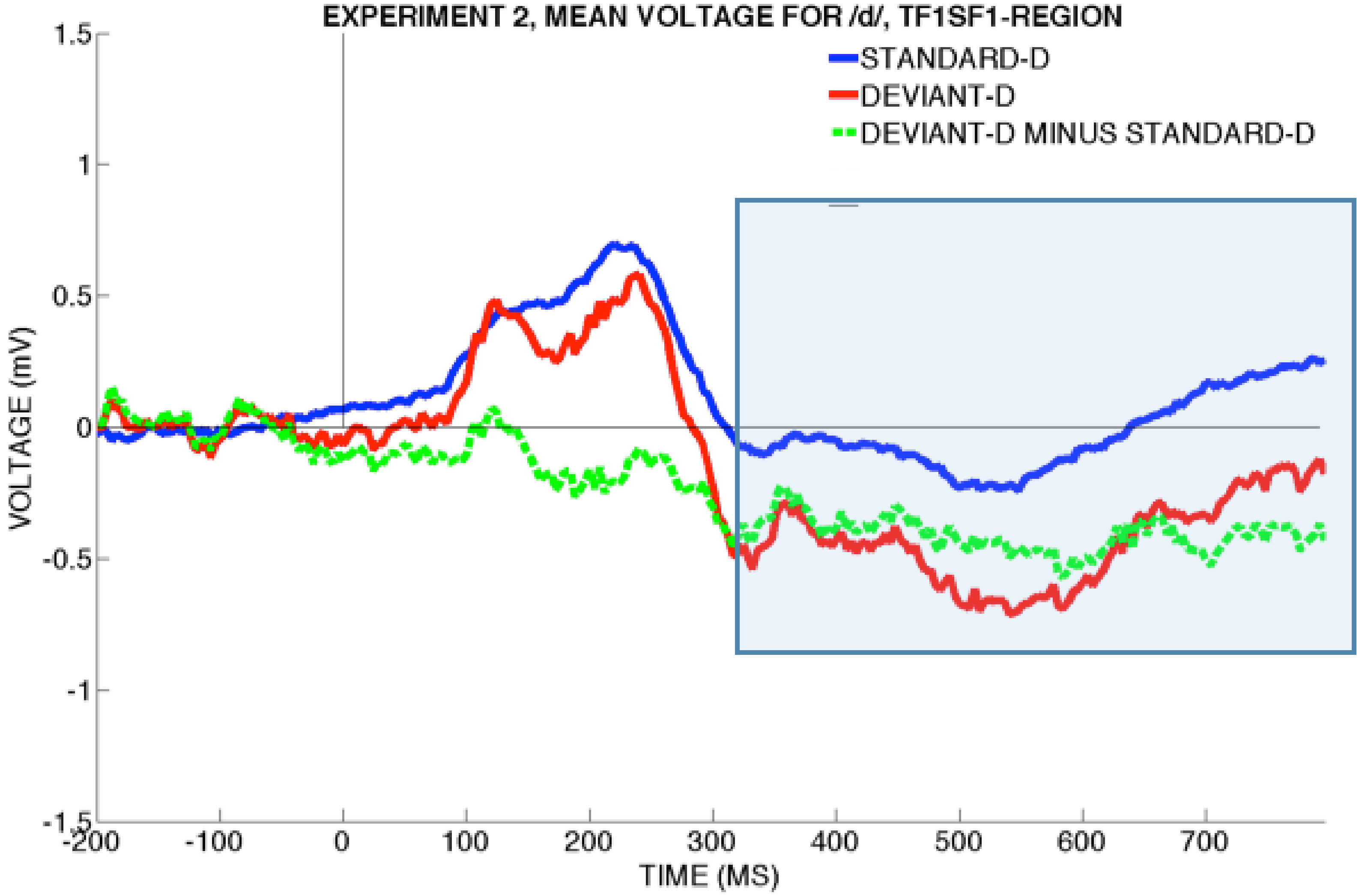


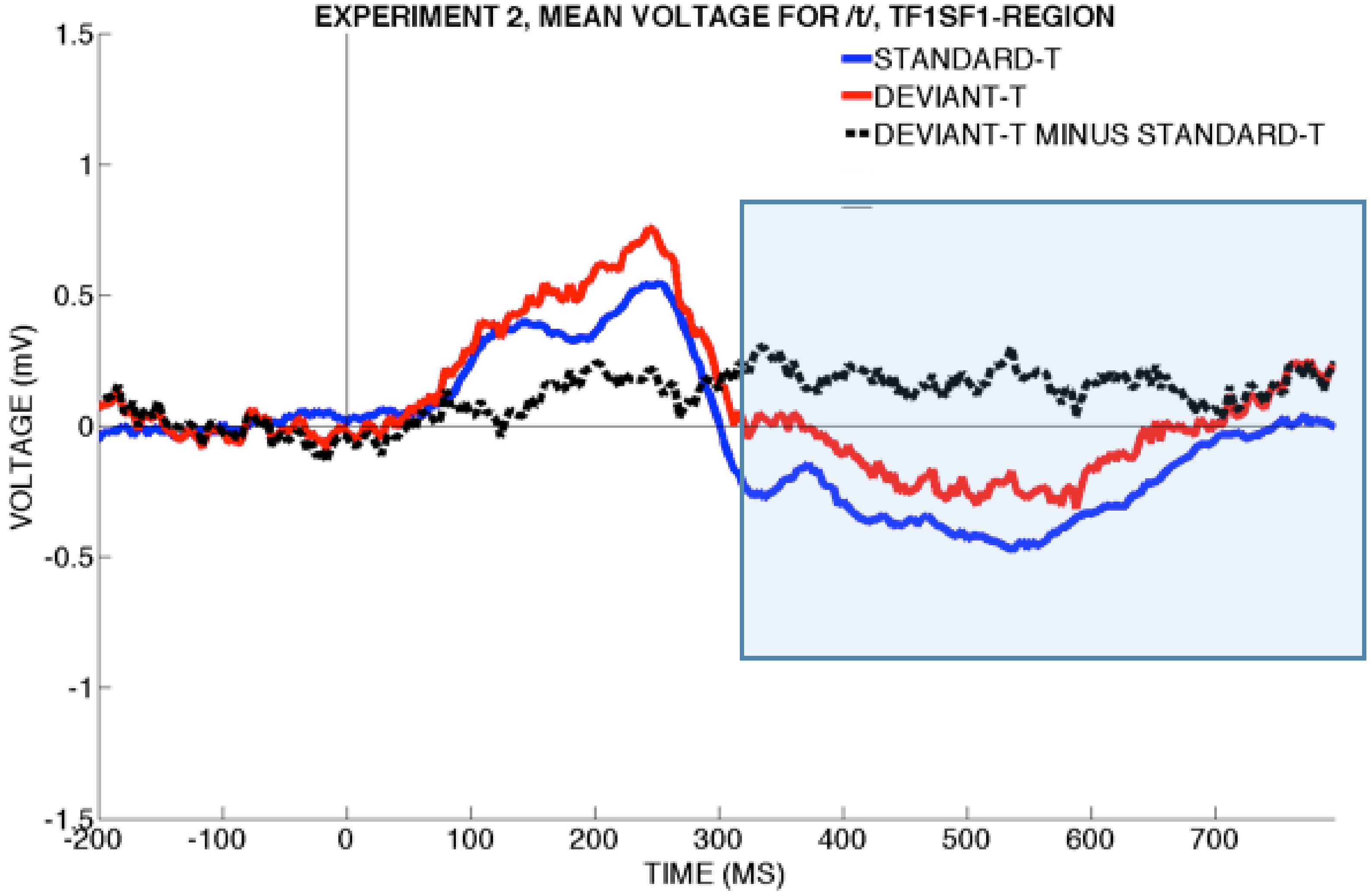
EXPERIMENT 2, MEAN VOLTAGE FOR /t/, TF2SF2-REGION



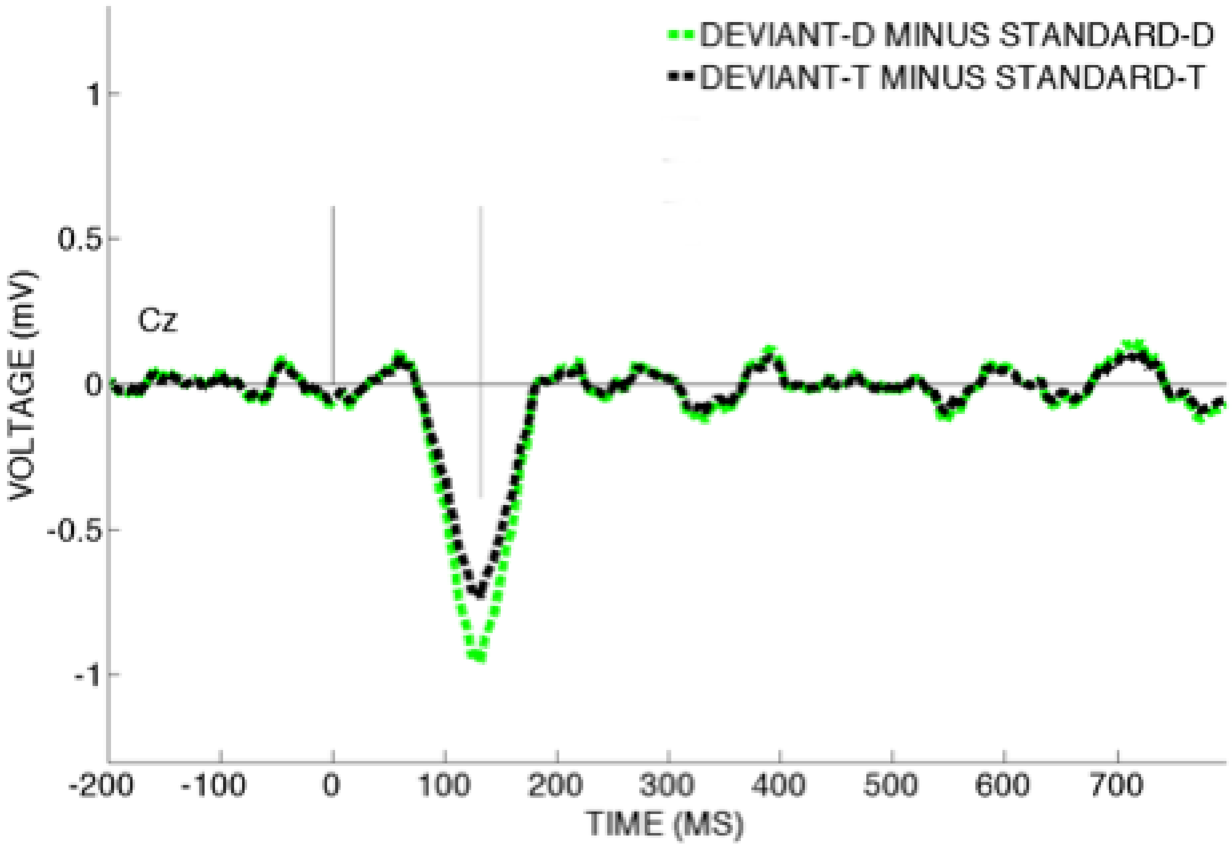

EXPERIMENT 2, TF1SF1 FACTOR DIFFERENCE WAVES


EXPERIMENT 2, TF1SF1-REGION,
MAIN EFFECT OF MISMATCH @776ms

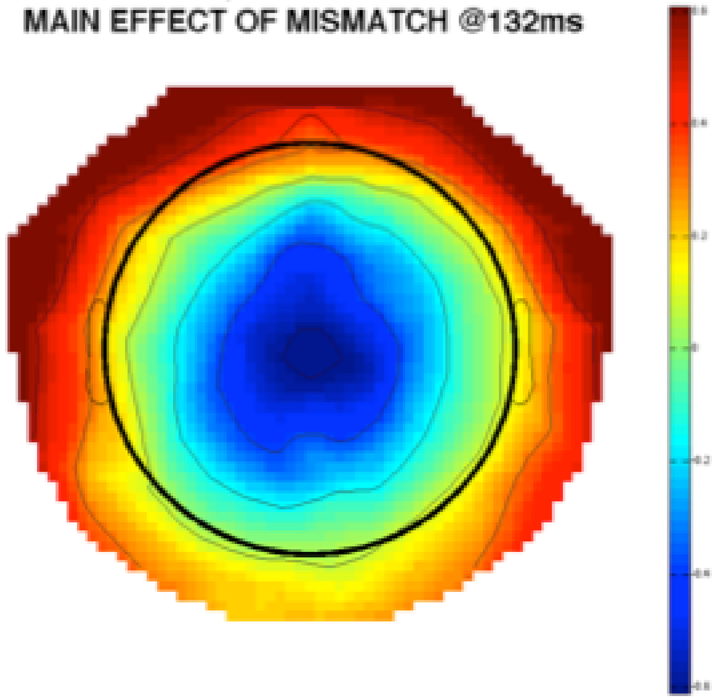

EXPERIMENT 2, MEAN VOLTAGE FOR /d/, TF1SF1-REGION


EXPERIMENT 2, MEAN VOLTAGE FOR /t/, TF1SF1-REGION

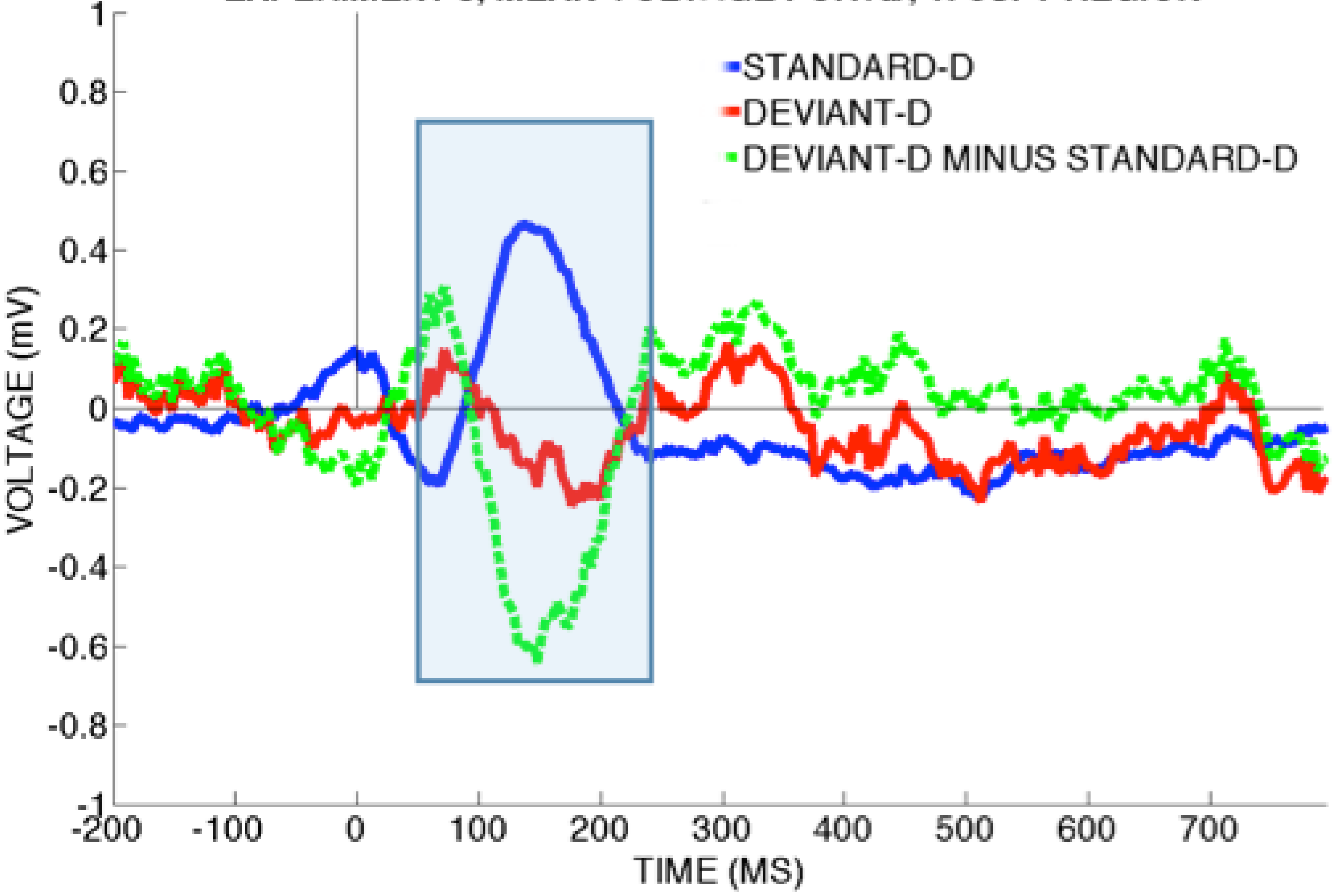
EXPERIMENT 3, TF5SF1 FACTOR SCORE DIFFERENCE WAVES



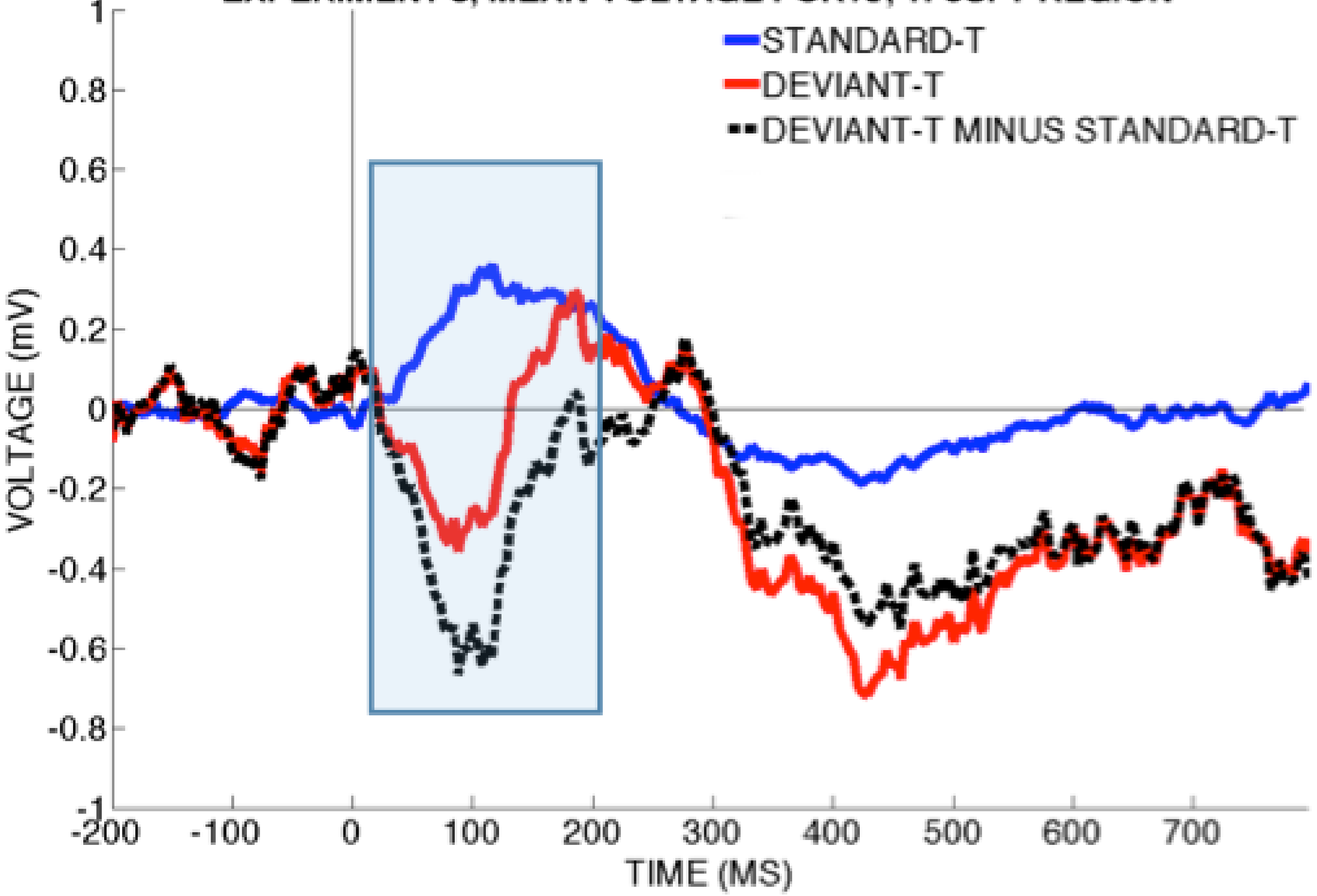
EXPERIMNENT 3, TF5SF1 FACTOR SCORES
MAIN EFFECT OF MISMATCH @132ms

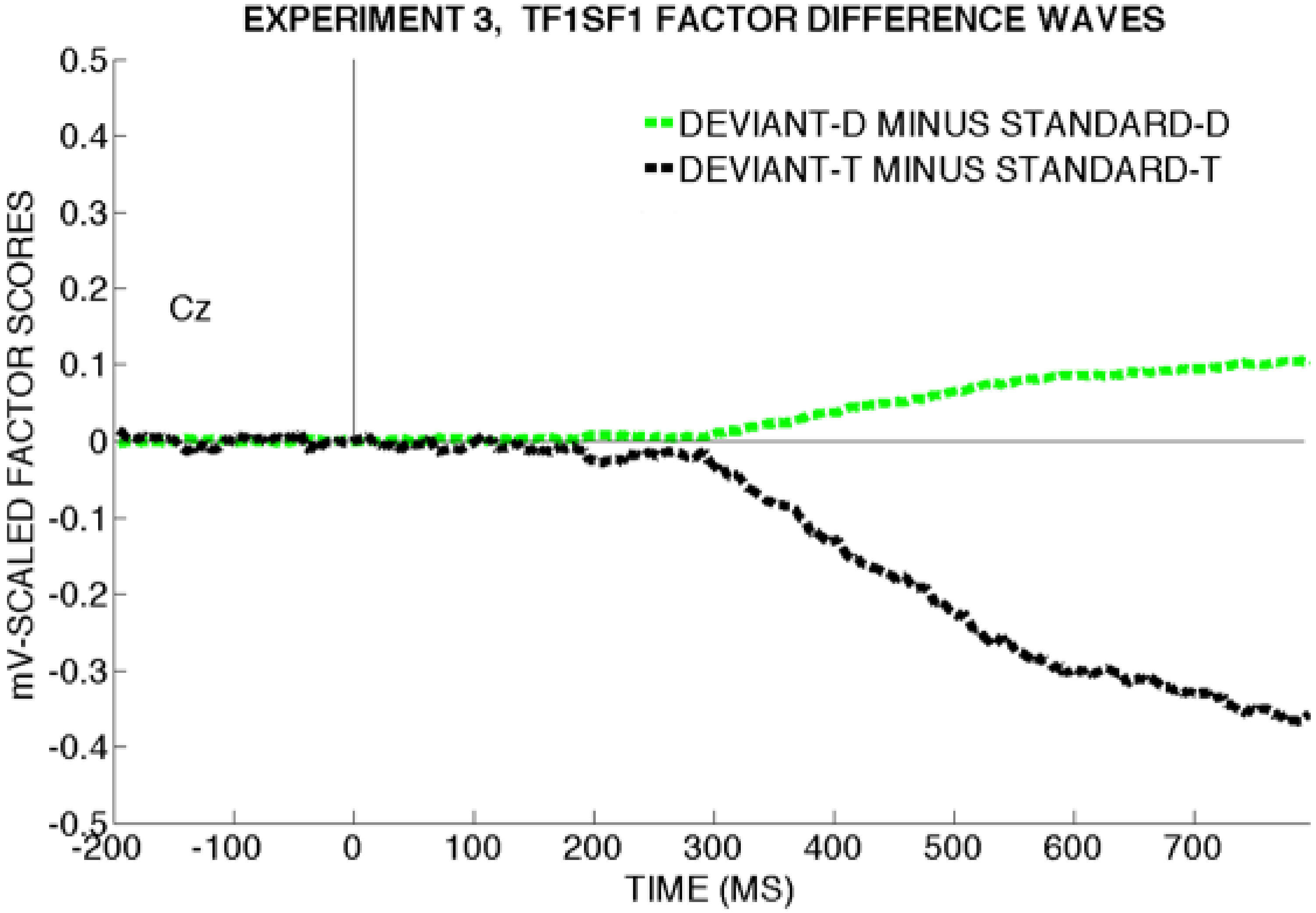


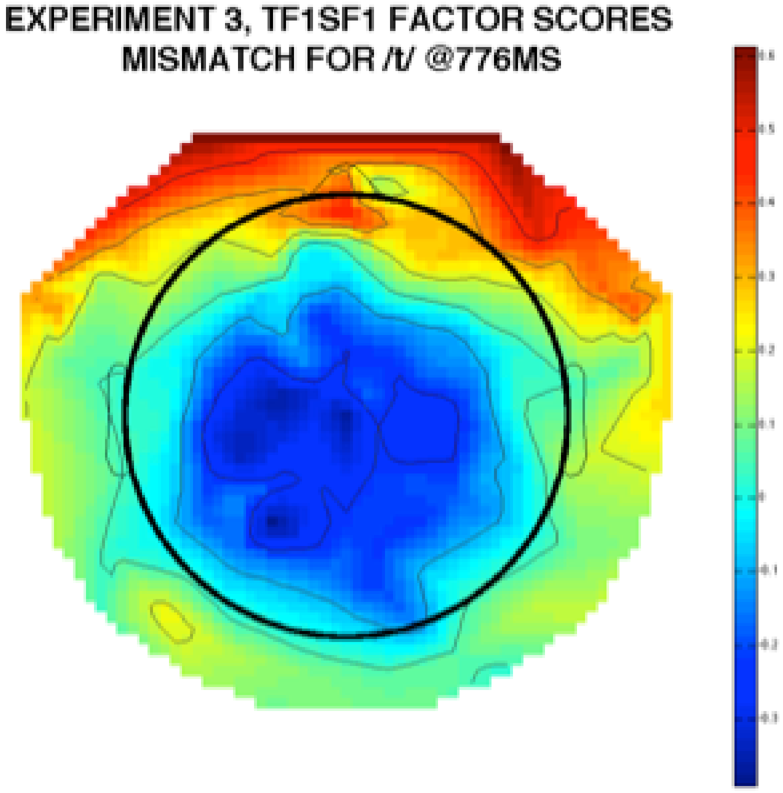
EXPERIMENT 3, MEAN VOLTAGE FOR /d/, TF5SF1-REGION

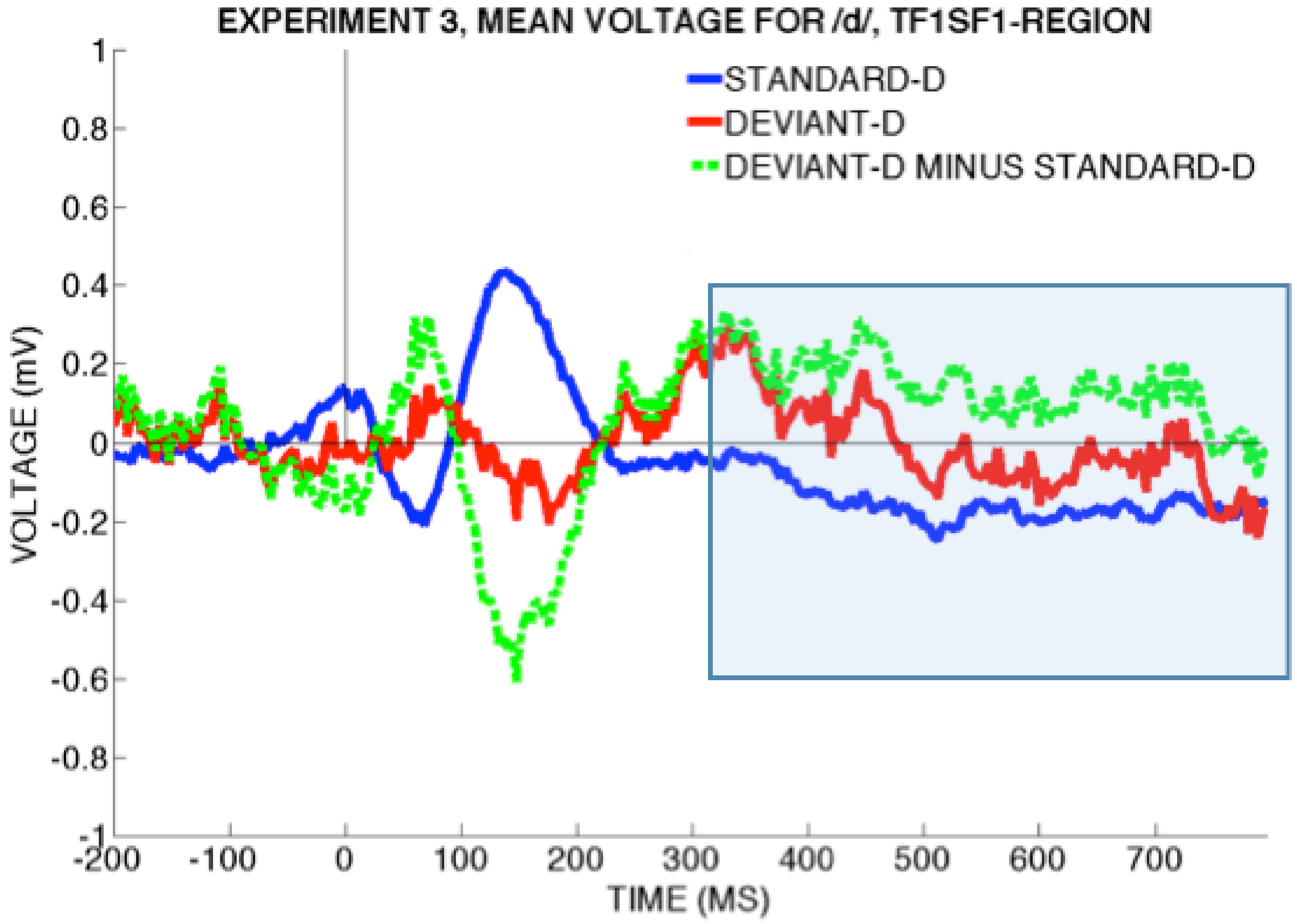


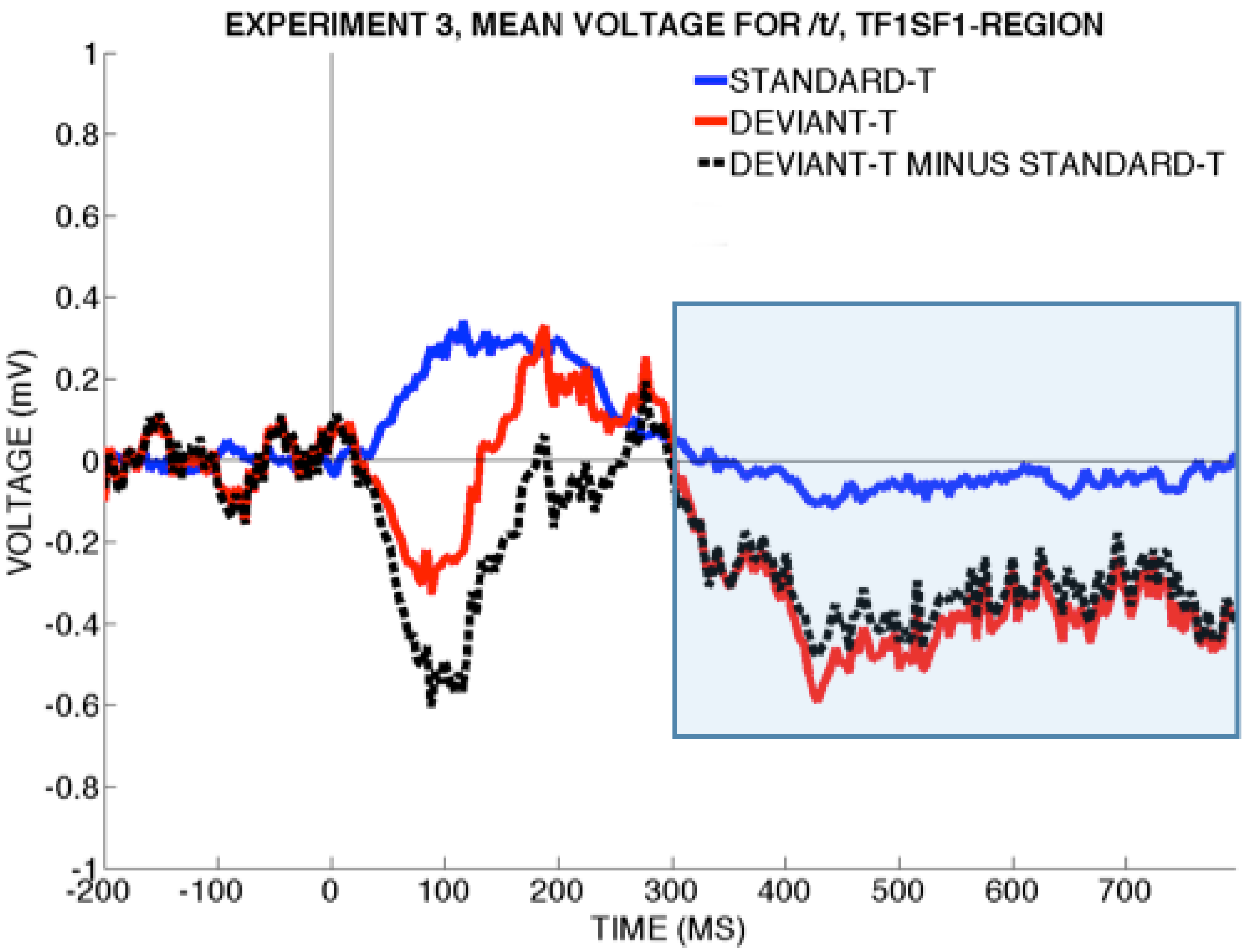
EXPERIMENT 3, MEAN VOLTAGE FOR /t/, TF5SF1-REGION

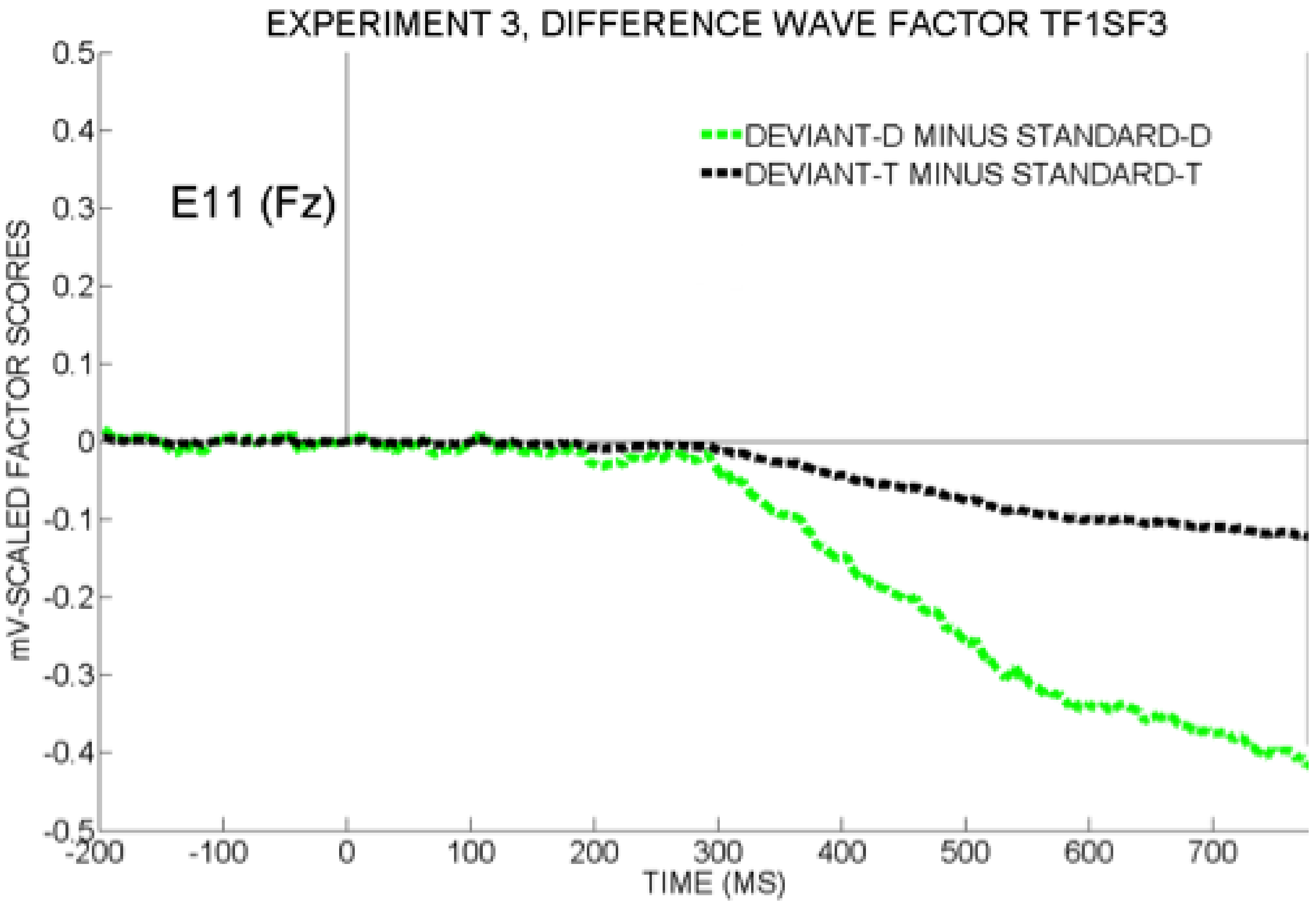


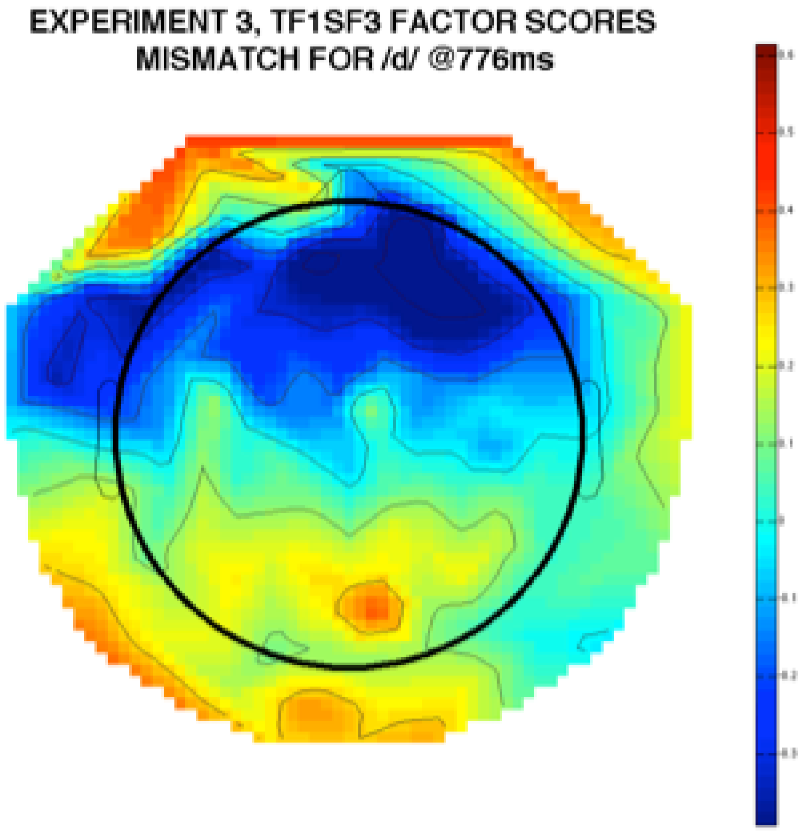

EXPERIMENT 3, TF1SF1 FACTOR DIFFERENCE WAVES
DEVIANT-D MINUS STANDARD-D
DEVIANT-T MINUS STANDARD-T
Cz
mV-SCALED FACTOR SCORES
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
-0.5
-200
-100
0
100
200
300
400
500
600
700
TIME (MS)

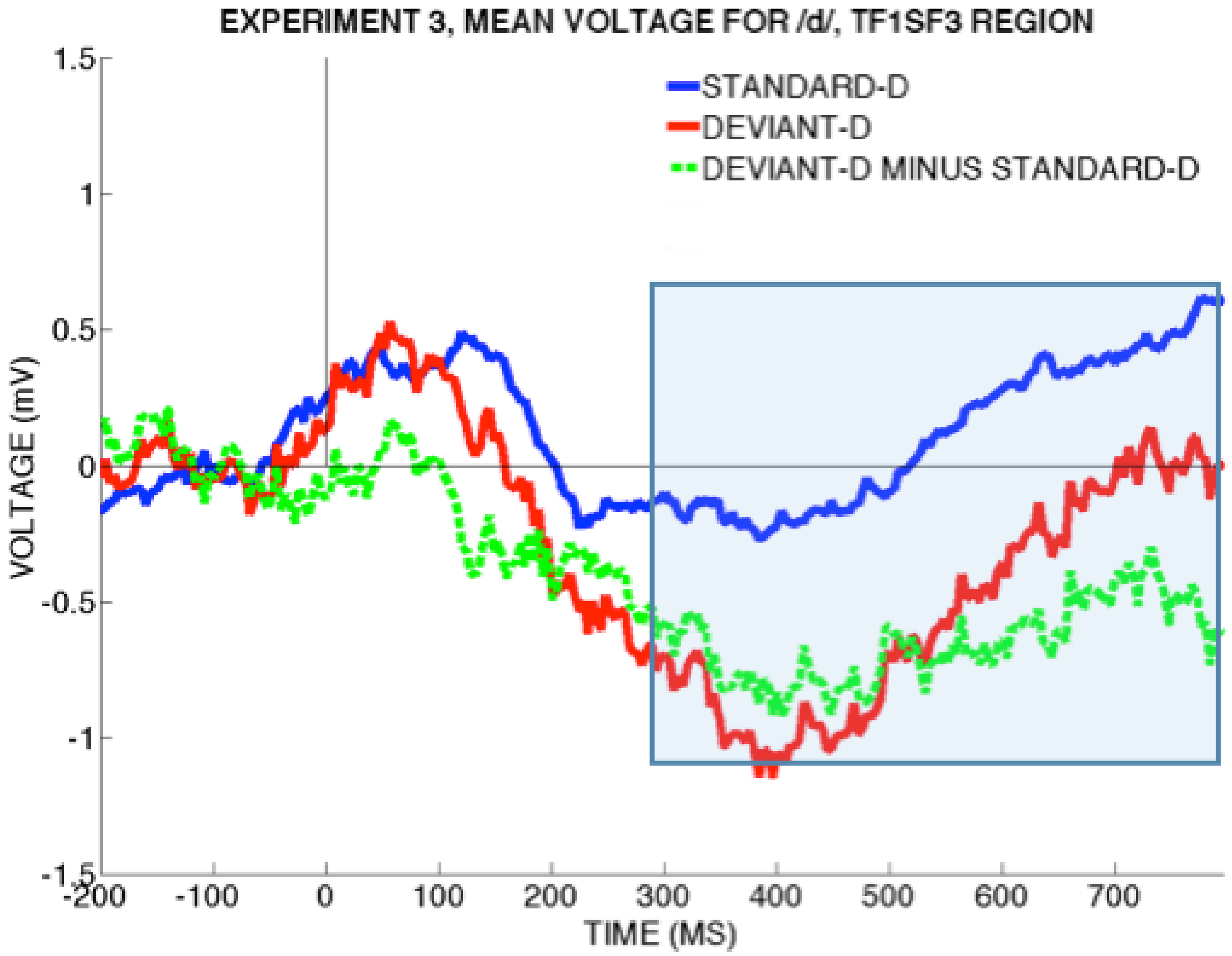

EXPERIMENT 3, TF1SF1 FACTOR SCORES
MISMATCH FOR /t/ @776MS

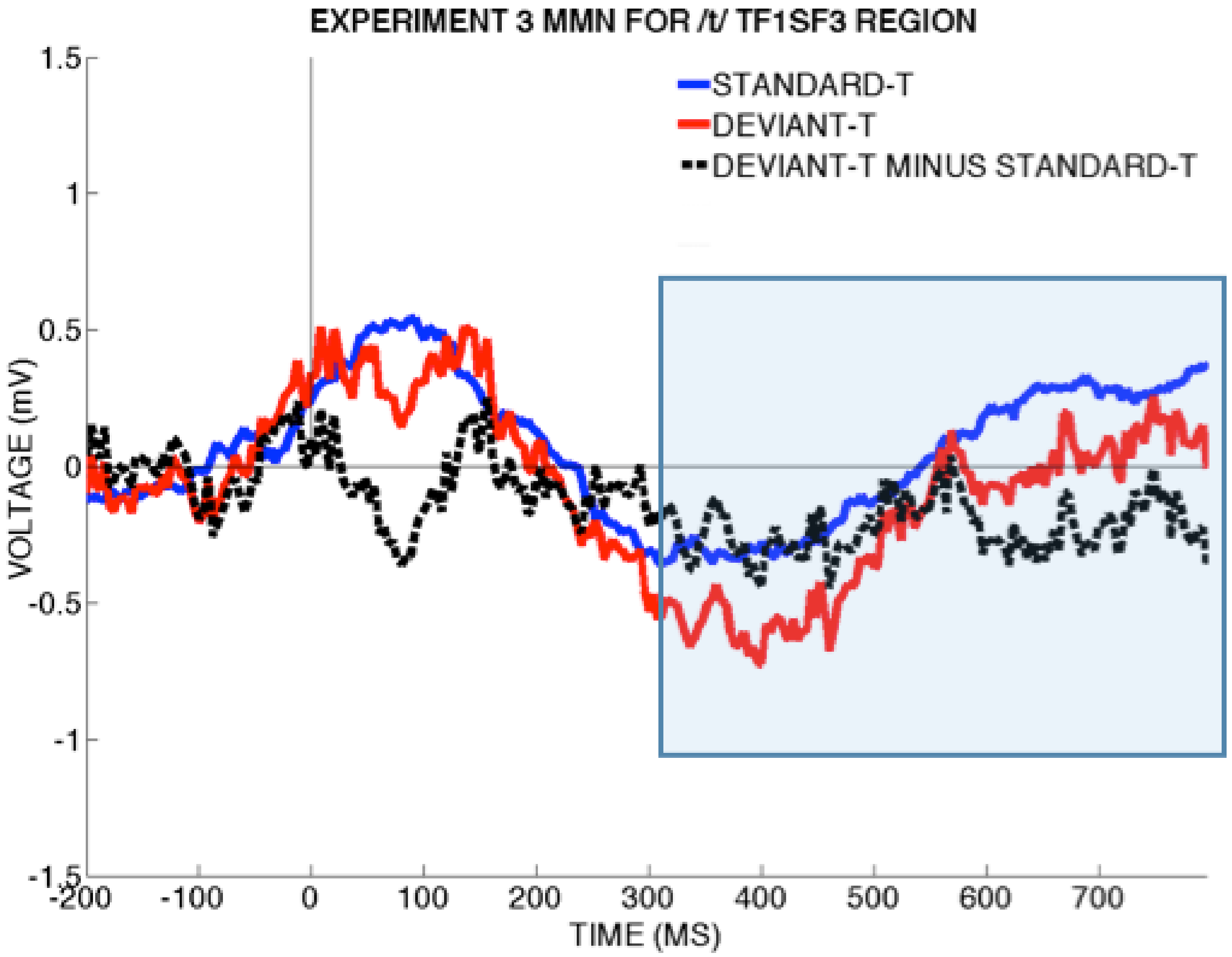

EXPERIMENT 3, MEAN VOLTAGE FOR /d/, TF1SF1-REGION
STANDARD-D
DEVIANT-D
DEVIANT-D MINUS STANDARD-D
VOLTAGE (mV)
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
-200
-100
0
100
200
300
400
500
600
700
TIME (MS)

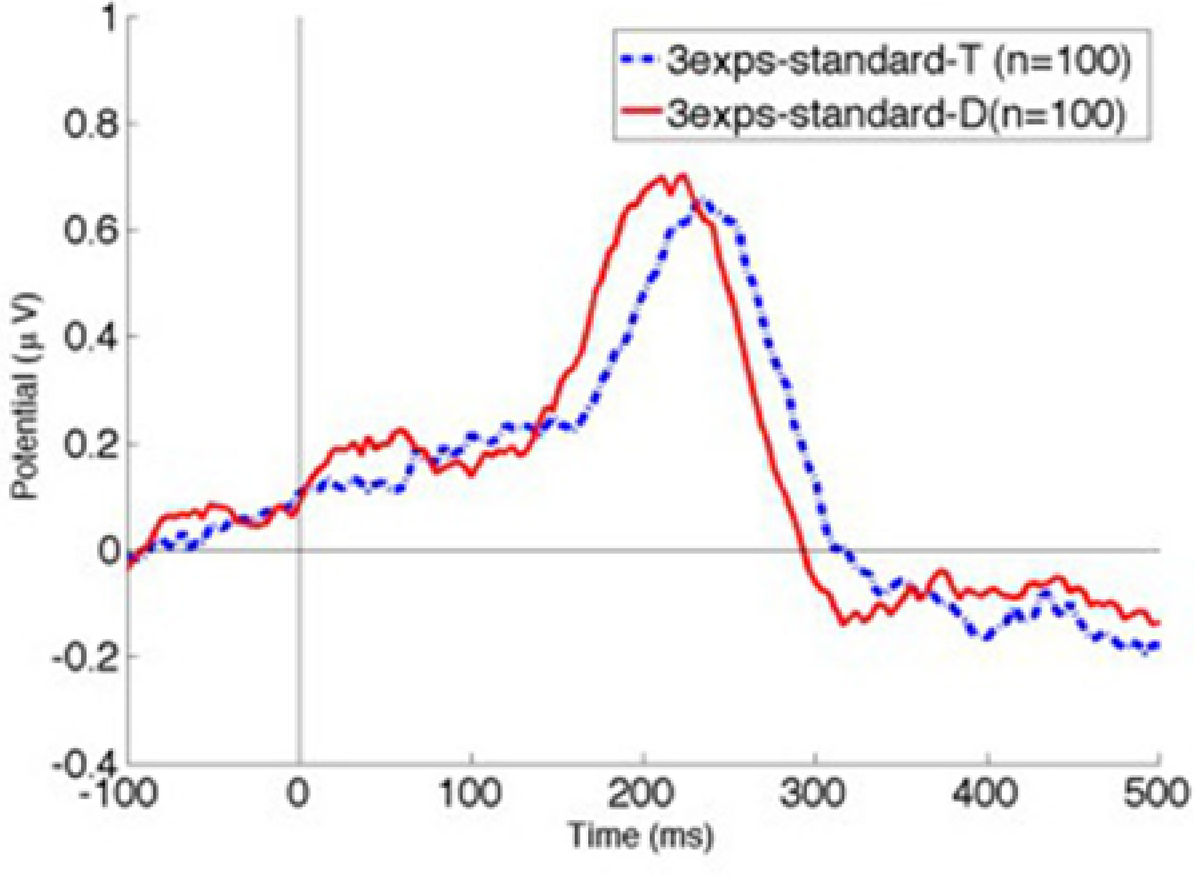

EXPERIMENT 3, MEAN VOLTAGE FOR /t/, TF1SF1-REGION
STANDARD-T
DEVIANT-T
DEVIANT-T MINUS STANDARD-T
VOLTAGE (mV)
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
-200
-100
0
100
200
300
400
500
600
700
TIME (MS)


EXPERIMENT 3, DIFFERENCE WAVE FACTOR TF1SF3


EXPERIMENT 3, TF1SF3 FACTOR SCORES
MISMATCH FOR /d/ @776ms


EXPERIMENT 3, MEAN VOLTAGE FOR /d/, TF1SF3 REGION


EXPERIMENT 3 MMN FOR /t/ TF1SF3 REGION


3exps-standard-T (n=100)
3exps-standard-D(n=100)
Potential (μ V)
Time (ms)
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-100
0
100
200
300
400
500

Long-term memory representations of phoneme feature matrices are copied into short-term memory during an MMN experiment. MMN generators access this information to produce asymmetric MMN responses at the scalp, due /d/ being an underspecified feature matrices whereas /t/ is fully specified. This distinction tapped into by the MMN ERP response.

Highlights

- Long-term memory representations are phonologically underspecified.
- Varying standards MMN paradigm tap into phoneme-level representations.
- Asymmetric MMN responses indicate underspecified memory traces for standards.
- Findings support laryngeal underspecification for /d/ but not /t/.
- Single-token standards paradigm results in symmetric MMN, not previously tested.