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

Phonemes, as opposed to phonetic units, are only coded for those features and feature values that are necessary to express the minimal oppositional contrast in the lexicon. In addition, underspecification theory says that phonemes that have unmarked feature values are even more abstract in that the feature is omitted from the representation altogether. Eulitz and Lahiri (2004) invented a new way of testing this theory, by showing that certain asymmetries in the Mismatch Negativity response to phoneme contrasts are predicted by underspecification. We expand on this research by demonstrating underspecification-driven asymmetry in the MMN response to laryngeal feature contrasts in English. We also extend the logic of previous experimentation by predicting (and observing) a lack of asymmetry in the MMN if the experimental paradigm encourages formation of phonetic memory traces. This result further strengthens the view that phonemes are more sparsely represented than phonetic units.

1. INTRODUCTION

1.1 Background: Phonological underspecification

The basic representational claims of underspecification trace their origins back to claims made in the framework of Generative Phonology, according to which memory representations of words consist of phonemes, which in turn are bundles of distinctive features [1], [2]. Distinctive features are a subset of phonetic features, namely the minimal subset required for distinguishing between the words in the lexicon. The full phonetic information required to produce 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 phonemic and phonetic representations 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.[1]

Underspecification theory takes this one step further, by incorporating an additional asymmetry into the representations, namely the idea that one value of a feature may be the "default", with the other value being "marked". To illustrate, suppose two words differ only in terms of nasality in one segment, and that the default value of nasality in the language is [-nasal], then only the word that has the more marked value needs to express that value in the memory representation. Default rules will then fill in the other value to produce a full phonetic representation. For example, if a language contains two vowels that contrast in nasality, say /ã/ and /a/, and the phoneme /ã/ is lexically specified for [+nasal], then the phoneme /a/ need not be specified for [-nasal], as /a/ is already distinguished from /ã/ in not having the feature [+nasal][2].

---

[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 cf. [45].

[2] This view of phonological representations is a notational equivalent of a view incorporating unary features [46], [47].

In the phonetic form, the [a] is equipped with [-nasal] by a redundancy rule that says that vowels not specified for [+nasal] must be [-nasal]. This is the basic hypothesis of Phonological Underspecification [3]–[6]. While this hypothesis has been criticized by some [7]–[9], several authors argue that it provides an unparalleled analysis of phonological patterns in many languages [3]–[5], [10], [11].

1.2 Cognitive neuroscience evidence for underspecification

Arguments for underspecification in linguistics have traditionally been based on the analysis of phonological patterns, but evidence has recently also come from techniques that access the brain's representations of speech sounds. Eulitz and Lahiri [12] used the Mismatch Negativity (MMN) paradigm to provide neurobiological evidence for underspecified phonemic representations. We now turn to a review of their findings and the logic of their experimental paradigm, which we emulate in the current study.

In the MMN paradigm, trains of "standard" sounds are presented to subjects. Each sound presentation elicits an Auditory Evoked Potential, which is a sequence of negative and positive waveforms—the N1-P2 complex (N1=the first negative dip in the waveform, and P2=the first positive shift). After a series of standards, the stimulus train is interrupted by a different stimulus (the "oddball" or "deviant" stimulus). The deviant is compared to the memory trace of the standard, and if the deviant is perceived as differing from the memory trace along some dimension, an attenuation of the amplitude of the Auditory Evoked Potential to the deviant stimulus is observed, compared to the standard stimulus waveform (between 100-300ms after stimulus onset). The difference wave obtained by subtracting the deviant waveform from the standard waveform represents the effect of mismatch, hence called Mismatch Negativity.

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 [13]–[15]. It has also been used to show that speakers are sensitive to more abstract phonemic contrasts [16]–[20]. In an experiment that we partially replicate in the current study, Phillips et al. [21] 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. Since the perceptual system encounters physically different stimuli, the only way for a single memory trace to be created, they reasoned, would be for the auditory system to use the phonemic representation (/d/ or /t/) as a memory trace. Indeed, a mismatch was observed when a deviant [d] stimulus was 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]. Crucially, Phillips et al. [21], also tested a second "acoustic" condition, where the VOT values for all the stimuli were increased by a constant, so that half standard stimuli (i.e. the frequent stimuli) were within the /d/ category and the other half were in the /t/ category. 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 condition, no mismatch was observed. This provides proof that by varying the standards within category, the generator of the MMN in the varying standards paradigm could only rely on phoneme-specific representations of the memory trace of the standards. Phillips et al suggested that this is direct evidence of categorical perception, superior to methods that rely on subjects making behavioral discriminatory responses to what are essentially phonetic distinctions.

Eulitz and Lahiri [12] employed the "varying standards" MMN paradigm to test even more detailed hypotheses about the nature of phonemic representations, specifically whether phonemes are underspecified. In what follows, we will illustrate the logic of Eulitz and Lahiri's claim. They suggest that front vowels are phonetically specified for the feature [CORONAL], but are phonologically underspecified for the feature, as [CORONAL] is phonologically unmarked; however, back vowels are both phonetically and phonologically marked for the feature [DORSAL]. Similarly, while rounded vowels are both phonetically and phonological specified for [LABIAL], unrounded vowels are specified neither phonologically or phonetically for any rounding features. So, in a language like German, the rounded mid front vowel [ø] is phonological specified for rounding but is phonologically underspecified for coronality (or frontness); thus it has the phonological representation [LABIAL]. In comparison, the rounded mid back vowel [o] is phonologically specified for both dorsality (or backness) and rounding; thus it has the phonological representation [DORSAL, LABIAL]. Depending on which sound is the standard and which is the deviant, underspecification leads to a surprising prediction for what counts as a mismatch. If the vowel [o] is the standard, since it is phonologically specified for [DORSAL, LABIAL], an incoming deviant [ø] will cause a mismatch – as there is a feature incompatibility between the standard phonemic representation [DORSAL, LABIAL] and the deviant acoustic input [CORONAL, LABIAL]. However, if the vowel [ø] is the standard, since it is only phonologically specified for [LABIAL], an incoming deviant [o] will not cause a mismatch – as there is no feature incompatibility between the phonemic representation [LABIAL] and the deviant acoustic input [DORSAL, LABIAL].[3] The logic of the argument and the results were substantiated for German vowels by [12], [22], [23], and the method has also been applied to the study of contrasts involving the place of articulation features distinguishing coronals from labials

---

[3] This can be formalized by modeling the memory trace comparison as feature unification, which is basically set union: $\{[+F]\} \cup \{[-F]\} = \{[+F], [-F]\}$ (a contradiction), but $\{[+F]\} \cup \varnothing = \{[+F]\}$ [48].

[24], [25].

1.3 The current study

While previous underspecification MMN studies showed that underspecification is a possible explanation for 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 observe a novel underspecification asymmetry in the voicing contrast in English, and we show that the asymmetry obtains only when the experimental conditions force the MMN generator to access phonological representations. In contrast, 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 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.

We demonstrate this by using 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 [26]–[29], but see Hwang et al. [30] for an opposing view. In order to test this underspecification hypothesis using the Eulitz/Lahiri experimental logic, we implemented an exact replication of Experiment 1 in [21][4], which tests the effect of presenting a stream of [t]

---

[4] In their study which employed MEG to demonstrate a phoneme-specific MMN generator, Phillips et al. [21] 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.

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. [21], 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 [12], 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/. 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 (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) is presented, the token stimulus [d] is phonetically represented as [-spread] and is compared to the memory representation of the phoneme /t/ specified as [+spread]. The direct contradiction of feature values (solid line in Figure 1) is predicted to generate an MMN response.

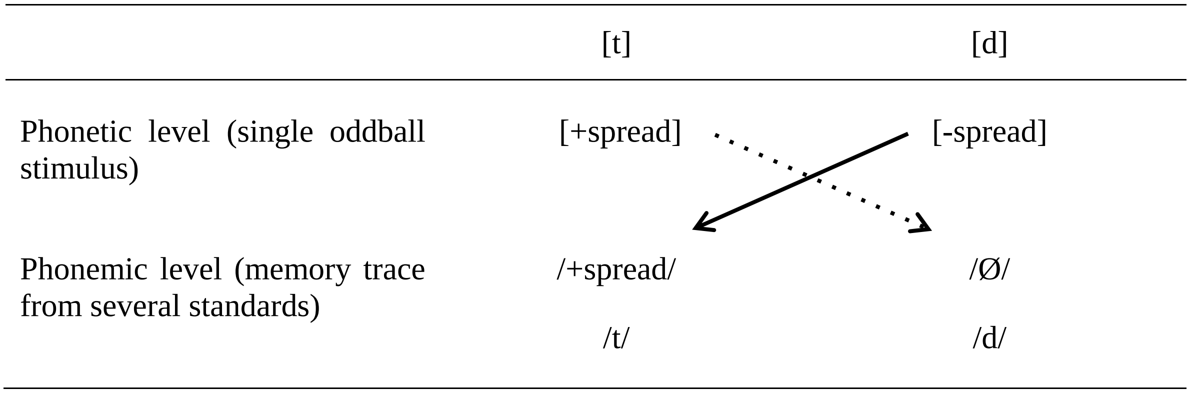


Figure 1: MMNs as per FUL. Ø = unspecified. Arrows indicate which pairs are compared by the perceptual mechanisms. Dotted arrow represents the "no mismatch" comparison.

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.

Critically, if the standards are not varied, there is no need for the memory trace to rely on a phonemic representation: If the exact same phonetic token is presented as the standard stimulus every time, a phonetic representation suffices as the memory representation. In this case, a fully specified phonetic token in the memory representation is always compared to a fully specified oddball stimulus. As both elements are fully specified, there will always be a direct feature conflict, irrespective of whether [t] or [d] is the oddball. Note that if the asymmetry that was observed in the phonemic condition was in fact due to a phonetic difference between the stimuli, then the same asymmetry should be observed again. This provides the test that was missing from previous underspecification studies.

We tested these predictions in three experiments. Experiment 1 tested the basic prediction of a different amplitude in MMN for /d/ vs. /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. Finally, in Experiment 3 we forced the MMN generator to rely only on phonetic information, by not varying the standards within category. This predicts that if phonetic information is used for comparing standards to deviants, there should be no asymmetry, because the MMN generator need not engage phonemic memory representations.

2. EXPERIMENT 1

2.1 Methods

2.1.1 Subjects

A total of 35 University of Delaware students were recruited as subjects and enrolled 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. [21], and were synthesized with an online version of the low-level Klatt synthesizer [31], [32]. In addition, we constructed target stimuli for a tracking task consisting of a male and female voice saying “ba” with different fundamental frequencies than the MMN stimuli.

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; we observed that the peak latency of the main effect P2 wave was about 16ms later for /t/ than for /d/, which most likely is related to the difference in mean VOT of the stimuli. A peak latency difference can give rise to a “fake” MMN. 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.

		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. [21]). This was done in a pre-test with a forced choice behavioral identification task using the stimuli described above. Six trials of each 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. 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 selected for that subject. The stimuli were delivered continuously, with a random number (between 2 and 7) of standards between each deviant. Each stimulus syllable lasted 290ms. 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 “Skype” emoticons, so that by the end of the experiment, the whole screen would be filled up. This provided some entertainment 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, cf. [33]. 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 prestimulus 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 [34] 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 [35], [36].

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, using difference waves as the dependent measure allows us to use Principal Component Analysis and Independent Component Analysis to focus on the temporal and spatial fluctuations of the mismatch effect itself rather than on the temporal and spatial distribution of the main effect amplitude change caused by general sound processing mechanisms.

MMN is typically observed as a negativity central on the scalp and during the 100-300ms time window. We first identified the temporal and spatial distribution of the experimental effect by conducting a sequential temporo-spatial PCA [34], [37], [38], and then used the results of the PCA to inform the selection of time windows and electrode regions in the raw voltage data. Each experiment was therefore analyzed both via PCA and via "traditional" analysis of the scalp-recorded voltages. We direct the reader to the literature on PCA decomposition [34], [37], [38] for a full justification and explanation of temporo-spatial PCA.

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 (from a subject pool of 135 subjects; SD= 5ms). 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, but we did not exclude left-handers, as most left-handed people have left-lateralized language function. 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 in each temporal factor[5]. 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

---

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

the scree test in combination with the Parallel Test [35], 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 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 [37]). 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 [34], [37]). 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. 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 [37]). For TF1, the first spatial factor TF1SF1[6] 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

---

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

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 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. 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:

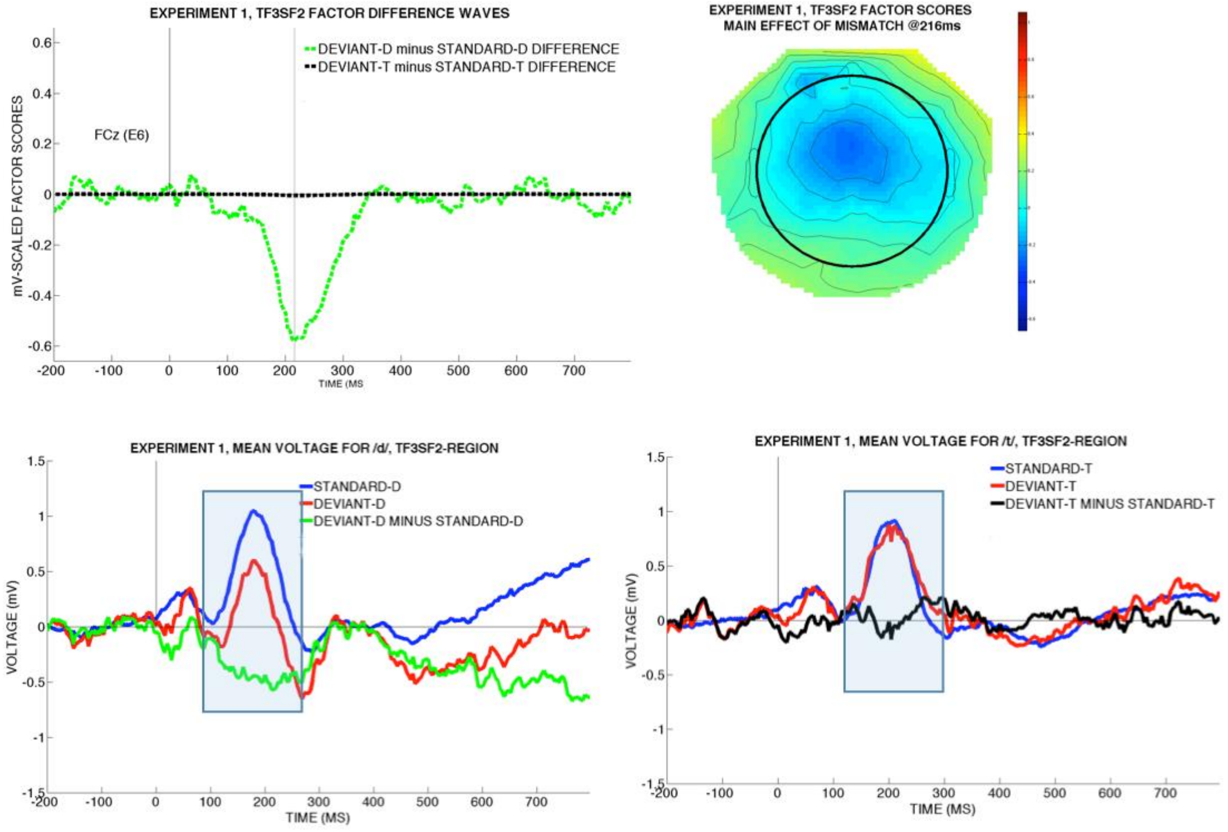


Figure 2: Experiment 1, early MMN effect. Upper panel: Time course (left) and spatial

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 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 (F1,22)=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.

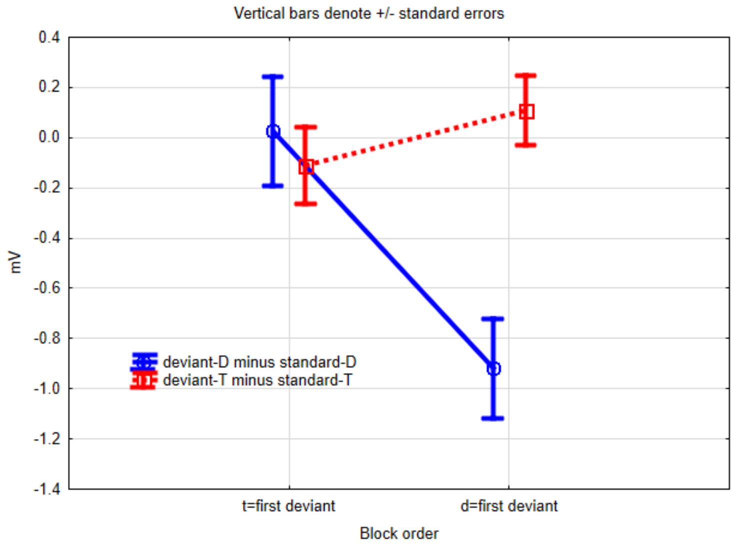


Figure 3: Interaction plot PHONEME (difference wave) x BLOCKORDER. Vertical bars denote +/-standard error.

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).

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 [16], [39]–[43], 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.

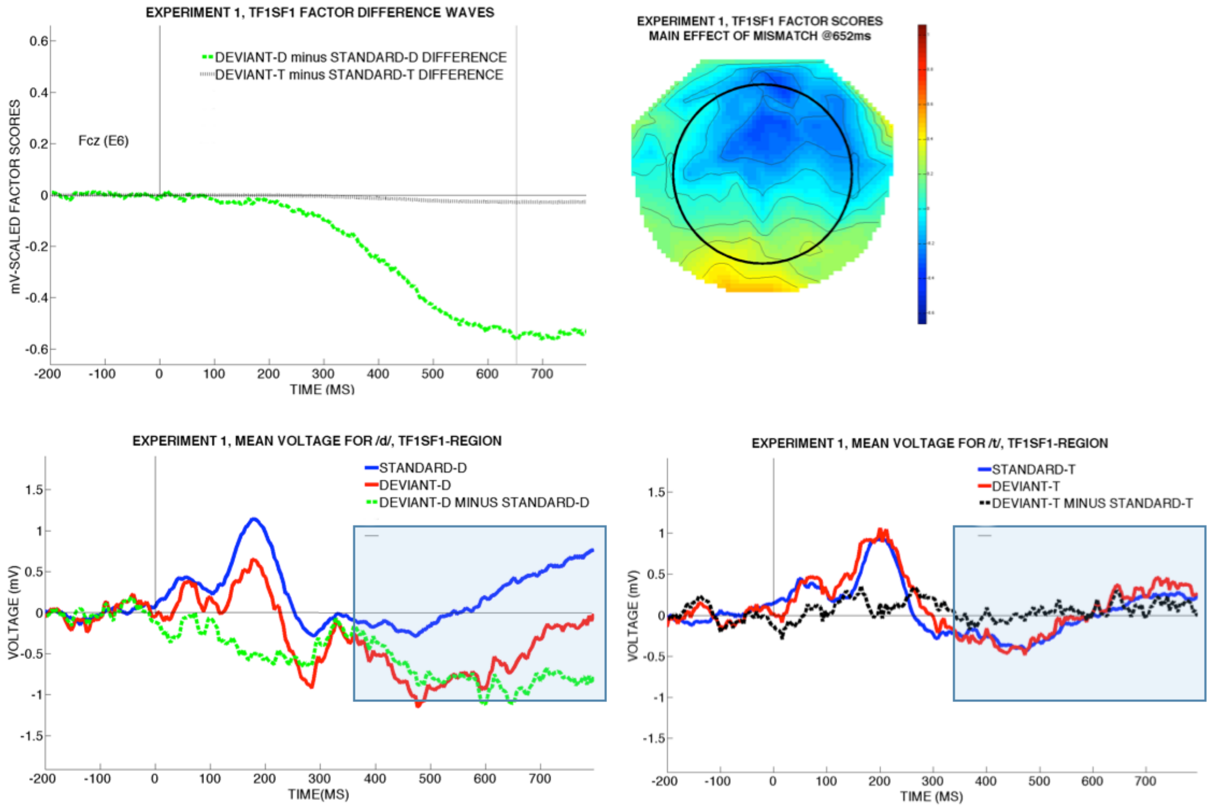


Figure 4: Experiment 1, Late Discriminative Negativity. Upper panel: Time course (left) and spatial distribution (right) of temporo-spatial factor decomposition; the topoplot shows the mean difference wave at the horizontal line in the waveform plot at 652ms. Lower panel: corresponding absolute voltage 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 [21].

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 enrolled in Experiment 2. Each subject received course credit for participation. Three subjects were excluded from analysis because they measured with 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.

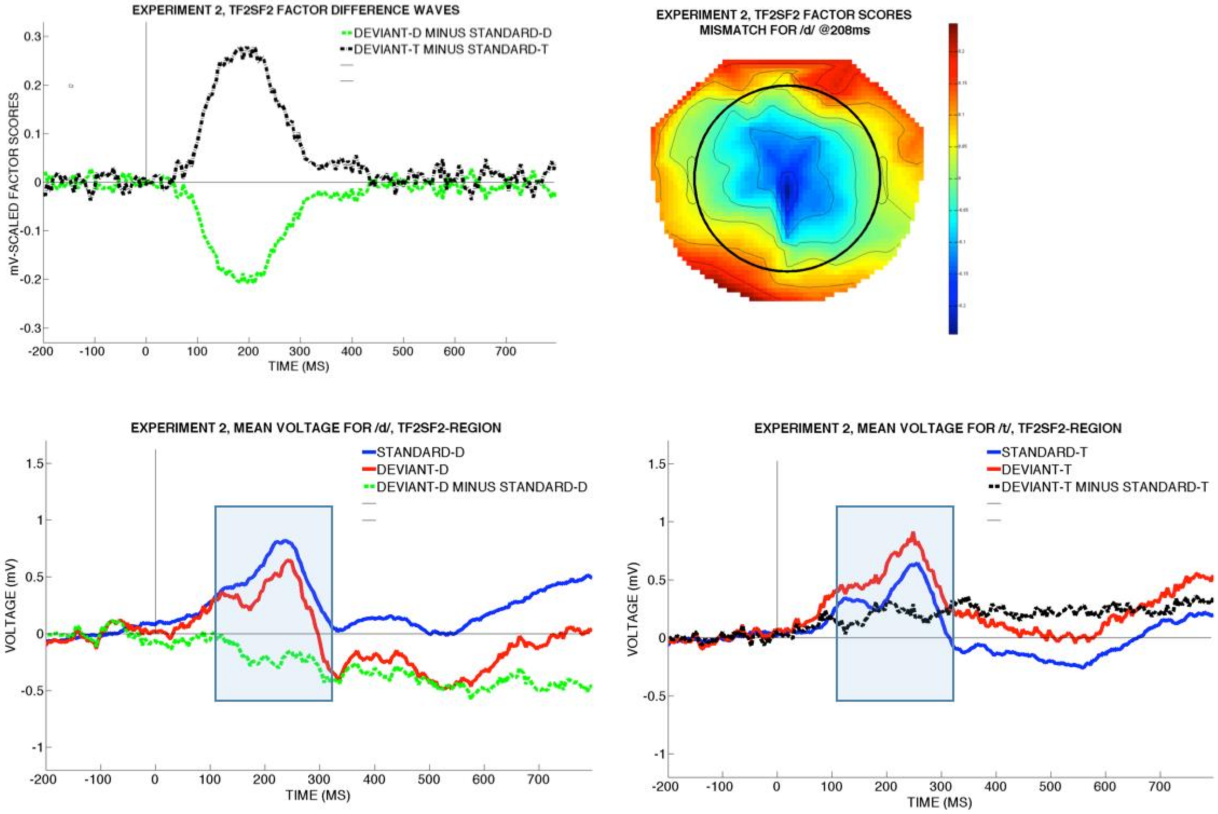


Figure 5: Experiment 2 MMN. Upper panel: Time course (left) and spatial distribution (right) of temporo-spatial factor decomposition; the topoplot shows the mean difference wave at the horizontal line in the waveform plot at 208ms. Lower panel: corresponding absolute voltage waves and difference waves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

This analysis was replicated with an ANOVA of the 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).

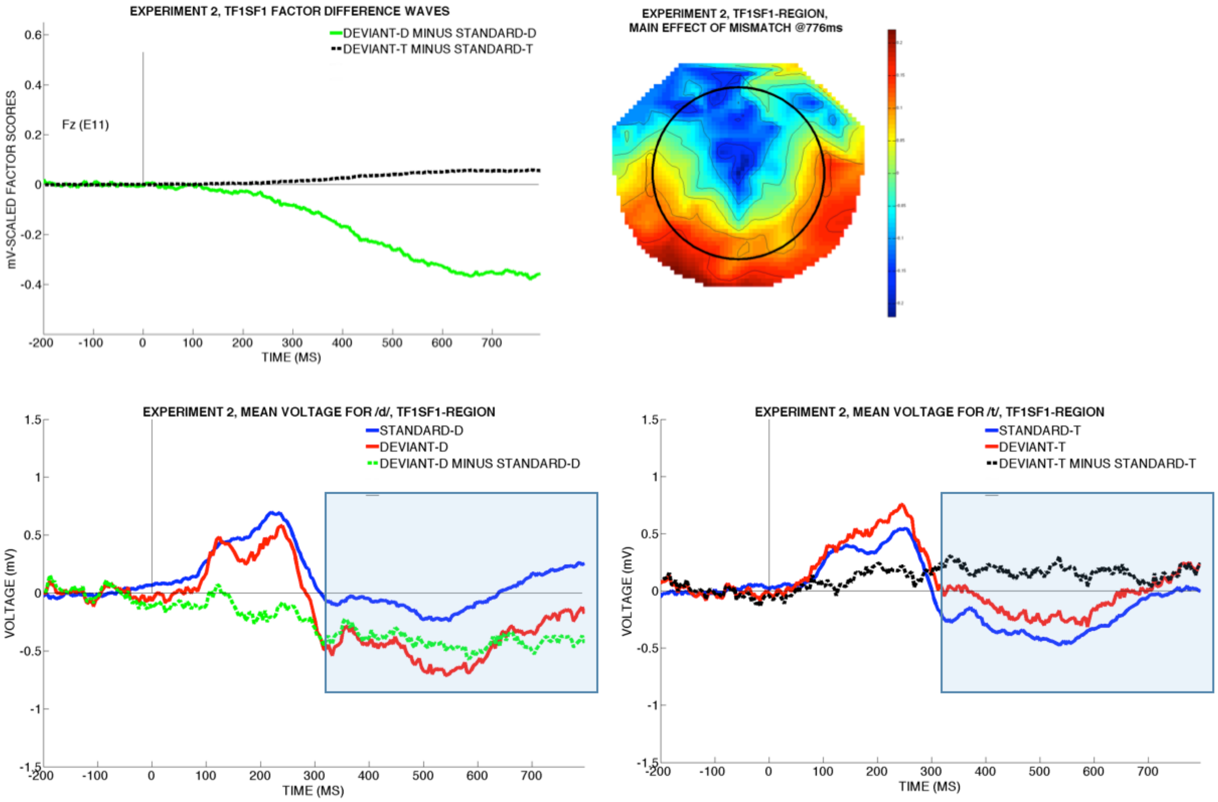


Figure 6: Experiment 2 Late Discriminatory Negativity. Upper panel: Time course (left) and spatial distribution (right) of temporo-spatial factor decomposition; the topoplot shows the mean difference wave at the horizontal line in the waveform plot at 776ms. Lower panel: corresponding absolute voltage 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 0.4 (0.6 resulted in a small set of electrodes so we relaxed the criterion to sample from a larger area) 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. 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/.

This experiment also allows 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? Of relevance here is the finding reported by [44], based on intracranial recordings of the auditory cortex in monkeys, that /da/-/ta/ stimuli give rise to two different responses from auditory cortex. One response is time-locked to stimulus onset, and another response is time-locked to the onset of voicing. For stimuli with 0ms VOT, there is a single “on” response, whereas for 60ms VOT, there are two “on”-responses, one the same as for 0ms VOT (i.e. /d/), and the second response is delayed by an interval equal to the VOT (and presumably encodes onset of voicing). Thus, a single response corresponded to /d/ and a double response corresponded to /t/.

It is unknown (at least to us) how these neural response patterns (if present in humans) are related to MMN amplitude. Conceivably, a “single on” response could give rise to a bigger mismatch than the double-on response, because the neural response is more homogeneous. However, if this were true, we would expect it to have the same effect in the single token paradigm just outlined. This last experiment therefore asks two questions: The first is, can we validate the assumptions of the varying token paradigm that it encourages phoneme representation of the memory trace (and inversely, that a single token paradigm encourages the formation of an allophonic memory trace)? The second question is, do intrinsic neural response patterns to VOT differences give rise to the asymmetry in MMN observed in the first two experiments?

4.1 Methods

4.1.1 Subjects

A total of 36 University of Delaware students were enrolled 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 two fixed tokens. Subjects were engaged in the same behavioral tracking task as in

Experiment 1; i.e., tracking “ba”-syllables and making male/female decisions about the voices.

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. 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).

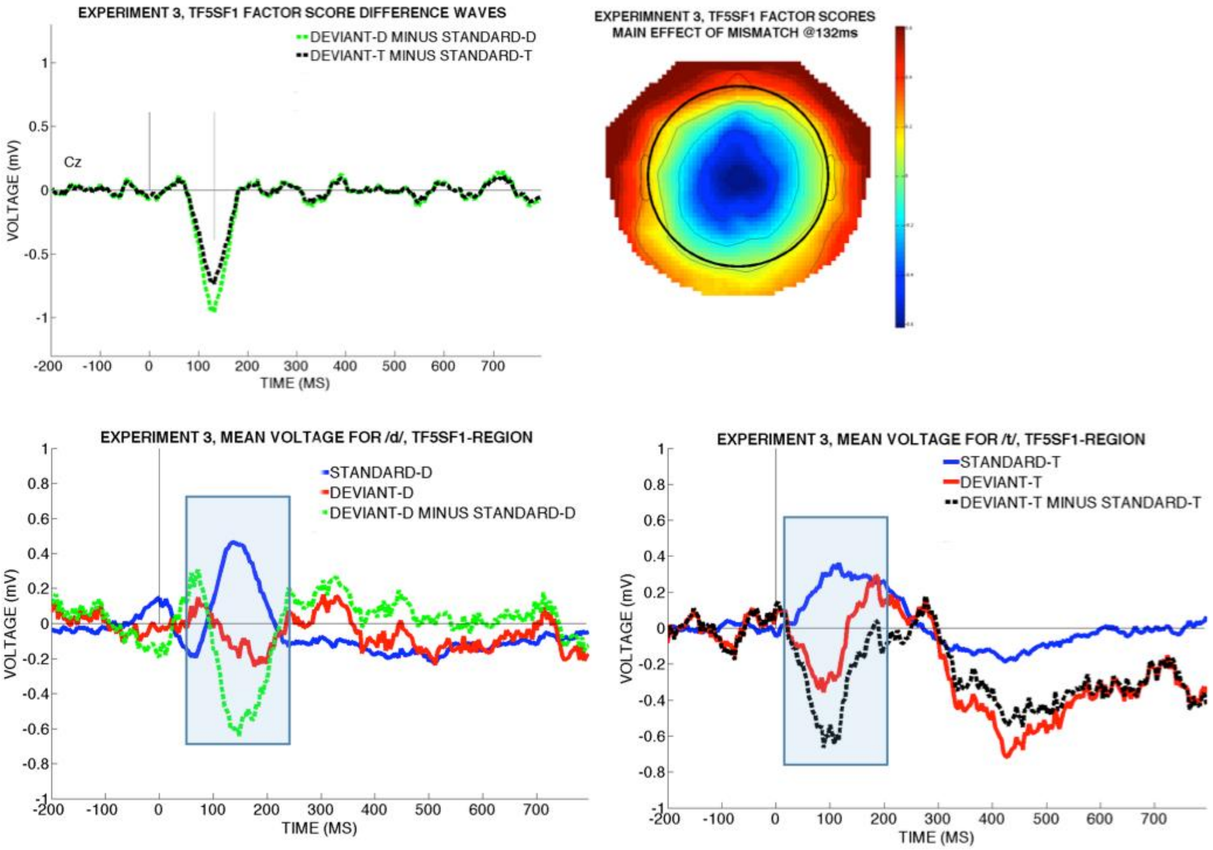


Figure 7: Experiment 3, phonetic MMN. Upper panel: Time course (left) and spatial distribution (right) of temporo-spatial factor decomposition; the topoplot shows the mean difference wave at the horizontal line in the waveform plot at 132ms. Lower panel: corresponding absolute voltage waves and difference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

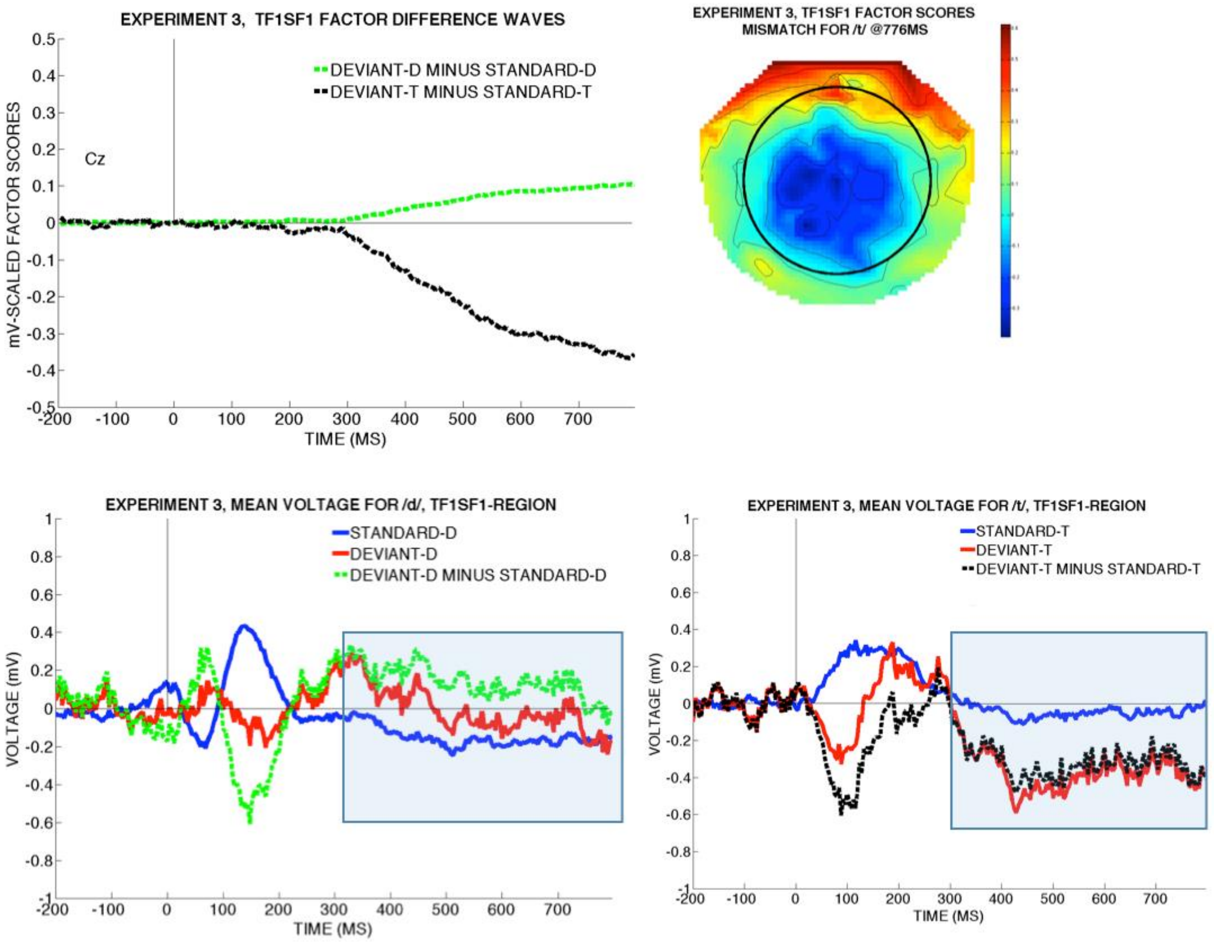


Figure 8: Experiment 3, Late Discriminatory Negativity for /t/. Upper panel: Time course (left) and spatial distribution (right) of temporo-spatial factor decomposition; the topoplot shows the mean difference wave at the horizontal line in the waveform plot at 776ms. Lower panel: corresponding absolute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

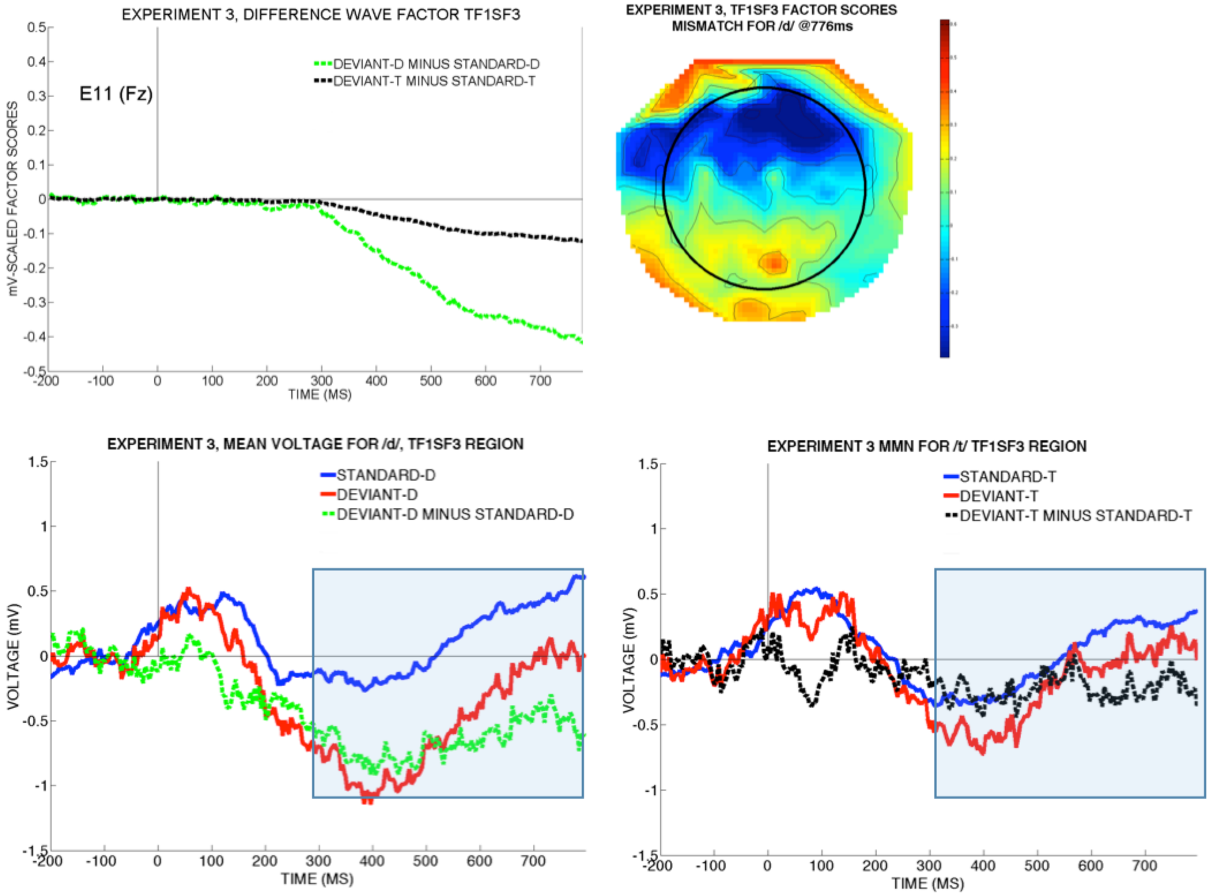


Figure 9: Experiment 3, Late Discriminatory Negativity for /d/. Upper panel: Time course (left) and spatial distribution (right) of temporo-spatial factor decomposition; the topoplot shows the mean difference wave at the horizontal line in the waveform plot at 776ms. Lower panel: corresponding absolute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

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.

3. CONCLUSIONS

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 addressed the possibility of the asymmetric results in Experiment 1 being due to the distractor task, and whether the asymmetry only appeared under attention to the auditory stream. 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 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.

Author Note

We acknowledge the generous support of this research by a University of Delaware UNIDEL grant to the first author. We also thanks 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).

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

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

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

[6] 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.

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

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

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

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

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] S. Ylinen, “Cortical Representation for Phonological Quantity,” University of Helsinki, Helsinki, 2006.

[21] 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.

[22] 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.

[23] 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.

[24] 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.

[25] M. A. Walter and V. Hacquard, “MEG Evidence for Phonological Underspecification,” 2004.

[26] 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.

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

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

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

[30] 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.

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

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

[33] 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.

[34] 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.

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

[36] 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.

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

[38] 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.

[39] 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.

[40] 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).

[41] 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.

[42] 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.

[43] 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.

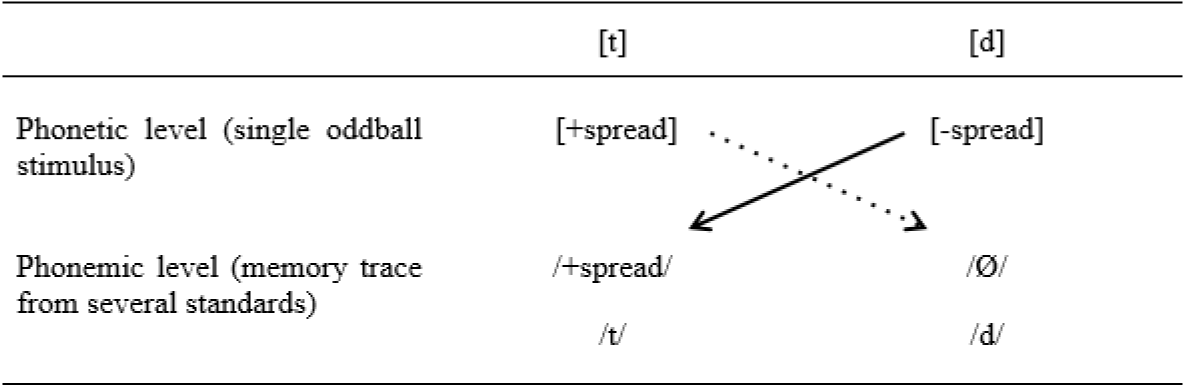
[44] M. Steinschneider, C. E. Schroeder, J. C. Arezzo, and H. G. Vaughan, “Physiologic Correlates of the Voice Onset Time Boundary in Primary Auditory Cortex (A1) of the Awake Monkey: Temporal Response Patterns,” *Brain Lang.*, vol. 48, no. 3, pp. 326–340, 1995.

[45] J. Pierrehumbert, “Exemplar dynamics: Word frequency, lenition and contrast,” in *Frequency effects and the emergence of linguistic structure*, J. Bybee and P. Hoppe, Eds. Amsterdam: John Benjamins, 2001, pp. 137–157.

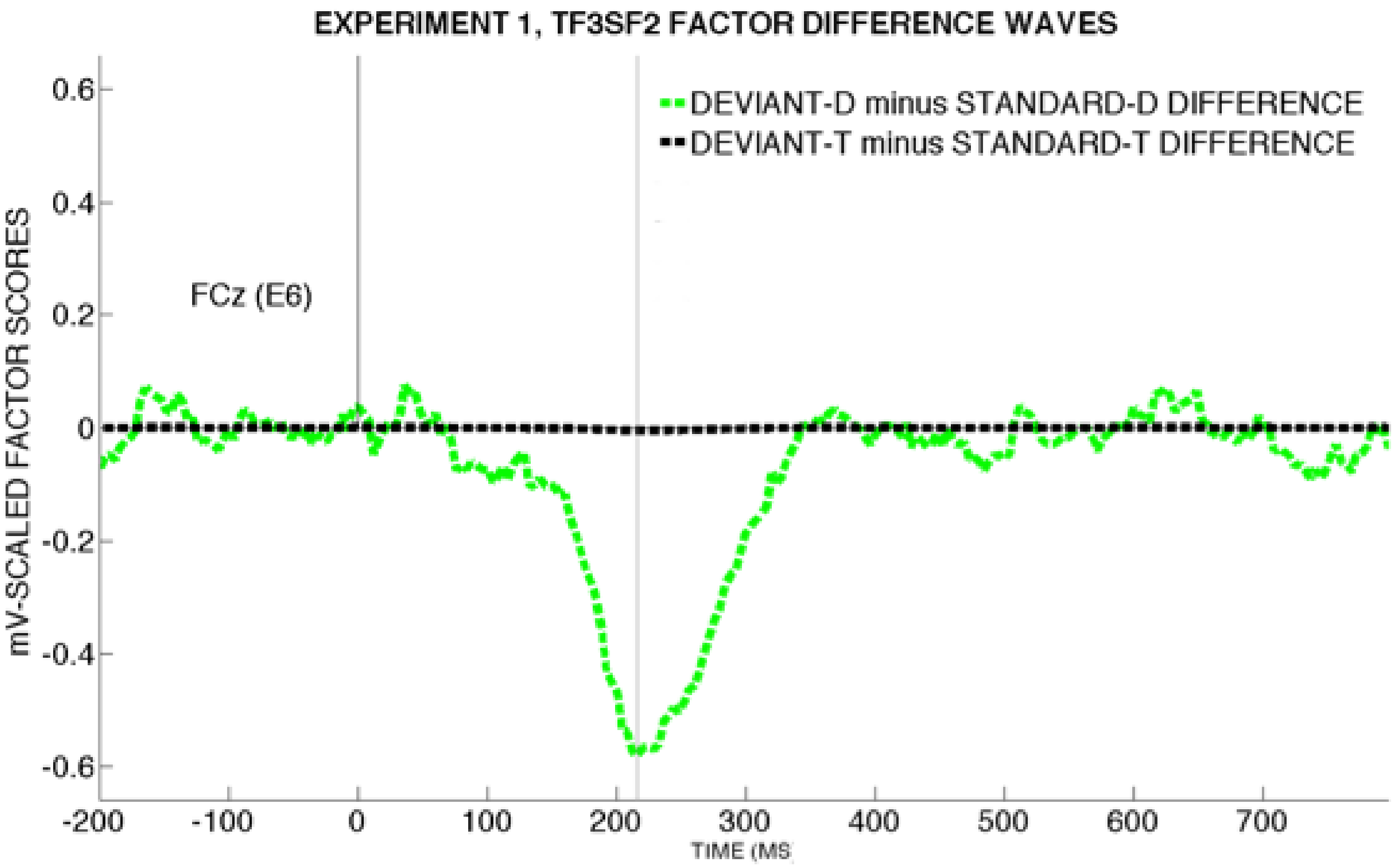
[46] J. M. Anderson and C. J. Ewen, *Principles of Dependency Phonology*. Cambridge: Cambridge University Press, 1987.

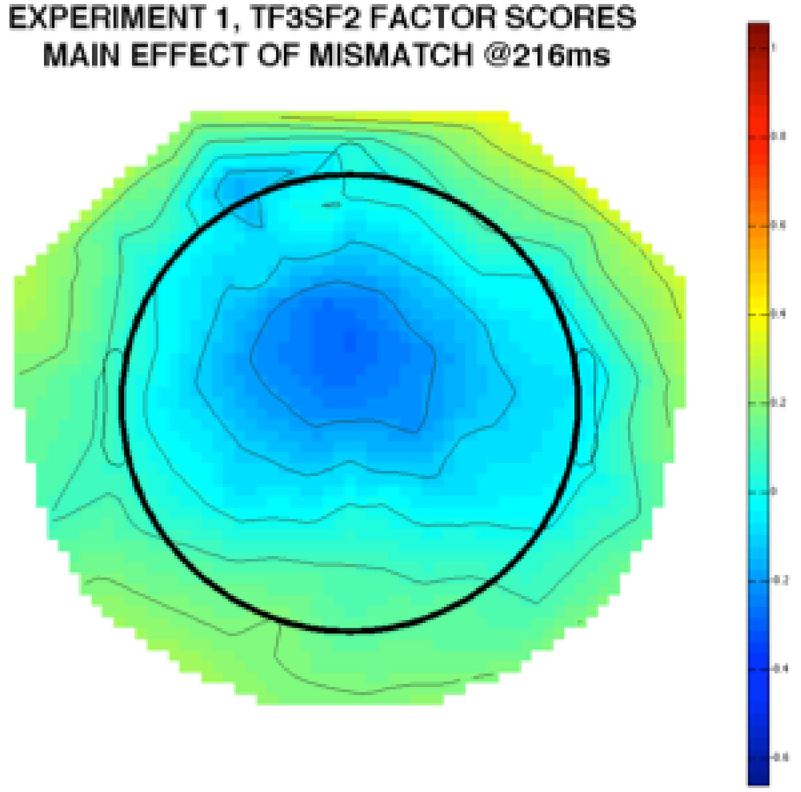
[47] J. Kaye, J. Lowenstamm, and J.-R. Vergnaud, “The Internal Structure of Phonological Elements: A Theory of Charm and Government,” *Phonol. Yearb.*, vol. 2, pp. 305–328, 1985.

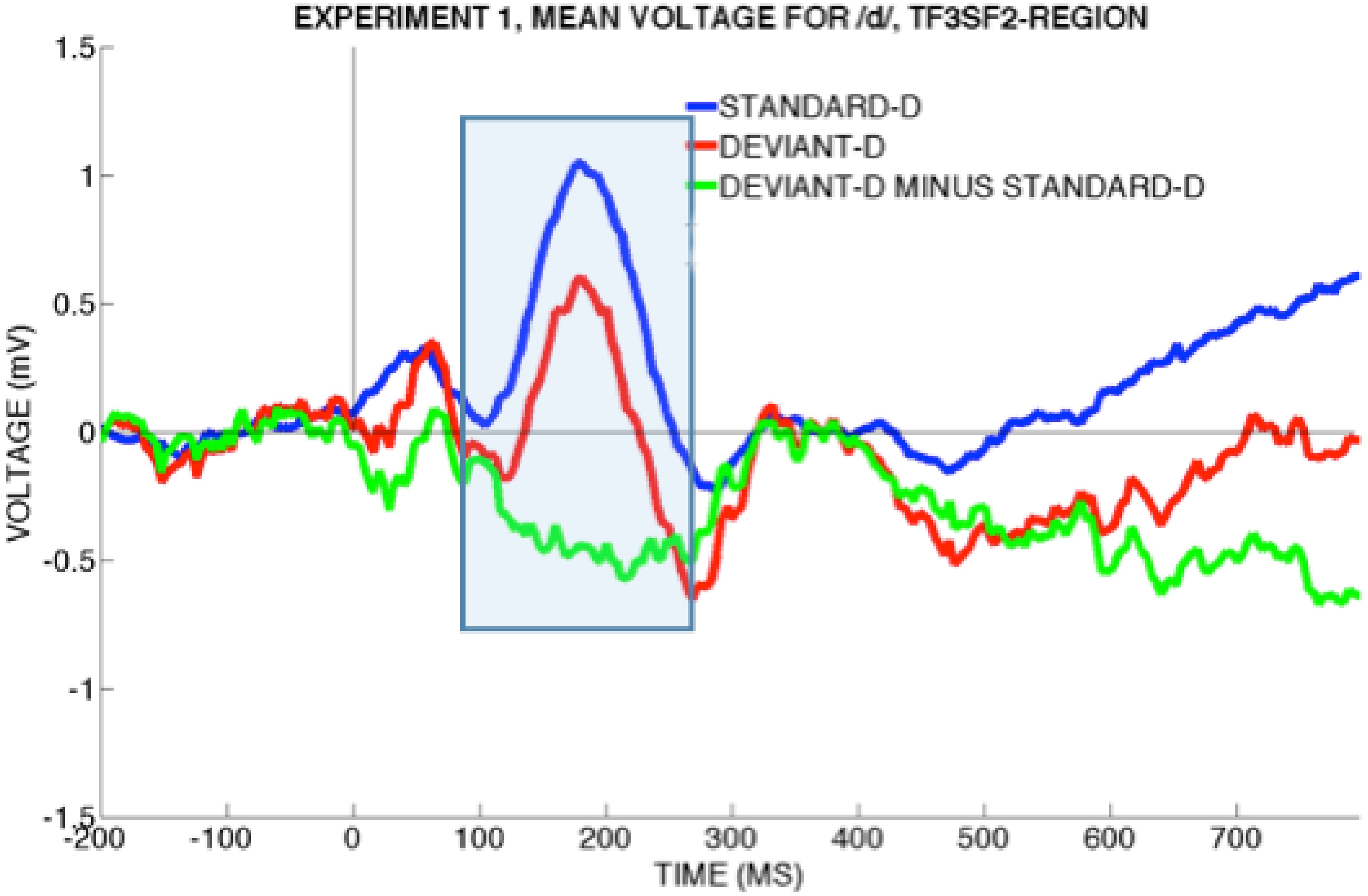
[48] S. M. Shieber, *An Introduction to Unification-Based Approaches to Grammar*, vol. 4. Stanford, CA: Center for the Study of Language and Information, 1986.

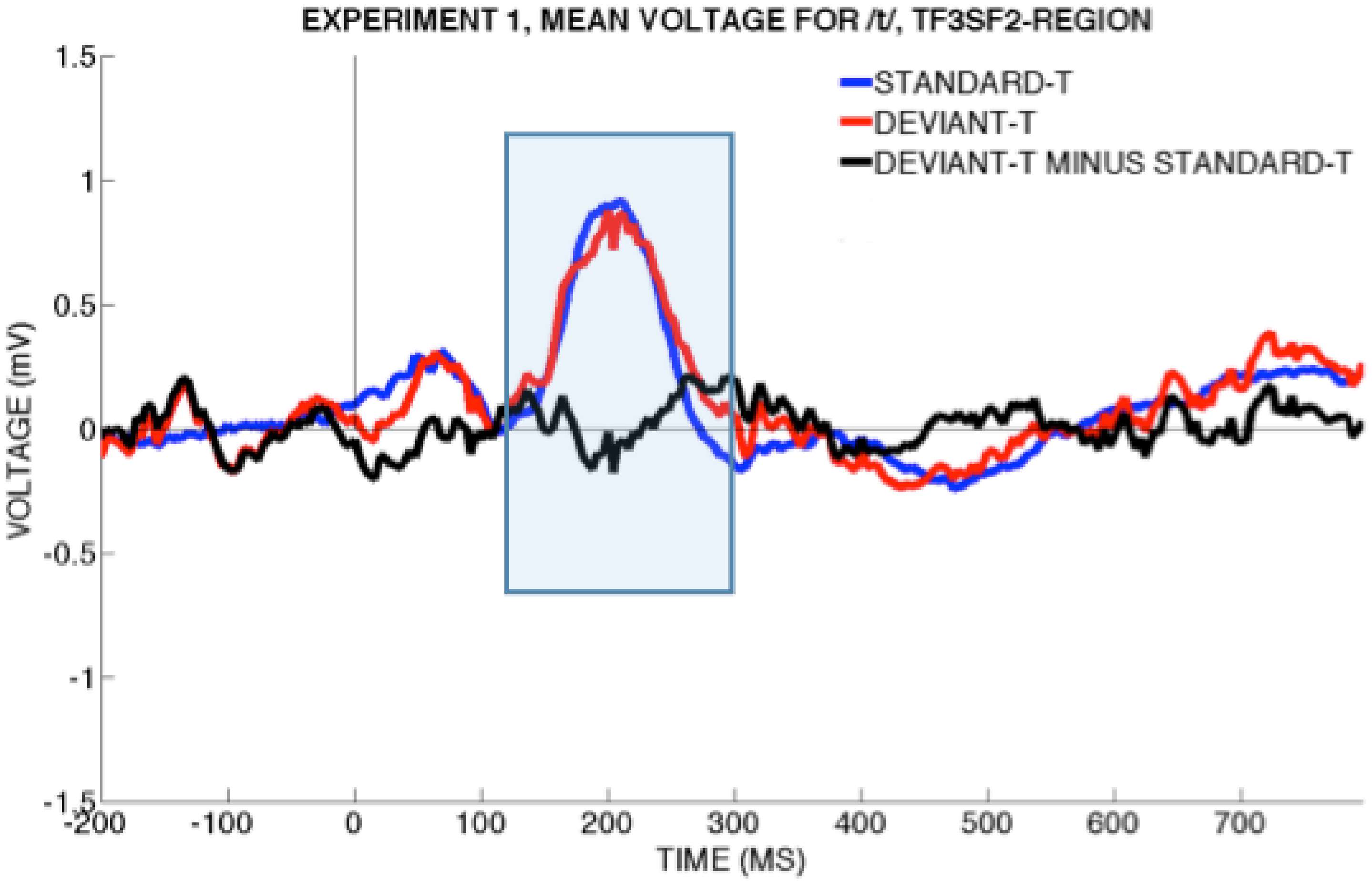


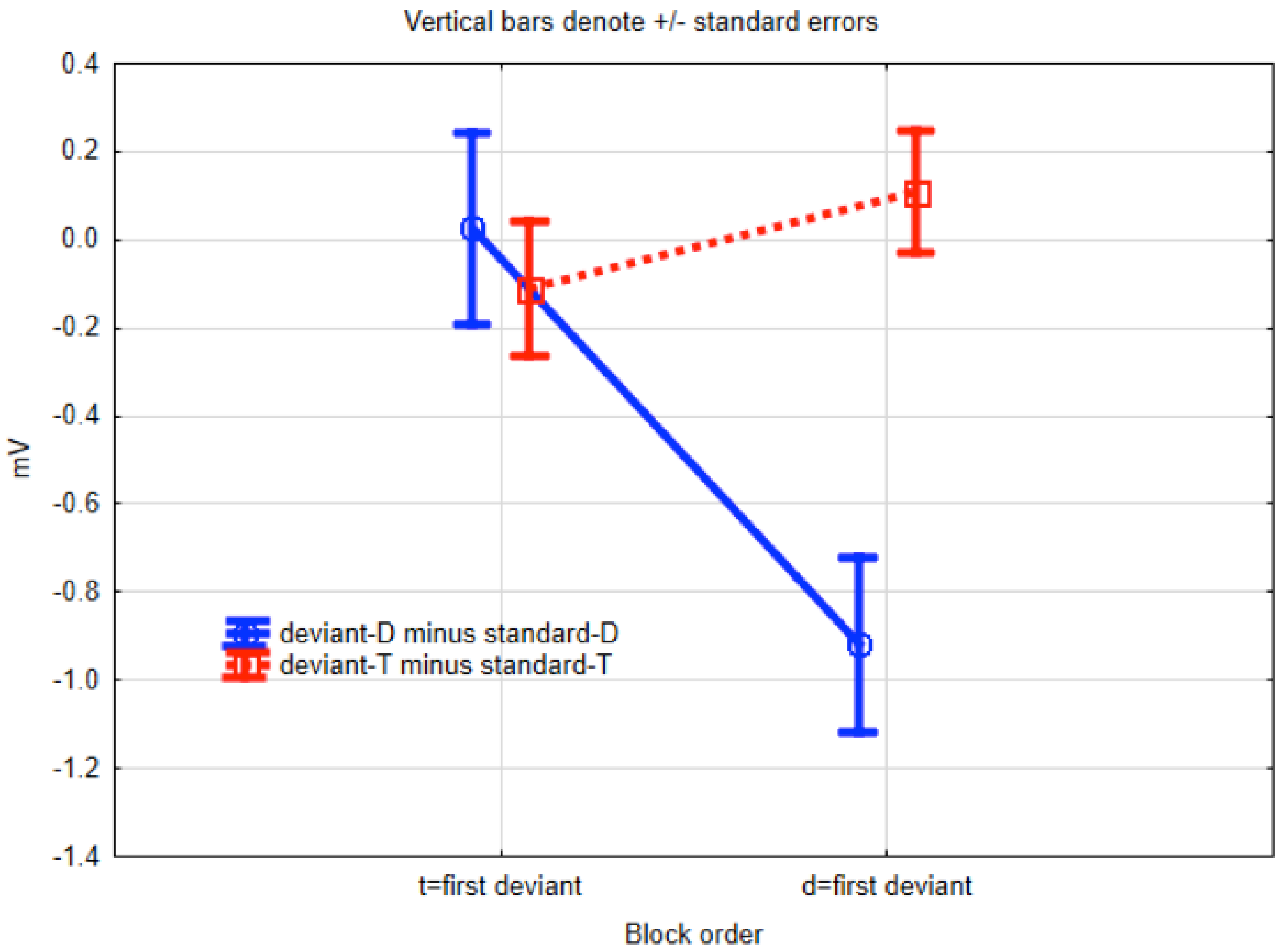
	[t]	[d]
Phonetic level (single oddball stimulus)	[+spread]	[-spread]
Phonemic level (memory trace from several standards)	/+spread/ /t/	/Ø/ /d/

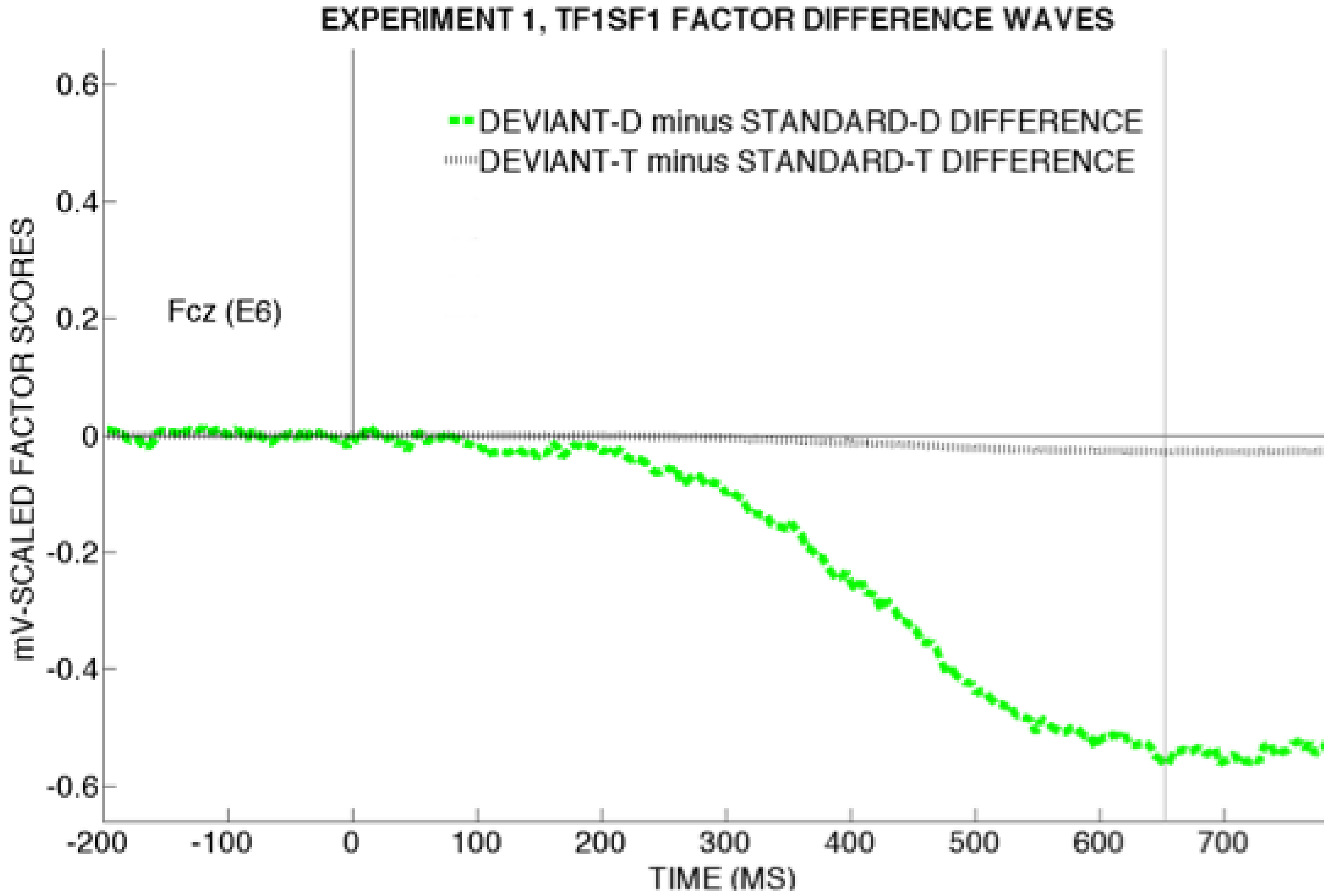

EXPERIMENT 1, TF3SF2 FACTOR DIFFERENCE WAVES
DEVIANT-D minus STANDARD-D DIFFERENCE
DEVIANT-T minus STANDARD-T DIFFERENCE
FCz (E6)
mV-SCALED FACTOR SCORES
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-200
-100
0
100
200
300
400
500
600
700
TIME (MS

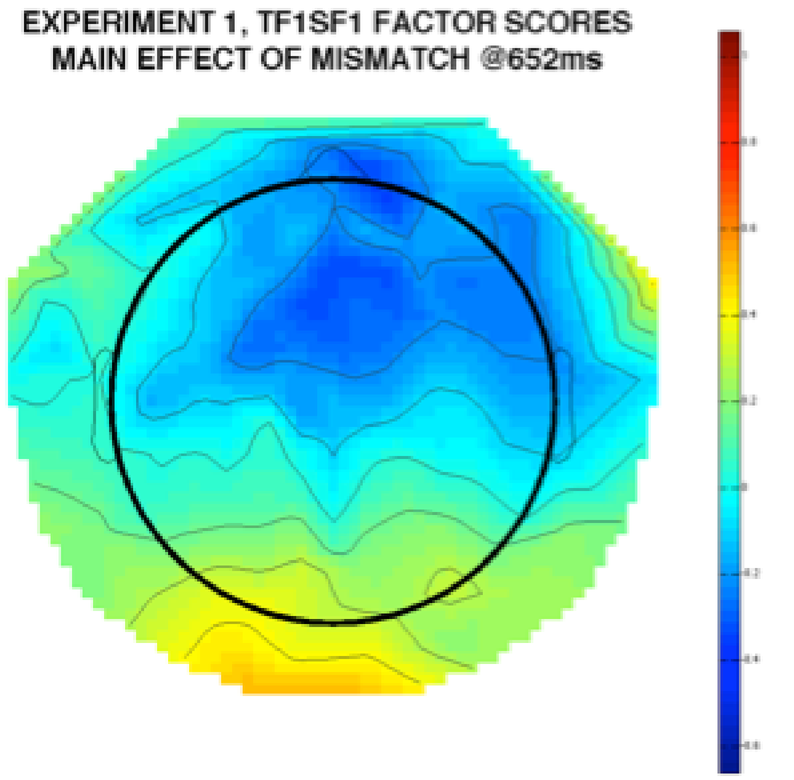

EXPERIMENT 1, TF3SF2 FACTOR SCORES
MAIN EFFECT OF MISMATCH @216ms

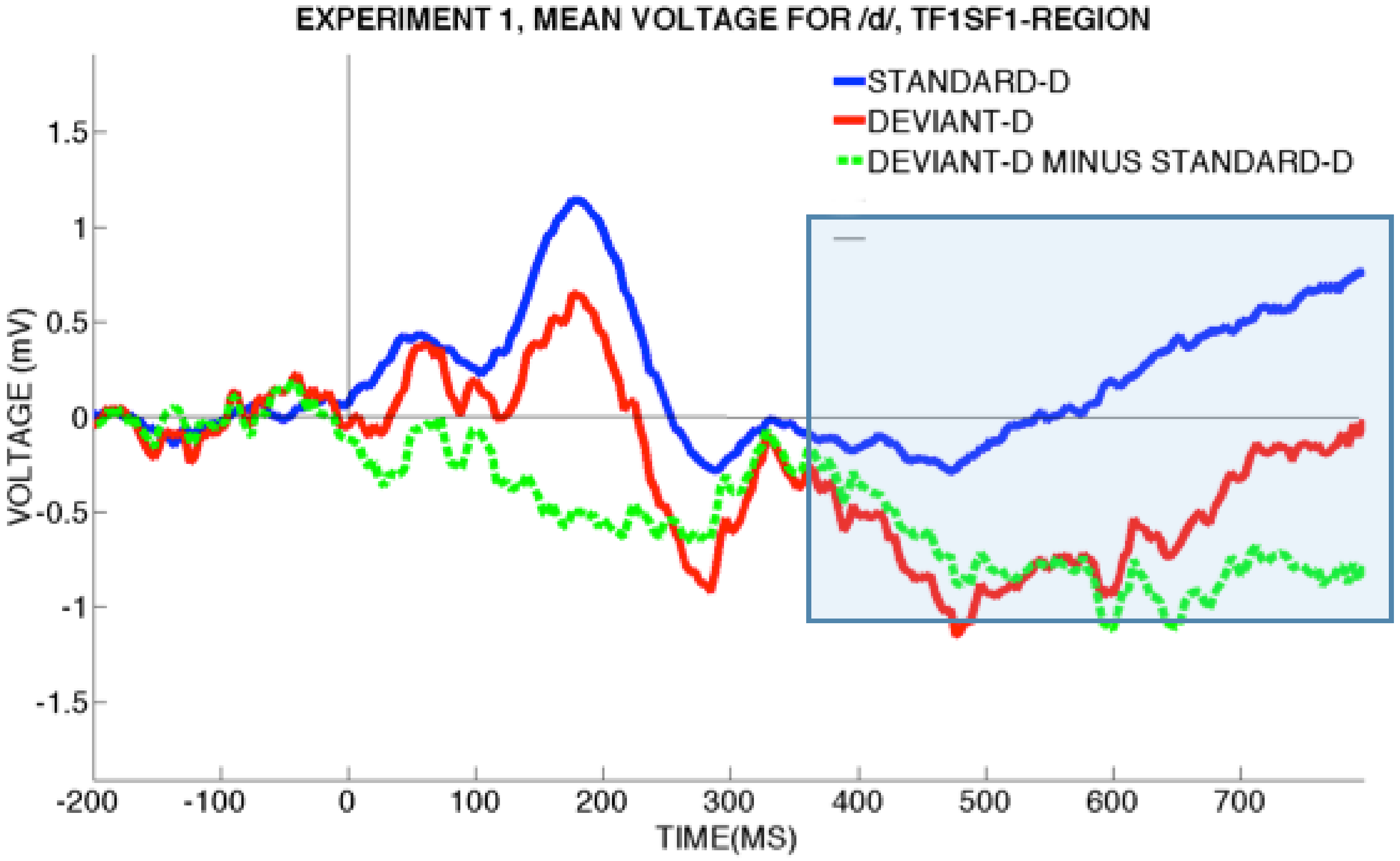

EXPERIMENT 1, MEAN VOLTAGE FOR /d/, TF3SF2-REGION
STANDARD-D
DEVIANT-D
DEVIANT-D MINUS STANDARD-D
VOLTAGE (mV)
1.5
1
0.5
0
-0.5
-1
-1.5
-200
-100
0
100
200
300
400
500
600
700
TIME (MS)

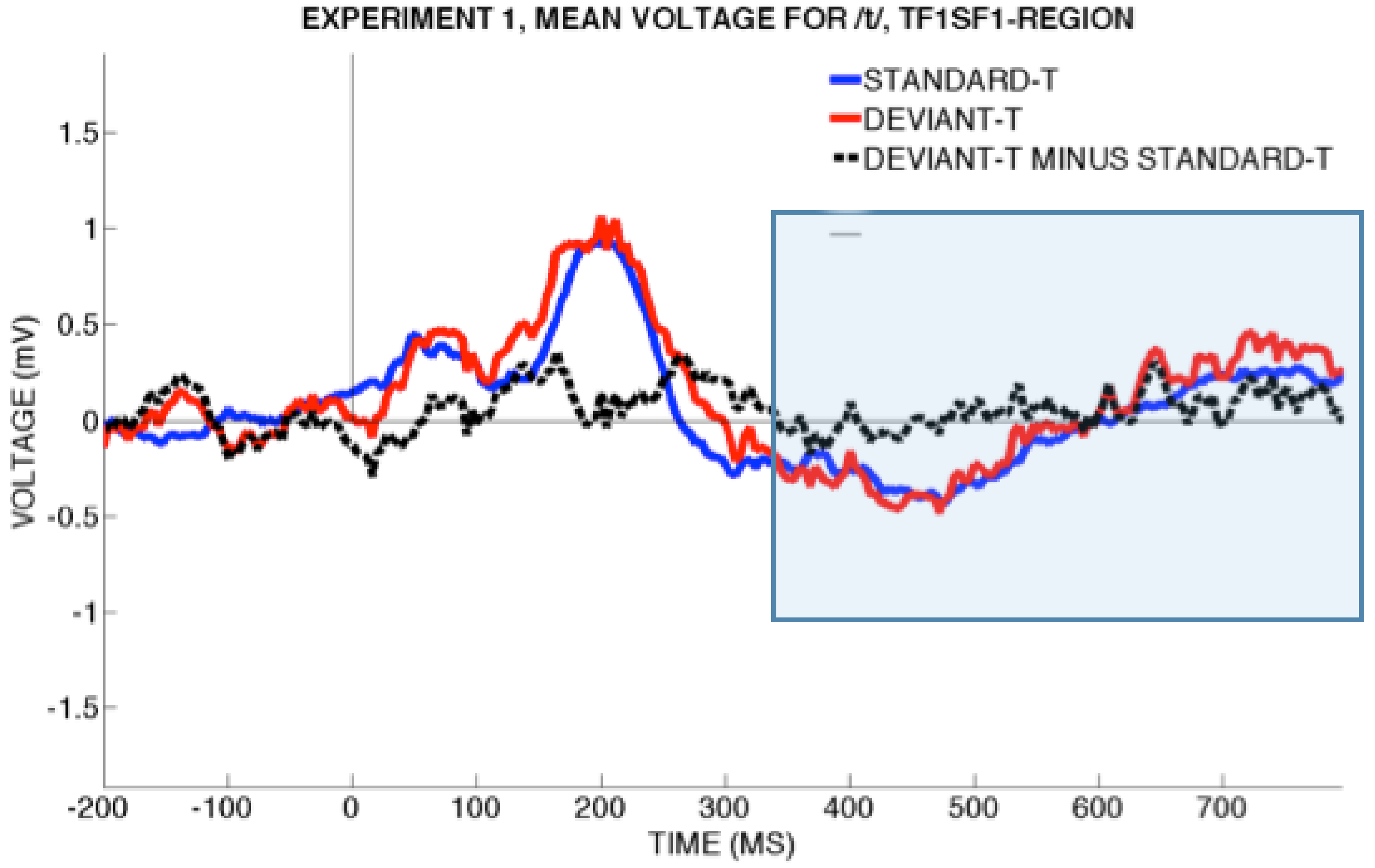

EXPERIMENT 1, MEAN VOLTAGE FOR /t/, TF3SF2-REGION
STANDARD-T
DEVIANT-T
DEVIANT-T MINUS STANDARD-T
VOLTAGE (mV)
1.5
1
0.5
0
-0.5
-1
-1.5
-200
-100
0
100
200
300
400
500
600
700
TIME (MS)

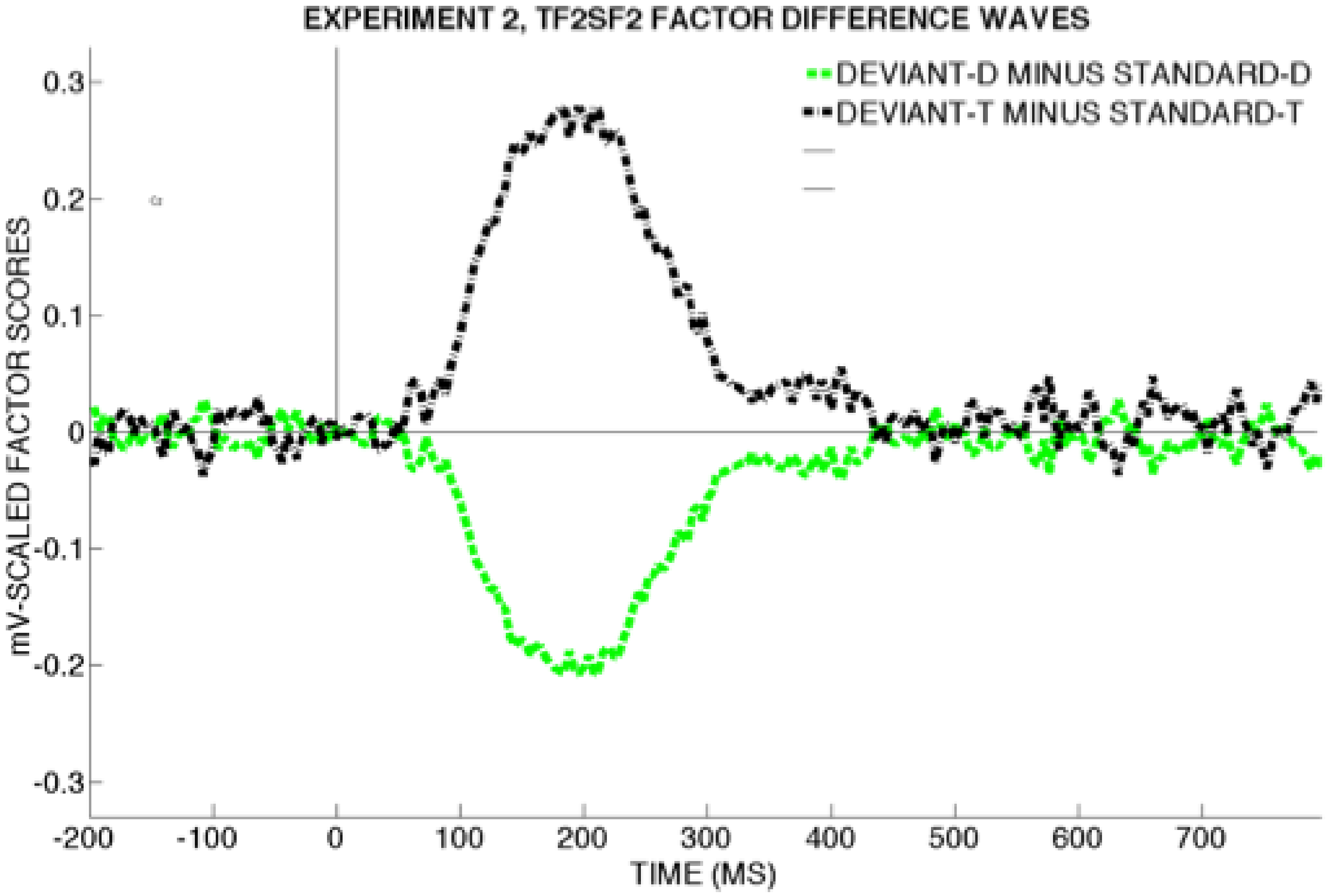

Vertical bars denote +/- standard errors
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.2
-1.4
mV
deviant-D minus standard-D
deviant-T minus standard-T
t=first deviant
d=first deviant
Block order

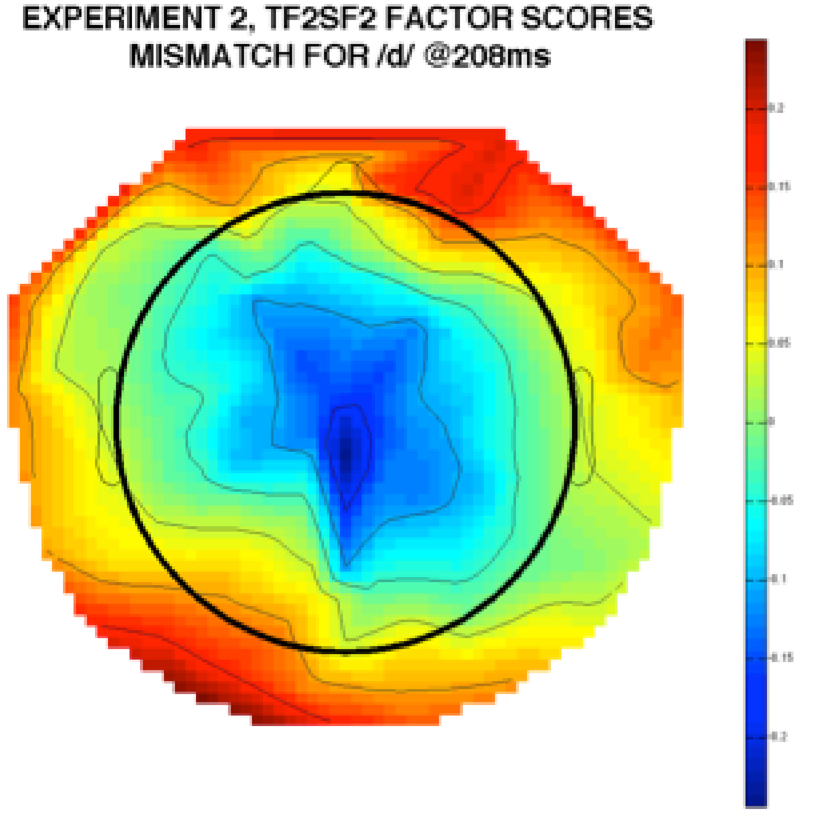

EXPERIMENT 1, TF1SF1 FACTOR DIFFERENCE WAVES

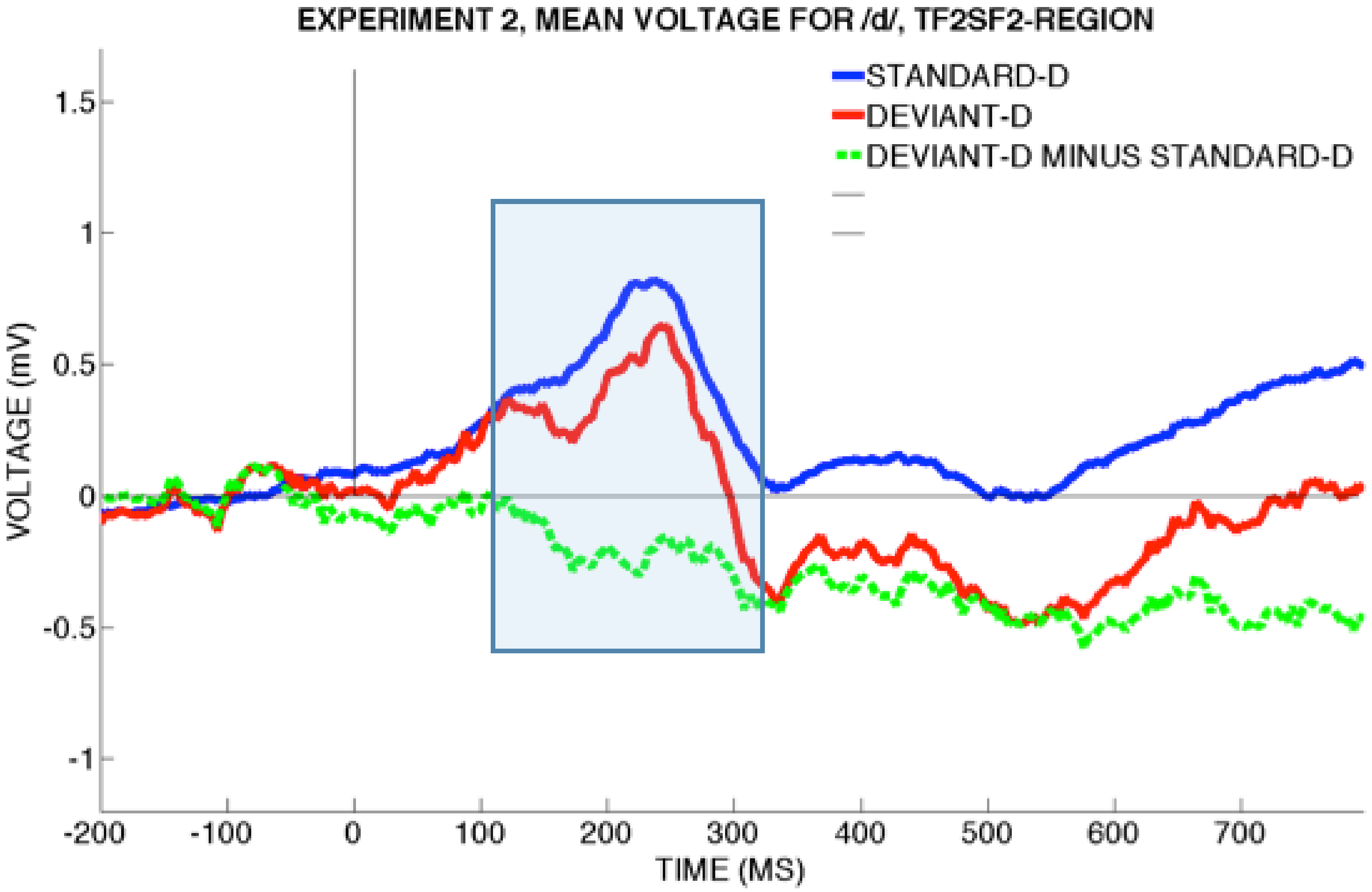

EXPERIMENT 1, TF1SF1 FACTOR SCORES
MAIN EFFECT OF MISMATCH @652ms

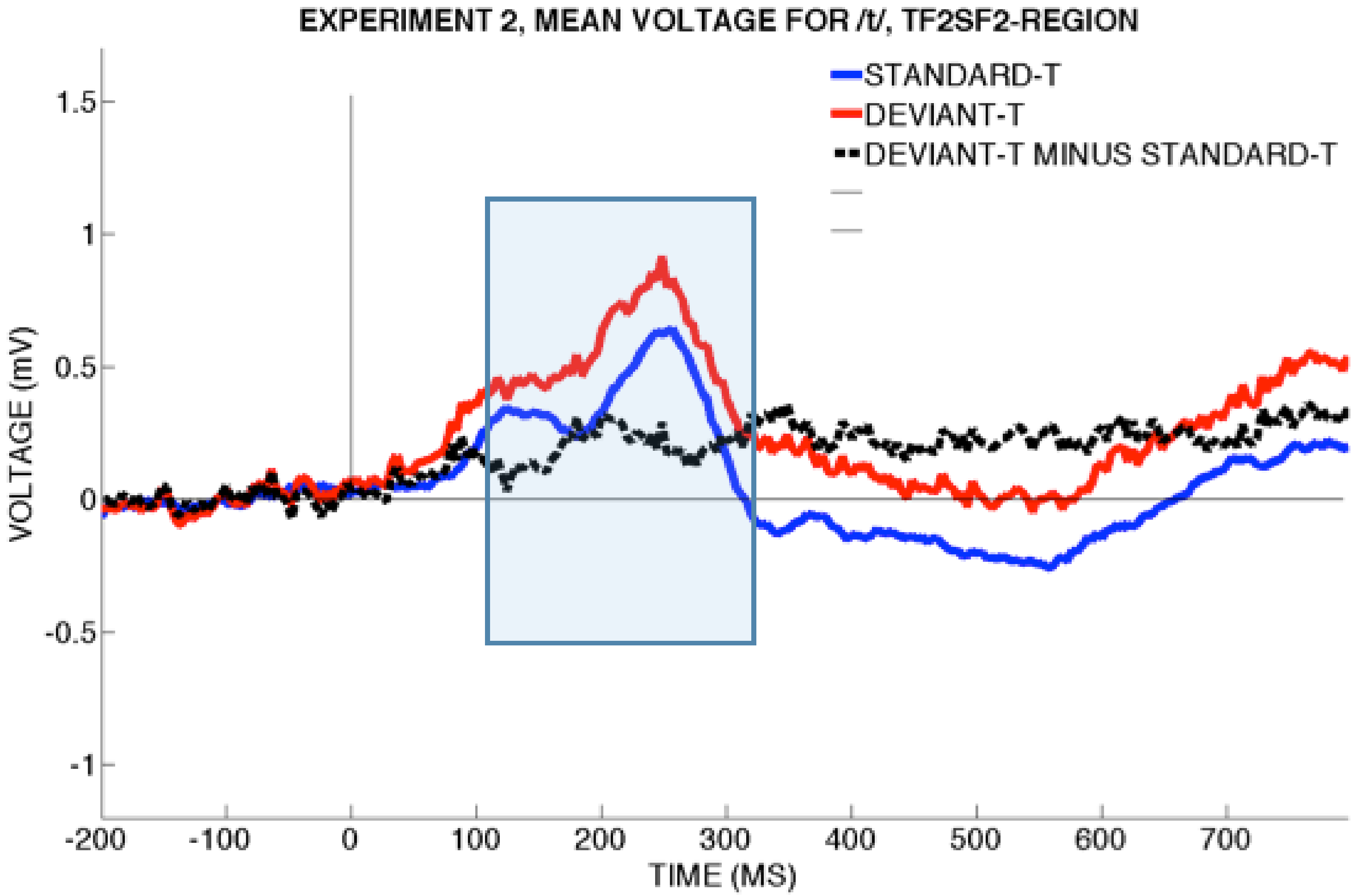

EXPERIMENT 1, MEAN VOLTAGE FOR /d/, TF1SF1-REGION

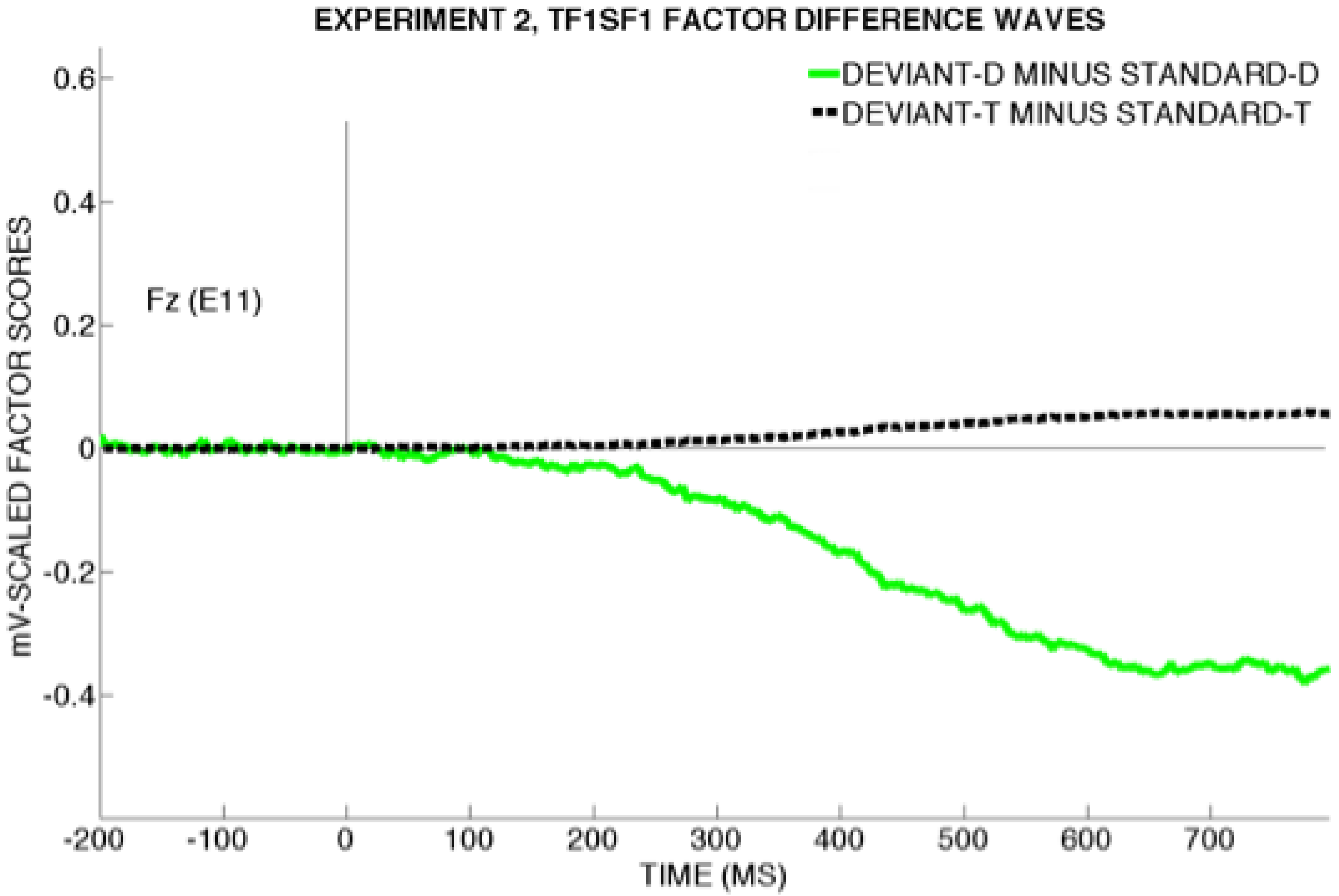

EXPERIMENT 1, MEAN VOLTAGE FOR /t/, TF1SF1-REGION

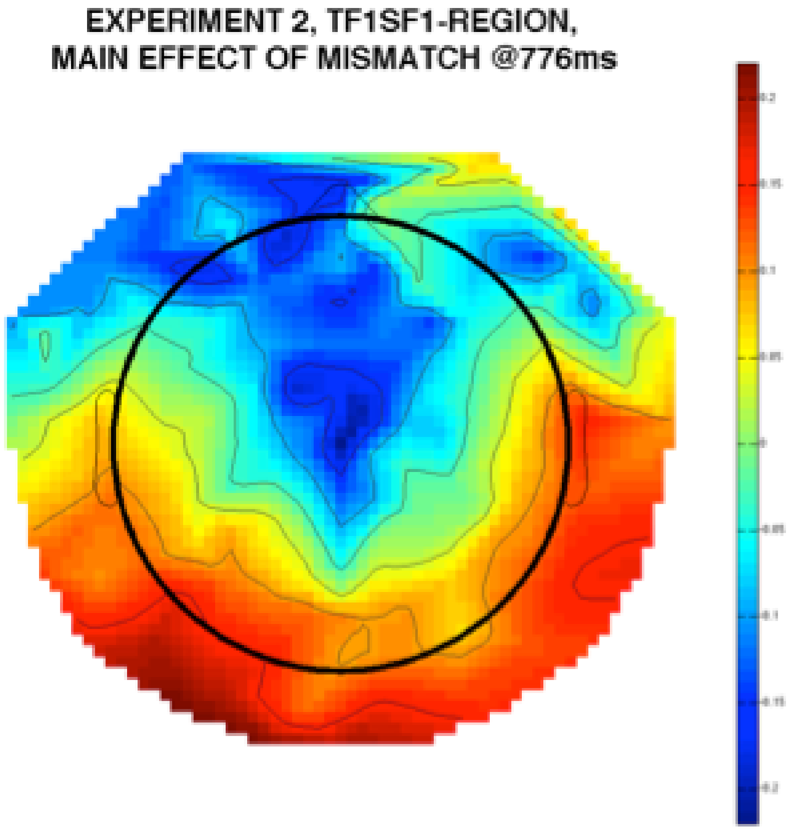

EXPERIMENT 2, TF2SF2 FACTOR DIFFERENCE WAVES

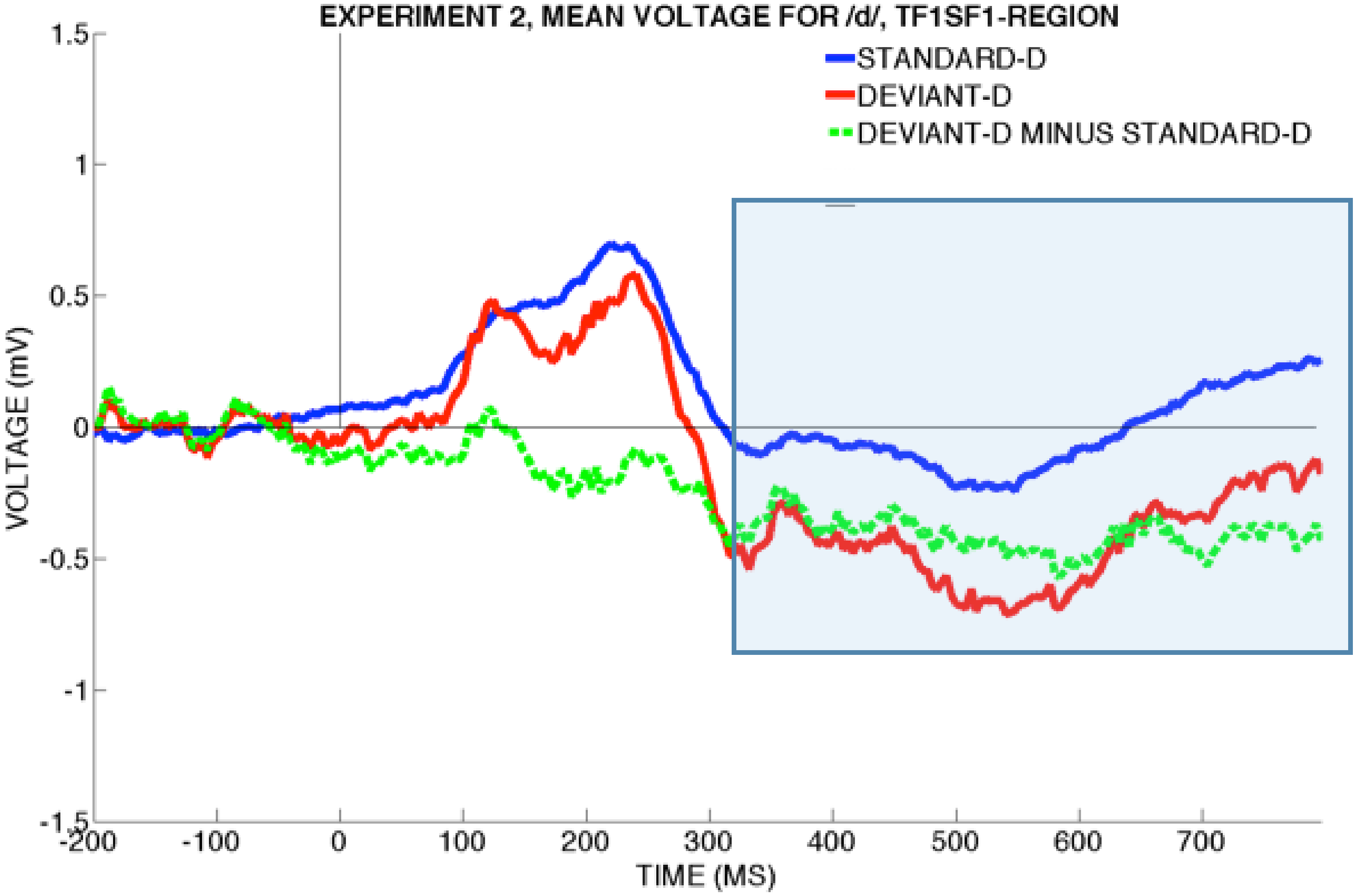

EXPERIMENT 2, TF2SF2 FACTOR SCORES
MISMATCH FOR /d/ @208ms

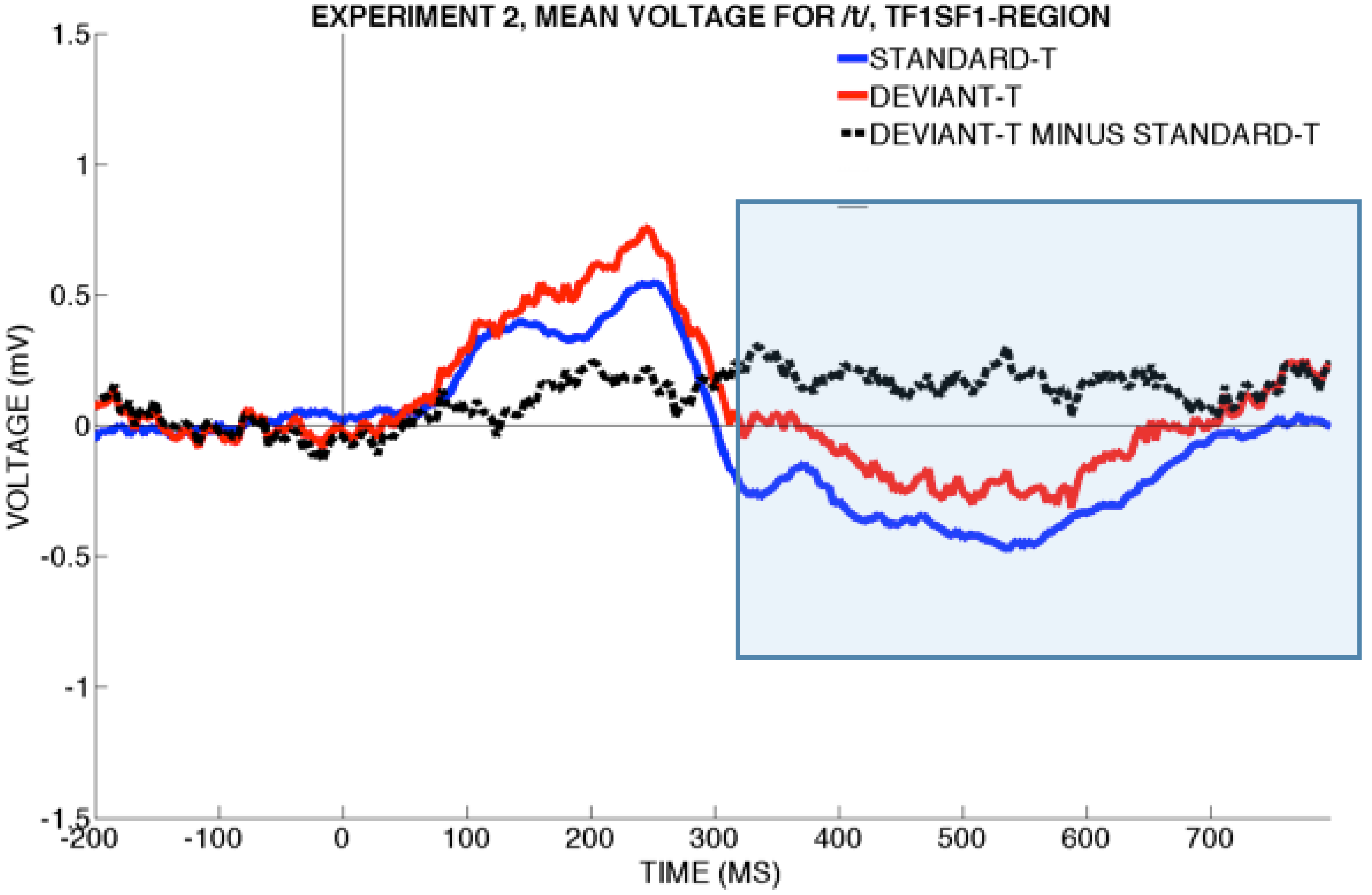

EXPERIMENT 2, MEAN VOLTAGE FOR /d/, TF2SF2-REGION

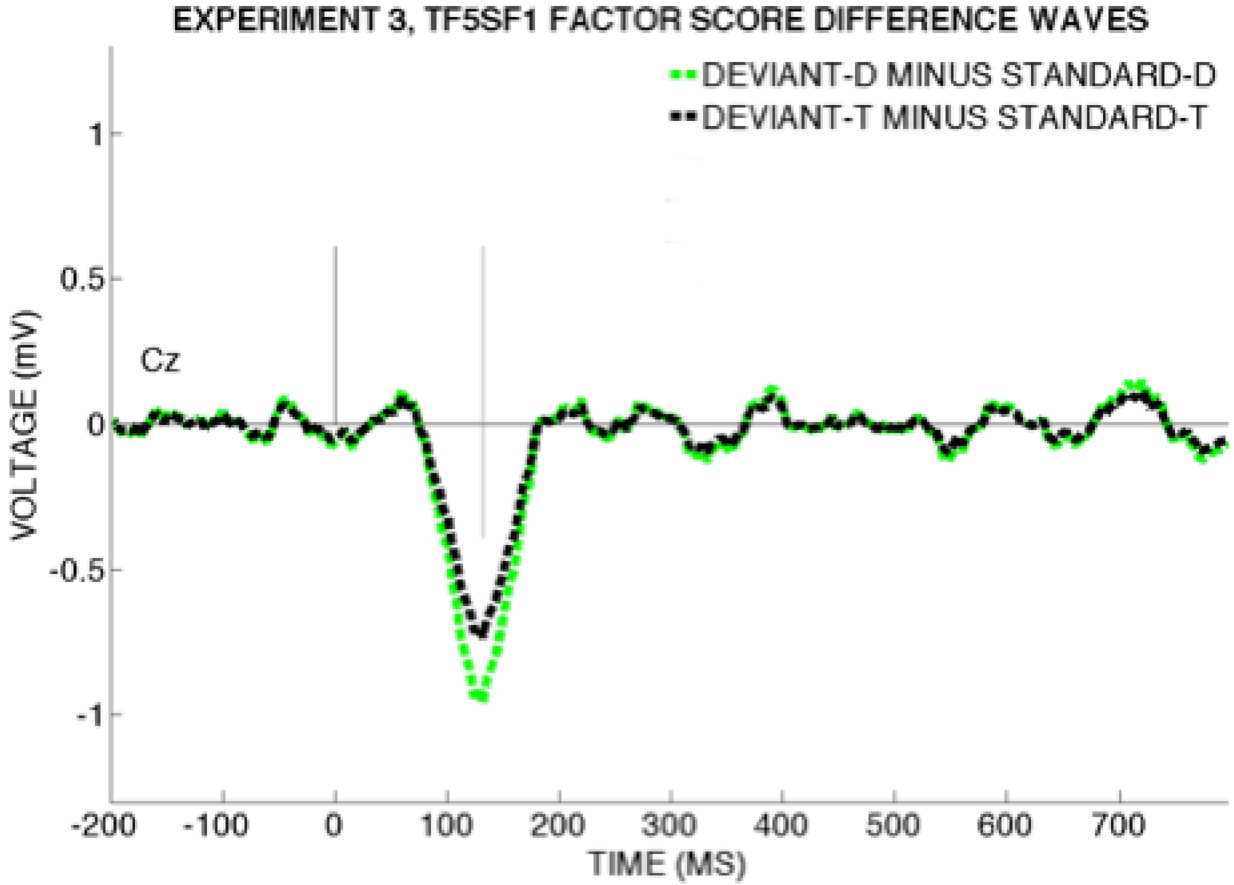

EXPERIMENT 2, MEAN VOLTAGE FOR /t/, TF2SF2-REGION

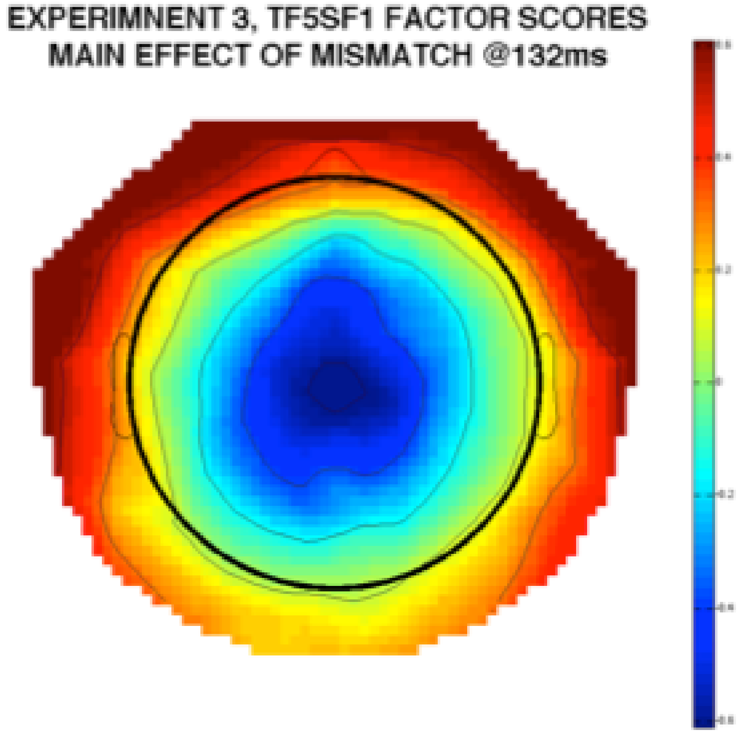

EXPERIMENT 2, TF1SF1 FACTOR DIFFERENCE WAVES

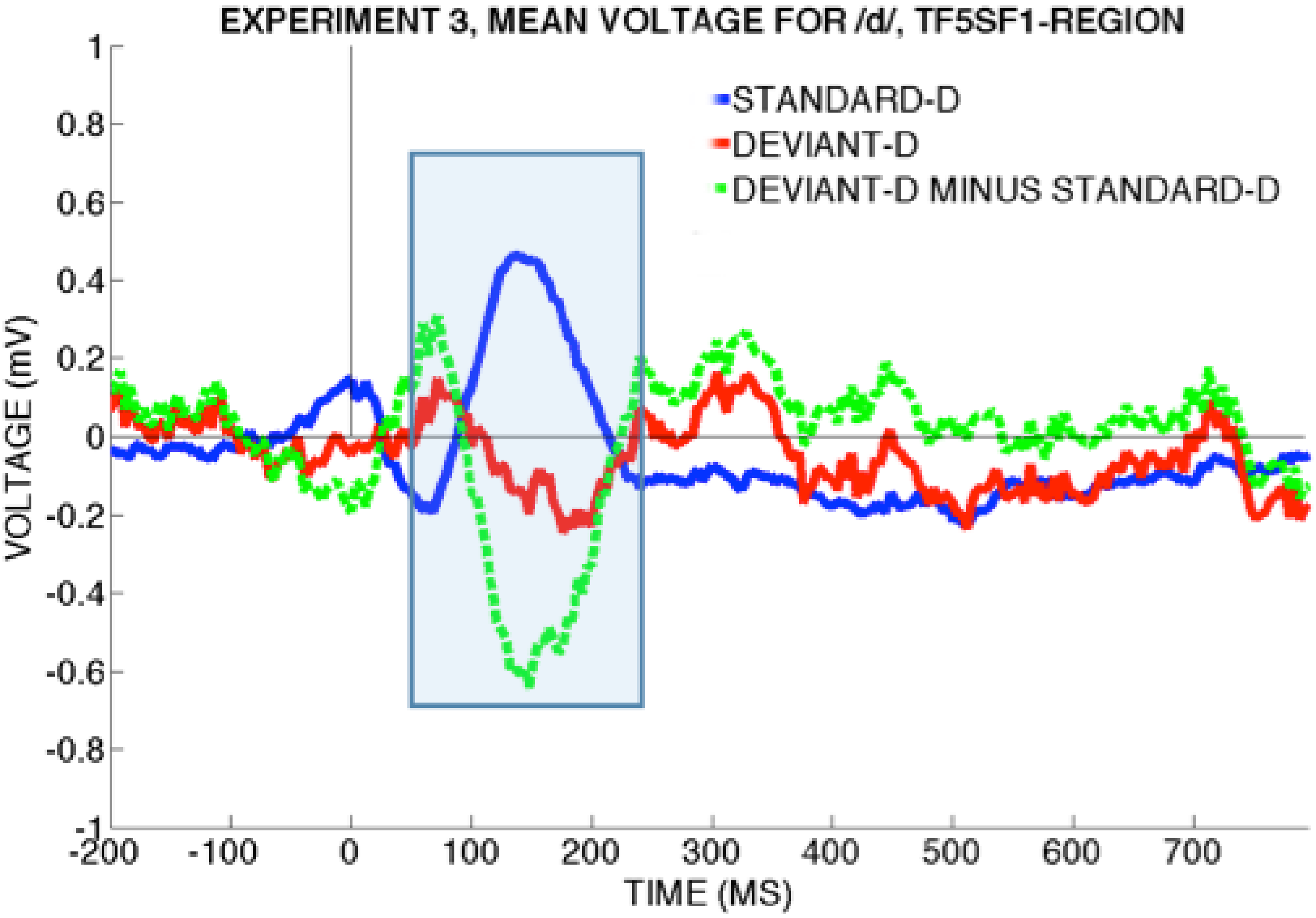

EXPERIMENT 2, TF1SF1-REGION,
MAIN EFFECT OF MISMATCH @776ms

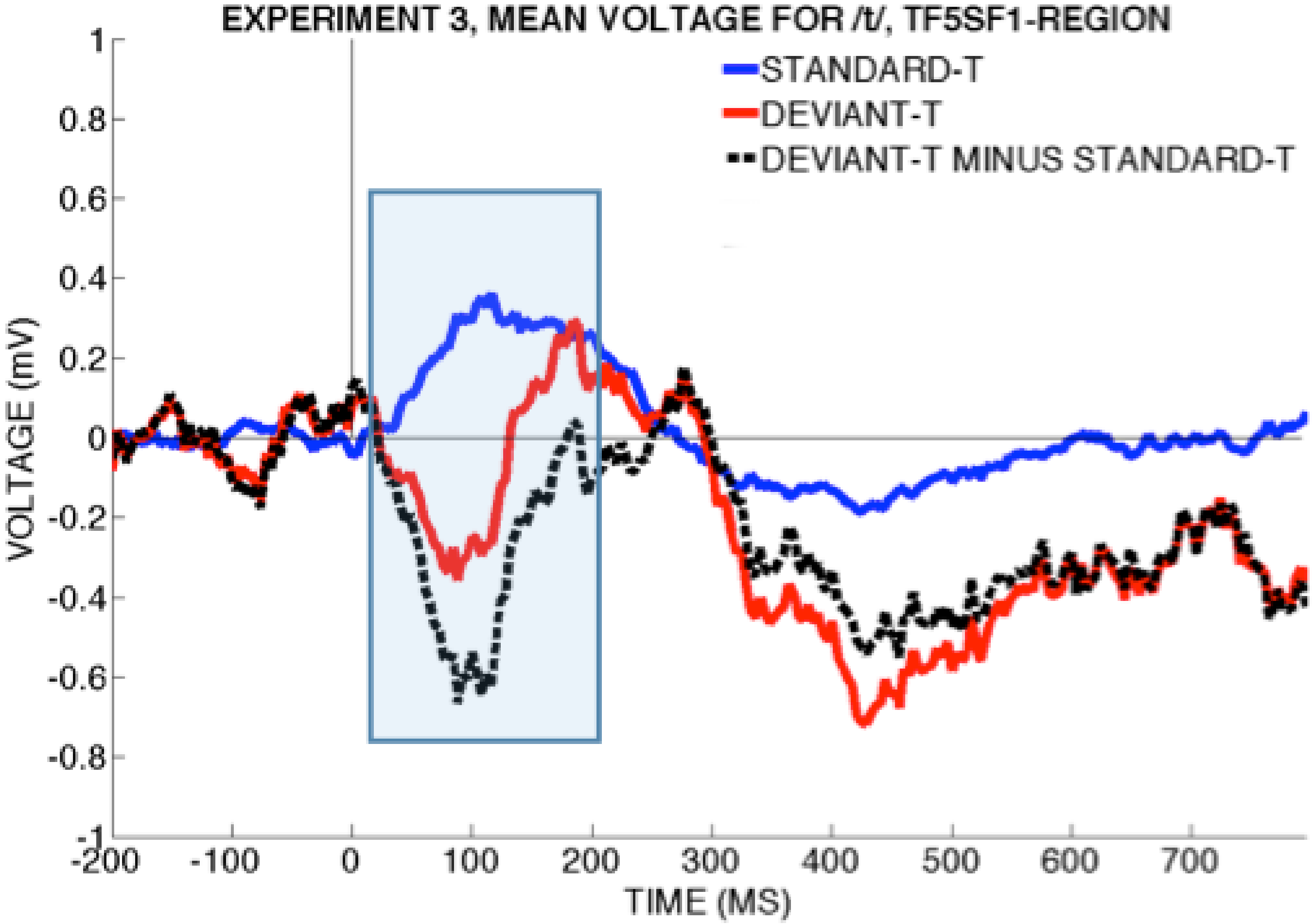

EXPERIMENT 2, MEAN VOLTAGE FOR /d/, TF1SF1-REGION

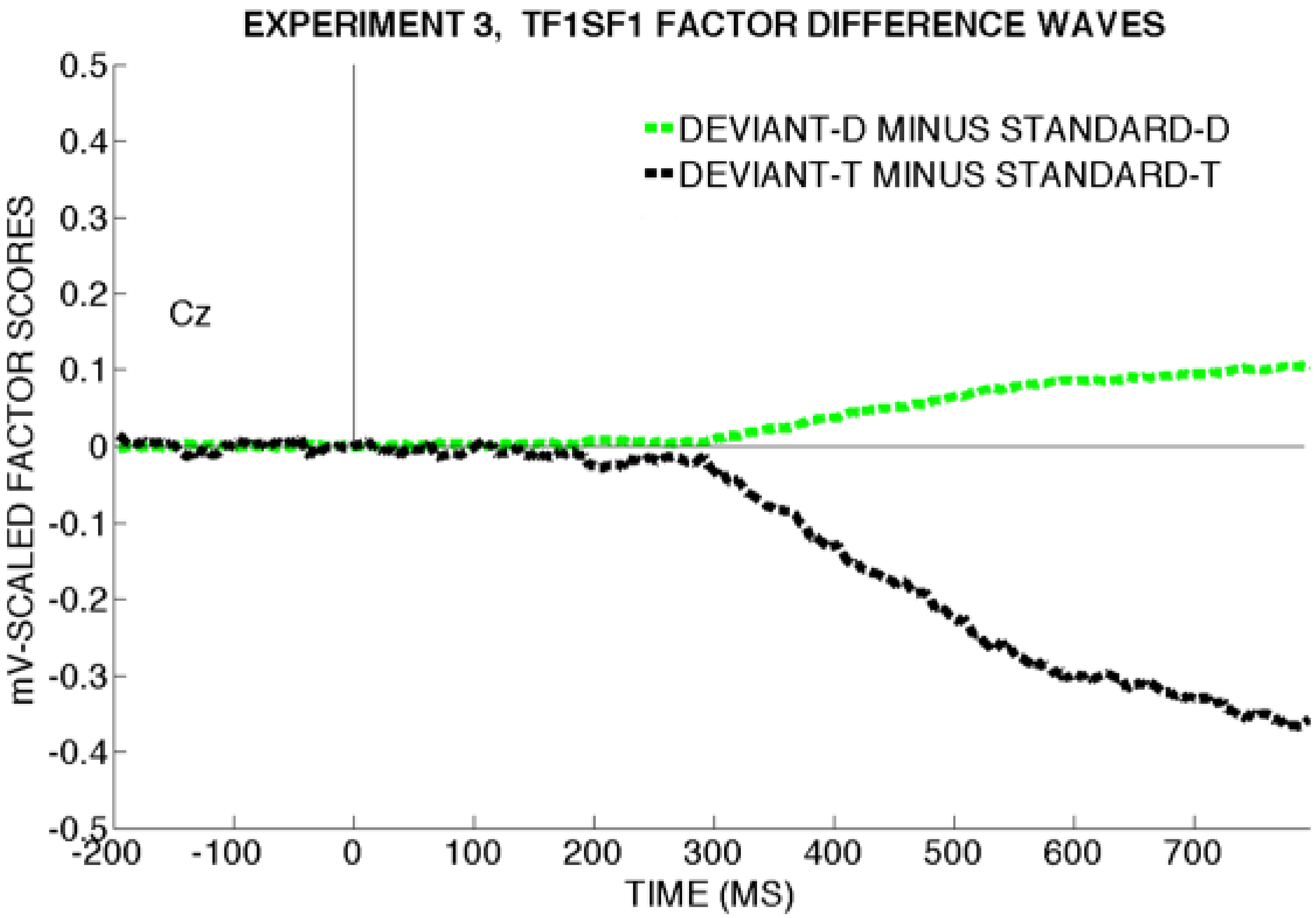

EXPERIMENT 2, MEAN VOLTAGE FOR /t/, TF1SF1-REGION

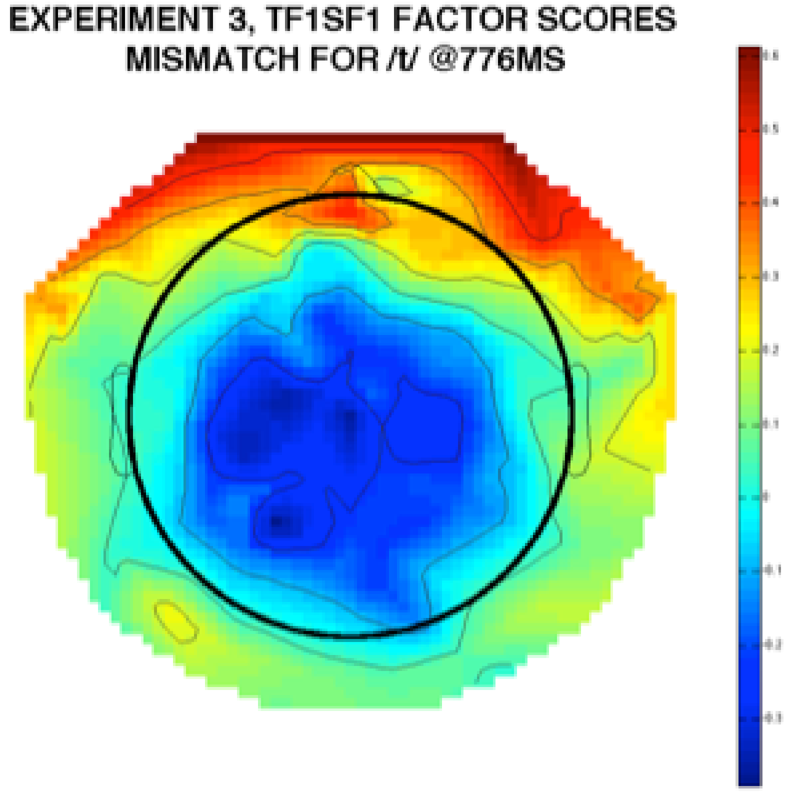

EXPERIMENT 3, TF5SF1 FACTOR SCORE DIFFERENCE WAVES

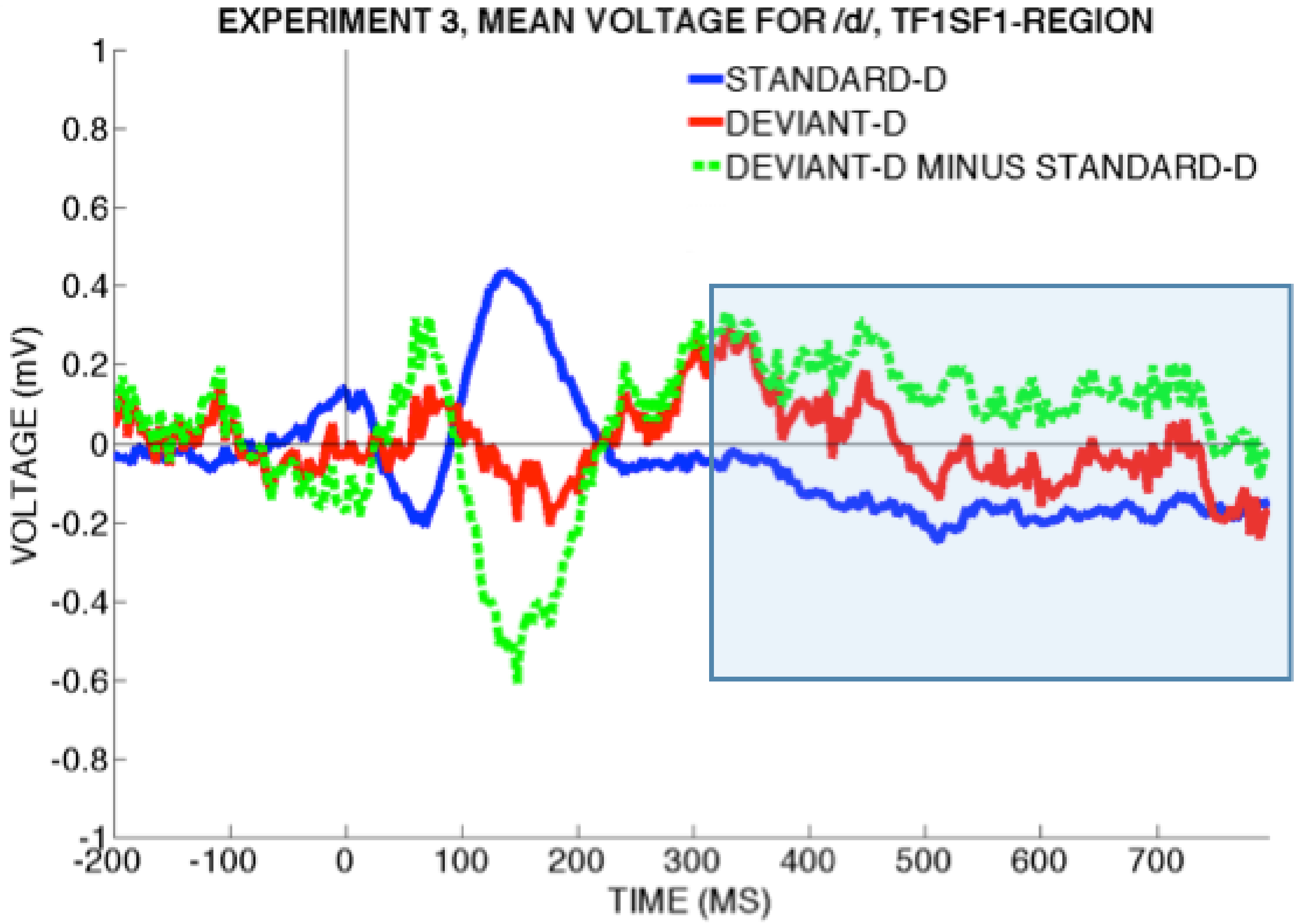

EXPERIMNENT 3, TF5SF1 FACTOR SCORES
MAIN EFFECT OF MISMATCH @132ms

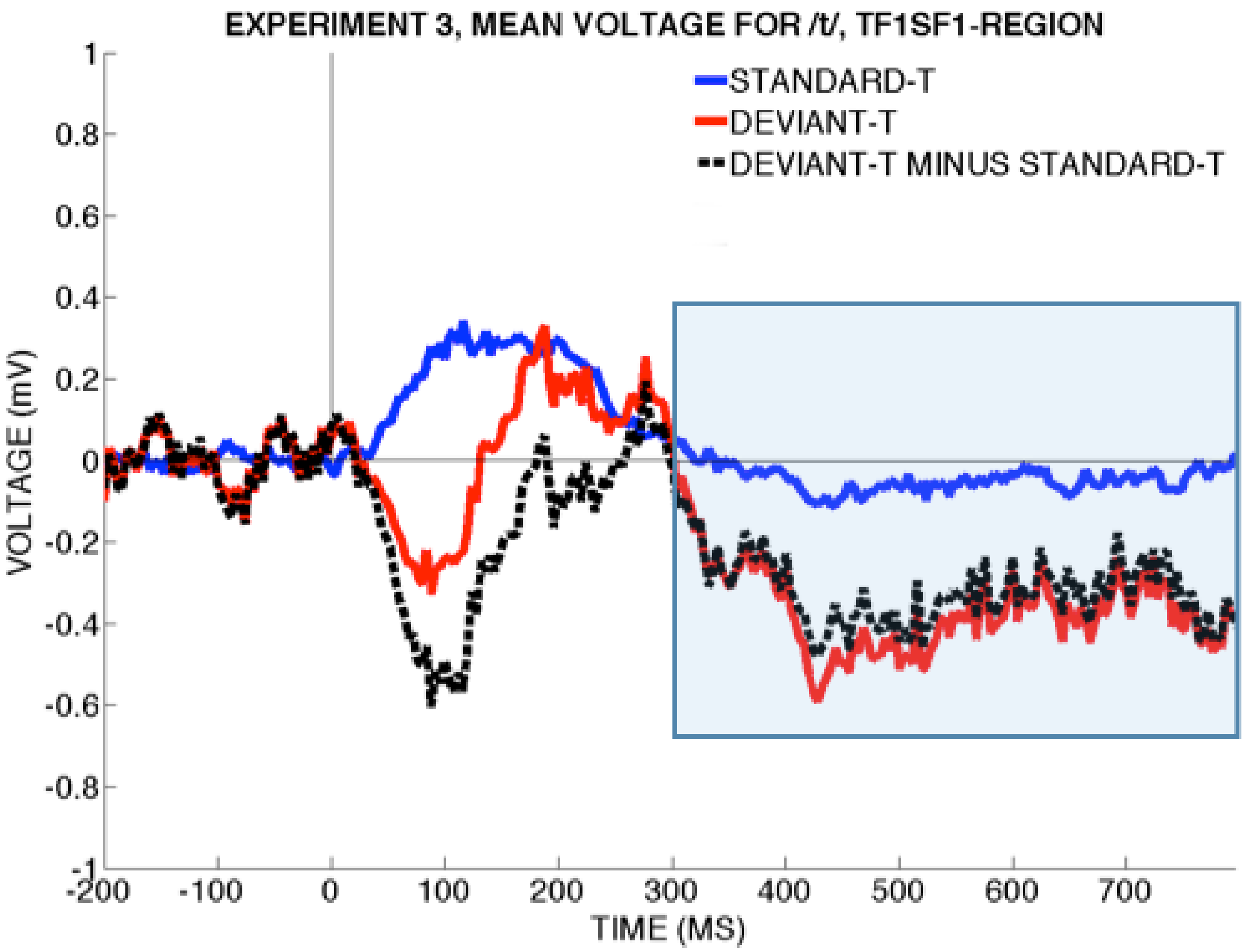

EXPERIMENT 3, MEAN VOLTAGE FOR /d/, TF5SF1-REGION


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

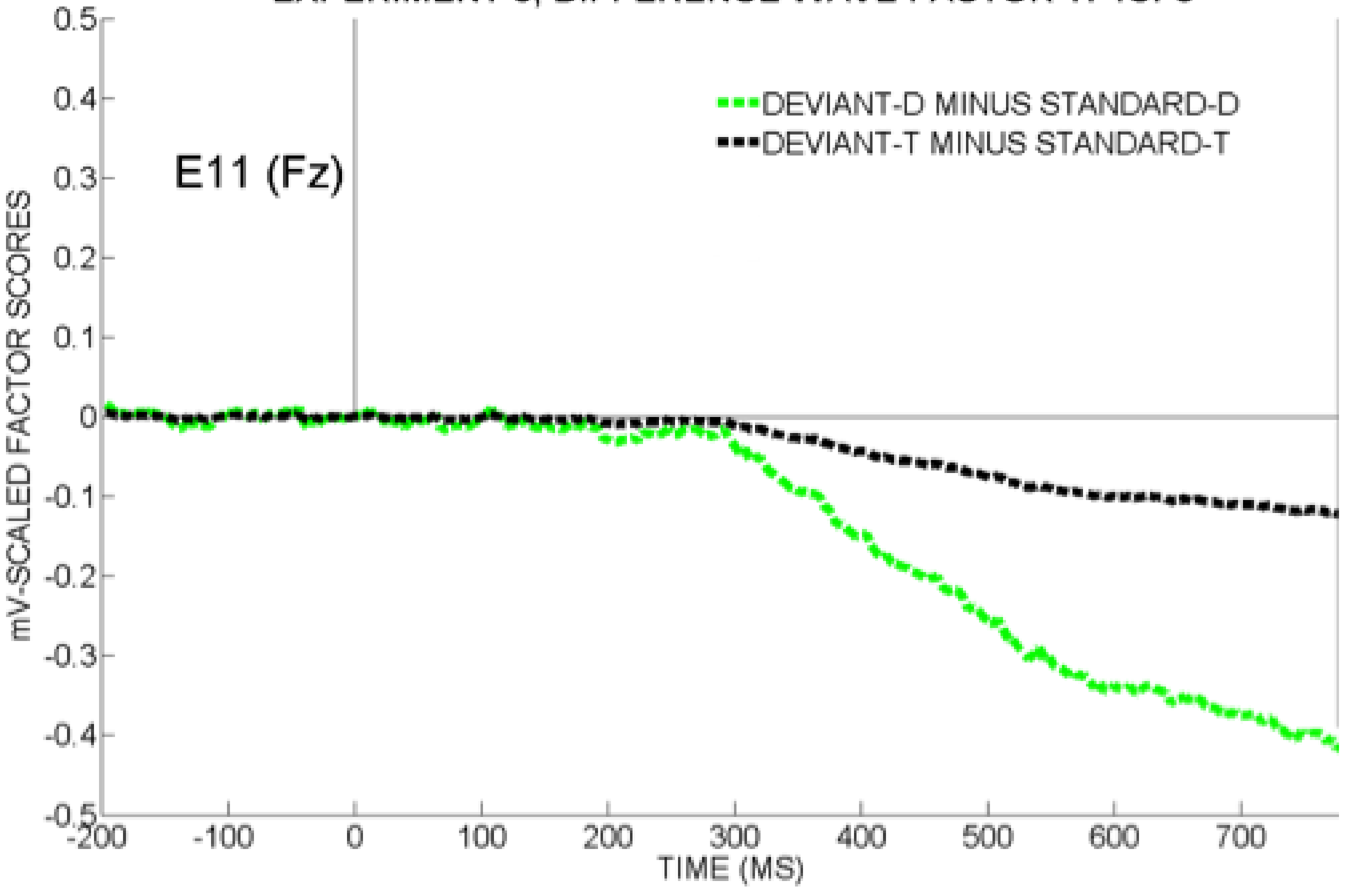

EXPERIMENT 3, TF1SF1 FACTOR DIFFERENCE WAVES
DEVIANT-D MINUS STANDARD-D
DEVIANT-T MINUS STANDARD-T
Cz
mV-SCALED FACTOR SCORES
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
-0.5
-200
-100
0
100
200
300
400
500
600
700
TIME (MS)


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

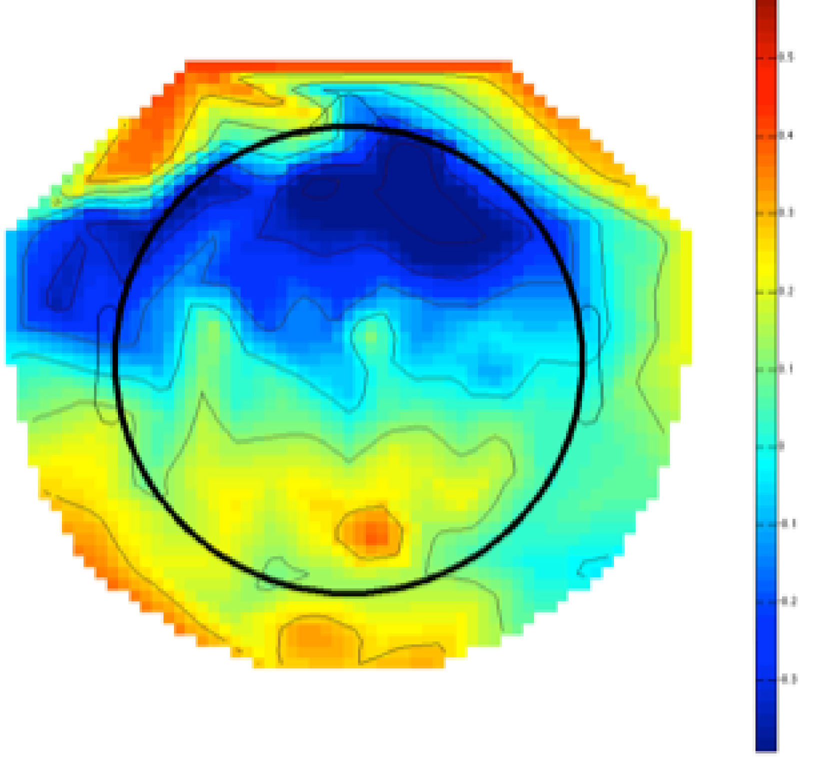

EXPERIMENT 3, MEAN VOLTAGE FOR /d/, TF1SF1-REGION
STANDARD-D
DEVIANT-D
DEVIANT-D MINUS STANDARD-D
VOLTAGE (mV)
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
-200
-100
0
100
200
300
400
500
600
700
TIME (MS)


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)

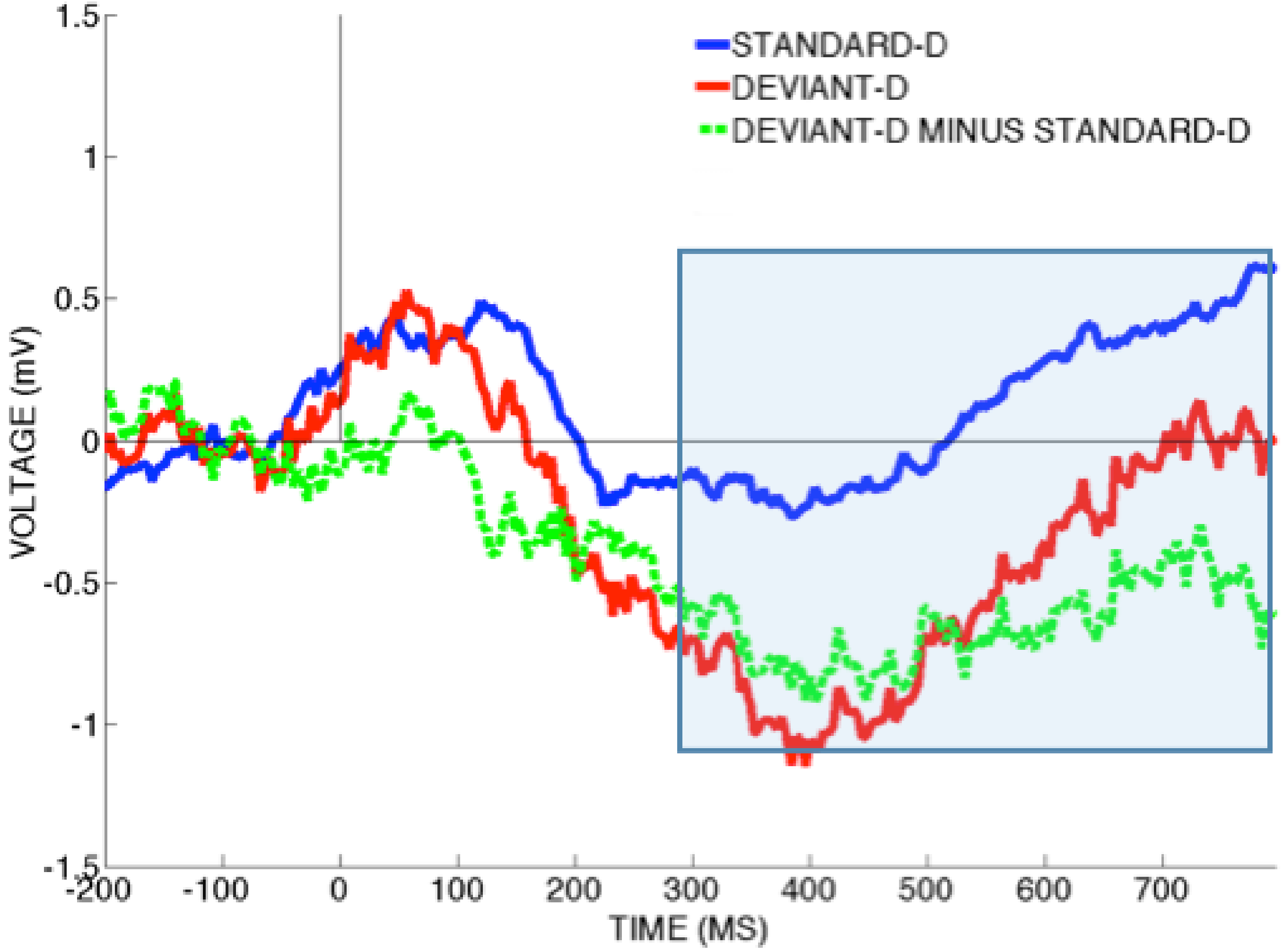
EXPERIMENT 3, DIFFERENCE WAVE FACTOR TF1SF3



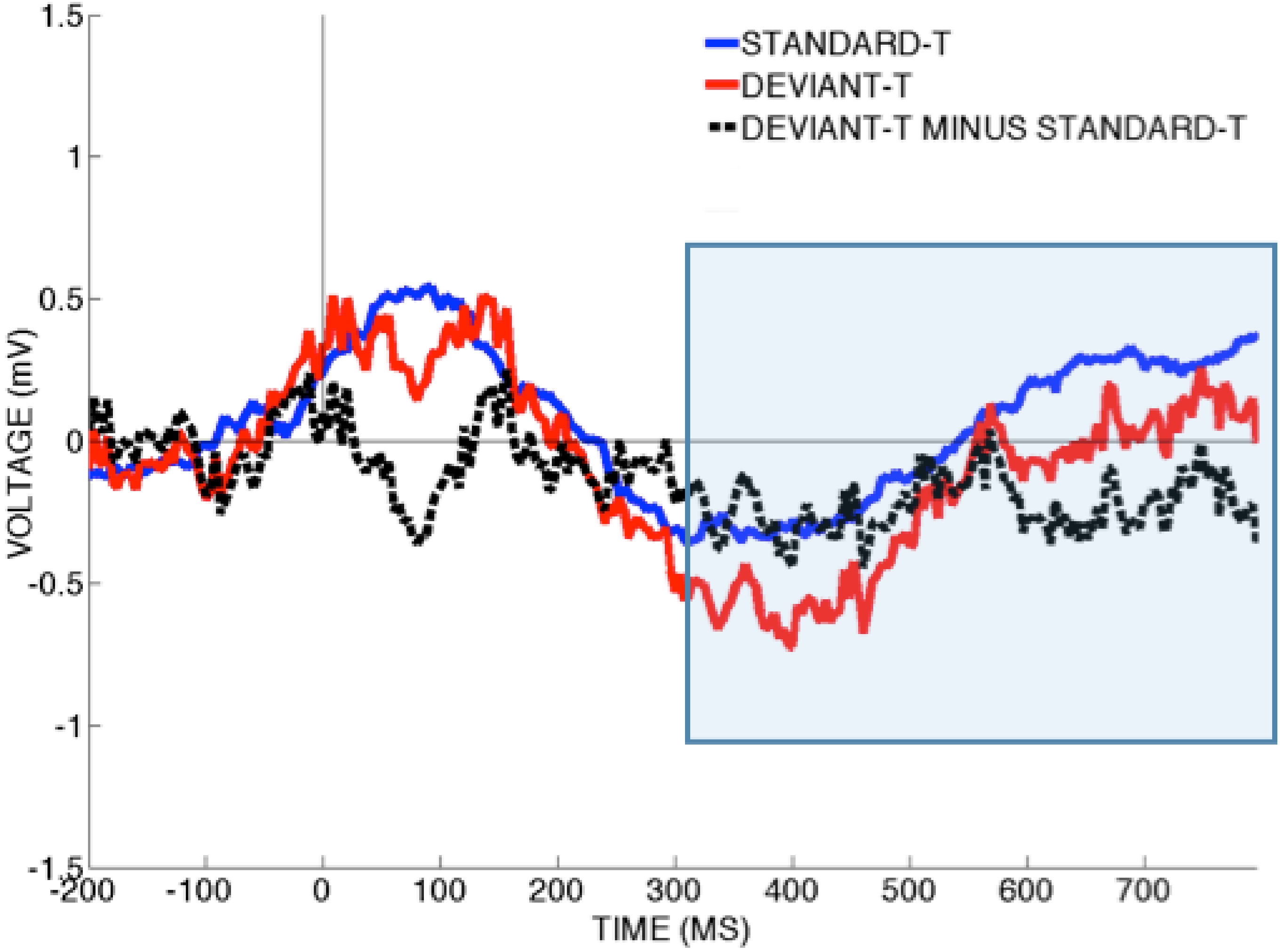
EXPERIMENT 3, TF1SF3 FACTOR SCORES
MISMATCH FOR /d/ @776ms



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



EXPERIMENT 3 MMN FOR /t/ TF1SF3 REGION



		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.