Brain Research 1732 (2020) 146664

Contents lists available at ScienceDirect

Brain Research

journal homepage: www.elsevier.com/locate/brainres

Research report

Abstractness of human speech sound representations

Arild Hestvik[a,*], Yasuaki Shinohara[b], Karthik Durvasula[c], Rinus G. Verdonschot[d], Hiromu Sakai[e]

[a] University of Delaware, Newark, DE, USA
[b] Waseda University, Tokyo, Japan
[c] Michigan State University, East Lansing, MI, USA
[d] Hiroshima University, Japan
[e] Waseda University, Tokyo, Japan

HIGHLIGHTS

- Phonemes are underspecified so that only one member of a minimal pair has a feature value for specific features.
- Japanese and English have the opposite underspecification pattern with respect to the voiced/voiceless stops.
- Previous study found Mismatch Negativity evidence for English /d/ being underspecified in long-term memory.
- The current study finds Mismatch Negativity evidence for the opposite underspecification in Japanese.
- Languages differ in terms of which feature value is underspecified.

ARTICLE INFO

Keywords:
Phonemes
Phonetics
Underspecification
Mismatch negativity
Language-specificity

ABSTRACT

We argue, based on a study of brain responses to speech sound differences in Japanese, that memory encoding of functional speech sounds—phonemes—are highly abstract. As an example, we provide evidence for a theory where the consonants/p t k b d g/ are not only made up of symbolic features but are *underspecified* with respect to voicing or laryngeal features, and that languages differ with respect to which feature value is underspecified. In a previous study we showed that voiced stops are underspecified in English [Hestvik, A., & Durvasula, K. (2016). Neurobiological evidence for voicing underspecification in English. *Brain and Language*], as shown by asymmetries in Mismatch Negativity responses to /t/ and /d/. In the current study, we test the prediction that the *opposite* asymmetry should be observed in Japanese, if voiceless stops are underspecified in that language. Our results confirm this prediction. This matches a linguistic architecture where phonemes are highly abstract and do not encode actual physical characteristics of the corresponding speech sounds, but rather different subsets of abstract distinctive features.

1. Introduction

1.1. Phonemes, features and underspecification

A fundamental question in the scientific study of language is how speech sounds are represented in memory. The answer provided by generative grammar (Chomsky and Halle, 1968; Trubetskoy, 1969) is that single speech sounds are not represented in long-term memory as holistic units, but rather as bundles of abstract distinctive features, which provide the minimal set of binary oppositions needed for lexical differentiation. For example, in classical phonology /d/ is represented as [+voice, +obstruent, +coronal], and /t/ is represented as [-voice, +obstruent, +coronal]; the two units are only differentiated by the distinctive feature 'voice'. Underspecification Theory (Archangeli, 2008; Lahiri and Reetz, 2002; Lahiri and Reetz, 2010) increases abstractness by proposing that a phoneme may have a "default" value for a given feature, which therefore need not be specified in the memory representation. To use voicing as an example, only one phoneme in a minimal pair will have a specific value for that feature. The underspecified member of a pair then obtains its value in the mapping from lexical representations to phonetic representations by a general redundancy rule, as required by the principle of Full Interpretation (Chomsky, 1986, 1995). I.e., in order to *pronounce* a speech sound, all articulatory features must be specified.

Underspecification was brought into the realm of cognitive neuroscience by Eulitz and Lahiri (2004) and Lahiri and Reetz (2002). They

* Corresponding author.
E-mail address: hestvik@udel.edu (A. Hestvik).

https://doi.org/10.1016/j.brainres.2020.146664
Received 5 July 2019; Received in revised form 2 January 2020; Accepted 9 January 2020
Available online 10 January 2020

proposed a linking theory that relates memory representations of phonemes to amplitude modulation of the Mismatch Negativity (MMN) brain response (Garrido et al., 2009; Näätänen and Alho, 1997; Näätänen et al., 2005, 2007). The MMN is typically elicited by an "oddball" or deviant stimulus in a sequence of other stimuli that represents a generalization, pattern or rule. It thus serves as a neurophysiological measure of change detection. For example, if a sequence of 400 Hz pure tones is presented, the auditory system will develop a prediction that the next stimulus will also be a 400 Hz pure tone. If instead a 200 Hz tone is presented, this prediction is violated, and the prediction error is observed as a modulation of voltage in a time window overlapping with the Auditory Evoked Potential (AEP) (Steinschneider et al., 2011). The difference between the average voltage of the standards and that of the deviants is the MMN. The MMN is typically observed in the 150–300 ms time window (depending on stimulus and experimental paradigm parameters). In addition, the mismatch negativity effect may also be observed during the earlier time window of the exogeneous or "obligatory" response to sound onset, the N1 wave. When observed during the N1 time window, the MMN overlaps with and is independent of the N1 response to physical stimulus properties (Näätänen and Picton, 1987; Näätänen et al., 1988).

A critical part of Lahiri and Eulitz' linking theory rests on using a "varying standards" MMN paradigm: As demonstrated by Phillips et al. (Phillips et al., 2000), if a series of allophones of a phoneme is presented as the standards, the auditory cortex will recruit a memory representation of the phoneme to serve as the "unifying" memory representation of those standards. The memory representation of the standards is therefore the set of distinctive features defining the phoneme, which is a sparser or more general representation than a phonetic representation of the individual standard tokens. Suppose feature 1 ("F1") and feature 2 ("F2") are distinctive features, and feature 3 ("F3") is a non-distinctive phonetic feature. A change in feature F3 will then not cause an MMN, but a change in feature F1 or feature F2 will. Depending on which value of feature F2 is underspecified, an MMN will or will not be predicted. This is illustrated in Table 1 below, where the process can be seen as going from left to right in time. The first three rows illustrate the case of varying standards, but where only [+F1, −F2] are distinctive features, hence only these two features remain in the memory trace. This generates a prediction of a new sound with at least these two features. In row 1, the new deviant token matches in the two distinctive features but may differ in the third, redundant phonetic feature. The new deviant token matches the prediction so no MMN is generated. The second row is the same case, except here the phonetic deviant token has the opposite value for the redundant phonetic feature; again, no MMN is predicted. The third row illustrates the typical change detection case: Now, the deviant has a change with respect to one of the distinctive features in the memory trace, hence an MMN is predicted to be generated. The last and fourth row illustrates the Eulitz and Lahiri (2004) underspecification case: Here, −F2 is underspecified in the language; hence the memory trace has no specification for F2. Even if the deviant token has an opposite value (+F2) compared to the phonetic tokens (−F2) used to generate the memory trace, there is *no direct feature conflict* between memory trace information ∅ and +F2 in the deviant.

Critically, the effect of varying the standards within a phonemic category has the effect of setting up a comparison between a fully specified phonetic representation of a token and a sparser phonemic representation of the memory trace. Eulitz and Lahiri (2004), using the identity MMN paradigm, used this to compare each vowel to the other in a bidirectional way and thus tested predictions about underspecification-driven mismatch asymmetries. Specifically, if F2 above was underspecified for the phoneme used as the standard, it would not be represented in the memory trace. If a deviant was presented with the opposite feature value, and is compared to the memory trace, the comparison is of [+F1, ∅] to [+F1, +F2], and since ∅ and +F2 are not contradictory, no MMN is predicted. Eulitz and Lahiri used this technique to demonstrate that German coronal vowels are underspecified compared to labial and dorsal vowels, and their work opened up a new research area combining cognitive neuroscience with underspecification theory, and led to a series of studies probing coronal underspecification (Cornell et al., 2011, 2013; Scharinger et al., 2012) (but see Scharinger et al., 2011); tongue-height specification of English vowels (Scharinger et al., 2016), and voicing in stops (Hestvik and Durvasula, 2016). However, some MMN studies have failed to observe or replicate underspecification-predicted asymmetric MMNs related to the coronal place feature (Bonte et al., 2005; Tavabi et al., 2009) , and some studies have failed to find priming related predictions of underspecification (Gow, 2001), or asymmetries of lexical activation evidence for underspecification using eye-tracking (Mitterer, 2011). Thus, the question of whether there is sufficient psychological evidence for underspecification theory remains in question.

1.2. Previous study investigating the /d/-/t/ contrast in English

In Hestvik and Durvasula (2016), we applied Eulitz and Lahiri's experimental logic and linking theory for the English /d/-/t/ contrast, and observed MMN amplitude modulations that provided evidence for the claim that /d/ is underspecified and /t/ is specified in English. For the feature that distinguishes /d/ from /t/ in English, we followed the phonological theory of laryngeal realism and assumed the feature to be [spread glottis] (Avery and Idsardi, 2001; Beckman et al., 2018; Beckman et al., 2013; Harris, 1994; Honeybone et al., 2005; Iverson and Salmons, 1995; Jessen and Ringen, 2003; Kager et al., 2007). Laryngeal realism states that there are two different articulatory mechanisms that can differentiate /p t k/ from /b d g/: either vocal cord vibration (voice) or glottal width. In the latter case, the feature value [spread glottis] characterizes /p t k/ and [∅] characterizes /b d g/. The reason is that syllable-initial English /p t k/ are consistently aspirated, whereas word-initial voiced stops /b d g/ are not consistently voiced (Davidson, 2016; Docherty, 1992), hence /p t k/ must have the specified feature. Furthermore, the amount of aspiration, and thereby the duration of Voice Onset Time (VOT) of the voiceless stops series increases at slower speech rates (Kessinger and Blumstein, 1997), suggesting that the phonemes are represented with an "intended" [spread glottis] gesture for aspiration. In contrast, there is no modulation of VOT due to the speech rate for the voiced stops, suggesting that there is

Table 1
The four logically possible scenarios for feature conflict in a varying standards paradigm with underspecification. Only one case leads to an MNN.

varying standards in category *A*	memory trace	prediction	new token	predicted response:
token A_1 = [+F1, −F2, +F3] token A_N = [+F1, −F2, −F3]	[+F1, −F2]	token A_{N+1} = [+F1, −F2, αF3]	token A_{N+1} = [+F1, −F2, +F3]	No feature conflict between memory trace and new token: **No MMN**
token A_1 = [+F1, −F2, +F3] token A_N = [+F1, −F2, −F3]	[+F1, −F2]	token A_{N+1} = [+F1, −F2, αF3]	token A_{N+1} = [+F1, −F2, −F3]	New token is phonetic variant, no feature conflict: **No MMN**
token A_1 = [+F1, −F2, +F3] token A_N = [+F1, −F2, −F3]	[+F1, −F2]	token A_{N+1} = [+F1, −F2, αF3]	token A_{N+1} = [+F1, +F2, +F3]	New token has direct feature conflict with memory trace: **MMN is elicited**
token A_1 = [+F1, −F2, +F3] token A_N = [+F1, +F2, −F3]	[+F1, ∅]	token A_{N+1} = [+F1, ∅, αF3]	token A_{N+1} = [+F1, +F2, +F3]	F2 is underspecified, no feature conflict: **No MNN**

no intended gesture specified for /b d g/. Further evidence for /d/ being underspecified and /t/ being specified comes from the English phonological assimilation rule observed for plurality. Observationally, the plural morpheme is realized as [z] after sonorants and voiced obstruents (e.g. dog[z], bee[z], ban[z]), but as [s] after voiceless obstruents (e.g. cat[s], laugh[s], back[s]), suggesting assimilation with the stem final phoneme. Is /s/ or /z/ the underlying form? The answer comes from environments where there is no assimilation, namely after sibilant-final morphemes, when epenthesis breaks up the sequence. Here, the suffix surfaces as [z]: bush[ɪz], bus[ɪz], batch[ɪz]. Due to the inserted vowel, there is no feature spreading from the stem final consonant, hence the feature value is filled in with the default (underspecified) value [+voice], and surfaces as [z] (Avery and Idsardi, 2001; Beckman et al., 2013; Iverson and Salmons, 1995).

Hestvik and Durvasula (2016) observed that when /d/ and /t/ are contrasted in a varying standards MMN experiment, the MMN to /d/ as deviant is significantly greater than the MMN to /t/ as deviant. This follows from underspecification of /d/. If the phonetic [d] (non-spread glottis sound) is compared to the phoneme memory trace of /t/ which is specified for spread glottis ("voiceless"), there is a direct feature conflict and hence an MMN is elicited. On the other hand, when [t] is the deviant, it is compared to a phoneme memory trace of varying versions of [d]. This phoneme representation of /d/ is underspecified, without specification for laryngeal features; hence there is no feature conflict and no MMN is generated. This observation leads to the prediction that if another language were to have the opposite underspecification pattern for the same speech sound contrast, then the opposite MMN asymmetry should be observed. Japanese is such a language, as we discuss next.

1.3. Japanese underspecification and the /d/-/t/ contrast

According to Avery and Idsardi (2001), Japanese differs typologically from English in that it uses the *voicing* dimension for phonological contrast between /p t k/ and /b d g/. Classical phonological arguments based on the interaction between intervocalic voicing ("*rendaku*") and a phonotactic pattern labelled *Lyman's Law* (Ito and Mester, 1986) furthermore suggests that the voiceless series is underspecified, because only voiceless stops are subject to voicing assimilation, whereas voiced stops resist devoicing phonologically. The *Rendaku* rule only applies to the native, non-Sino-Japanese lexical items, and changes a voiceless morpheme-initial obstruent into its voiced version when compounding makes it intervocalic. This is illustrated in (a-b) below; and (c) shows that the rule is blocked if there is another voiced obstruent in the morpheme (cf. 1c), because voicing it would result in a violation of Lyman's Law:

/ori/ + /kami/ → [ori**g**ami] 'origami'
/jama/ + /koja/ → [jama**g**oya] 'mountain shack'
/jama/ + /kaẓi/ → [jama**k**aẓi] 'mountain fire'

Mester and Ito (1989) argued that this pattern can only be explained if the voiceless stops are underspecified, because that allows a rule to change an underspecified [voice] feature to [+voice]. There is no similar rule that turns voiced stops to voiceless, hence /b d g/ must have the specification [+voice] in the lexicon.[1]

The Japanese writing system also appears to reflect this phonological underspecification. In the *hiragana* and *katakana* syllable or mora orthographical symbols, the voiced stop series is derived from the voiceless symbols by adding a diacritic mark (*dakuten*) to the voiceless series, cf. こ [ko] vs. ご [go]. If we assume that phonological distinctions make an imprint on the writing system, then this is consistent with voiceless stops being the unmarked form and underspecified. Indeed, a recent analysis of *rendaku* (Kawahara, 2018) argues that it operates on orthographic representations. Kawahara's analysis agrees with the spirit of the idea that *rendaku* is a process that adds "voicing" to unvoiced elements. (See Kuroda (2002) for an opposing view.)

In sum, we assume a phonological theory where English and Japanese have diametrically opposite patterns of underspecification for obstruents (even though the specific laryngeal features differ). This contrast is illustrated in Fig. 1.

1.4. The current study

Given the phonological typology illustrated in Fig. 1, we can now make the following prediction: If Japanese listeners participated in the same experiments reported for English speaking participants in Hestvik and Durvasula (2016), then the observed MMN asymmetry should go in the opposite direction; cf. Fig. 2.

Specifically, if a sequence of varying standards [d] is presented (by varying VOT within category), the memory trace will contain the distinctive feature value [+voice]. A deviant stimulus [t] with the feature [-voice] will then generate an MMN. On the other hand, if a sequence of varying standards [t] is presented (by similarly varying VOT within category), then the memory trace will not contain a feature specification for voicing. A deviant [d] will therefore not generate an MMN response.

To test these predictions, we conducted a replication of the varying standards Experiment 1 in Hestvik and Durvasula (2016), using the exact same stimuli, thus only changing the language of the participants (see Section 5 for the exact details). Since these stimuli were not natural sounding Japanese syllables, the ERP recording was preceded by a behavioral identification task to determine that Japanese participants perceived the stimuli as falling into two categories; and to determine their categorical perception threshold in VOT in order to tailor the stimuli to that threshold.

2. Results

2.1. Pre-test behavioral results

The mean VOT threshold for the stimulus sequence used in Hestvik and Durvasula (2016) was 33 ms (SD 7 ms). Fig. 3 displays the mean identification function curve and the phoneme boundary on the VOT continuum for the 62 Japanese subjects who took part in the ERP experiment.

In our previous study of English-speaking participants, we "customized" the stimulus sequence to the measured individual thresholds of the participants, following Phillips et al., 2000). The mean VOT threshold was 40 ms, with a 3.6 ms standard deviation. Given the narrow variance, we decided in Hestvik and Durvasula (2016) to exclude 8 outliers and combine the data based on VOT thresholds of 35, 40 and 45 ms. Therefore, given the limited benefits of the customization approach, we abandoned it in the current study, and instead used a fixed stimulus set based on the mean observed VOT threshold. This will also have the benefit of reducing VOT-induced latency jitter in the data.

The mean VOT value was set at 35 ms based on the phoneme boundary estimate from the first 40 participants' identification results

[1] There is a process of vowel devoicing in Japanese adjacent to a voiceless consonant that might be used to argue that voiceless consonants are also specified for laryngeal features as [+spread glottis](Tsuchida, 1997; Tsuchida, 2001). However, as pointed out by Avery and Idsardi (Avery and Idsardi, 2001), this specific pattern is better analyzed as phonetic overlap, because the vowel devoicing is highly variable. This is characteristic of a phonetic process, rather than a strictly phonological process. In contrast, the *rendaku* facts are indicative of phonological specification of the relevant consonants because *rendaku* is restricted to a particular type of compound, and therefore cannot be a purely

(*footnote continued*)
phonetic process.

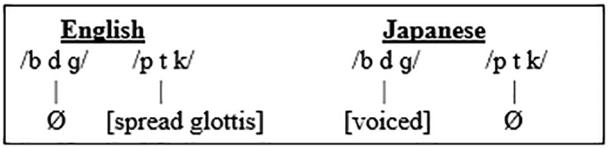

Fig. 1. The phonological contrast between English and Japanese with respect to underspecification and phonological coding of the /d/-/t/ contrast.

difference between the longest VOT of /d/ and the shortest VOT of /t/ was 20 ms, as was the case in Hestvik and Durvasula (2016). After starting the EEG recording session with the /d/ and /t/ stimuli which were selected based on the first 40 subjects' results, the behavioral identification experiment was continued for the remaining 22 subjects.

Note that this threshold was valid for the synthetic stimulus sequence we employed, and does not necessarily reflect exactly the VOT thresholds in Japanese dialects. Shimizu (1977) measured the threshold

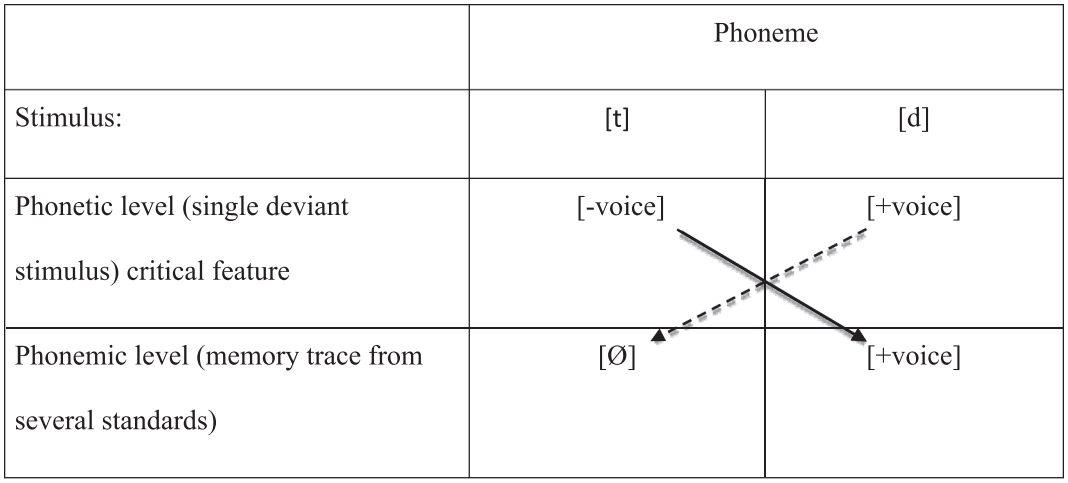

Fig. 2. Predictions for current experiment. Deviant [d] has no feature conflict with a phonemic memory trace for /t/, as indicated by the dotted line. Deviant [t] has a direct feature conflict with a /d/ phonemic memory trace, as indicated by the solid line.

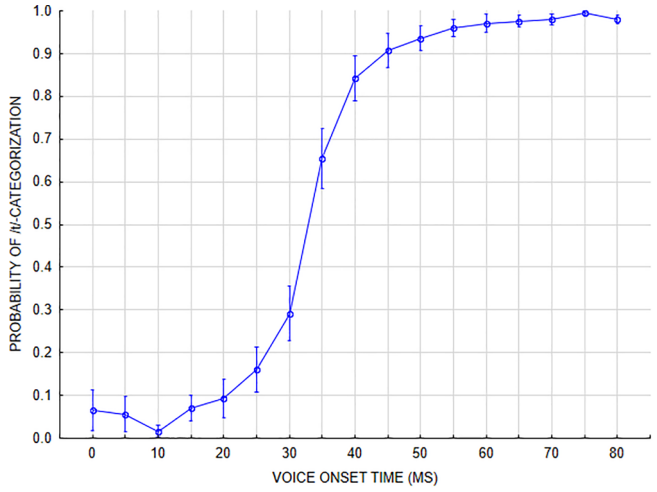

Fig. 3. Relationship between voice onset time and category identification by the Japanese participants.

(i.e. the stimulus with a VOT closest to the mean threshold of 33 ms).[2] We then selected four stimuli on each side of the VOT boundary to be used as representatives of /d/ and /t/ in the ERP experiment. For /d/, we used stimuli with VOTs of 10, 15, 20, and 25 ms, and for /t/, we used stimuli with VOTs of 45, 50, 55, and 60 ms.[3] Consequently, the for a synthetic VOT continuum for /d/-/t/ with 96 stimuli with VOT values from −40 ms to +80 ms with 12 participants. The mean reported threshold value was 26 ms (SD = 7, range = 12–35). The 95% confidence interval of this mean is 21–30 ms; our mean lies just outside this CI. One possible explanation for this difference is that the low intensity of the burst in our stimuli may have caused the subjects to require a longer VOT for the voiceless stop /t/ identification (Chao and Chen, 2008; Klatt, 1975). Another possible explanation for the difference is that Shimizu (1977) study was conducted 30 years before our study, so could be caused by generational differences in dialects. Takada (2011, 2012), using spontaneous speech corpora, observed significant variability among different age groups as well as dialect groups. A more recent laboratory elicitation study on Japanese VOT Riney et al. (2007) reported the mean VOT value of Japanese /t/ to be 28.5 ms. Since Riney et al. focused on voiceless stops, it is not clear what VOT threshold is expected from this result. Note also that their participants include both speakers from Tokyo dialect and Fukuoka dialect. Given that the dialectal variation is significant, as reported in Takada (2012), it is difficult to compare Riney's result to our result. Taking everything into consideration, we concluded that there is no consistent standard VOT threshold observation in the previous studies on Japanese. As we could not control for dialectal variability in VOT thresholds in our participant sampling, we instead relied on our own empirical measure of the actual threshold for our specific stimuli for the participants, using a standard behavioral identification test of the stimulus sequence used in Hestvik and Durvasula (2016) (see Section 5 below for details).

2.2. ERP results

2.2.1. Post-recording signal processing and artifact correction

After recording, the raw continuous EEG data were imported into the ERP PCA "EP" MatLab tool (Dien, 2010, 2017). The EP tool was used for all post-processing (including segmentation, baseline correction, artefact correction, and signal averaging). The data were first

[2] The reason we estimated the VOT threshold and selected stimuli based on the first 40 participants only, is that 22 participants were added at the end of the data collection to balance out the order groups. Their measured thresholds did however not alter the mean of the first 40, cf. Fig. 3.

[3] The mean identification rates, i.e., categorization as either /d/ or /t/, as a function of VOT: 10ms: 98%, 15ms: 93%; 20ms: 91%, 25ms: 84%; 45ms: 91%; 50ms: 94%; 55ms: 96%; 60ms: 97%. This shows that the participants clearly assigned the selected stimuli into two categories.

segmented into −200 ms to 800 ms epochs, for the four stimulus conditions (standard-d, standard-t, deviant-d, deviant-t). The data were baseline-corrected by subtracting the mean of the 200 ms baseline period from the whole segment, and then submitted to an automatic artifact correction procedure. Bad channels were identified by computing a moving average window of 80 ms; a channel was marked bad if the difference between minimum and maximum voltage exceeded 100 μV. A trial was marked bad if it contained more than 10% bad channels. Six neighboring electrodes were determined for each channel, and a channel was marked globally bad if it was below a threshold of predictability from the neighbors (less than r = 0.4) or if it differed by at least 30 μV from the most similar neighboring electrode. A channel was also declared globally bad if it was bad in more than 20% of the trials. Bad channels were replaced by the spline interpolation from the neighboring channels. Movement artifacts were subtracted after ICA decomposition, followed by remixing of the data. Trials with eye blinks were removed by the bad channel identification procedure. The data were finally re-referenced to the linked mastoids, and no additional offline filtering of the data was conducted.

After this procedure, one participant had very few remaining trials (only 13%). The remaining participants had on average 75% good trials (SD = 16%), ranging from 31% to 98% good trials. Picton et al. (2000) recommend using a cut-off value for proportions of good trials per cell for excluding participants. However, inspection of each of the 10 participants with as few as 31% good trials and at most 49% good trials revealed that even the participant with only 31% good trials had a very clear Auditory Evoked Potential and MMN. Rather than using a cut-off value for proportions of bad trials per cell to exclude subjects, we therefore decided to include all participants in the calculation of principal components and average waveforms, but tested at the point of inferential statistical analysis the effect of including or excluding low trial count participants.

2.2.2. PCA factor decomposition

One of the techniques recommended by Luck and Gaspelin (2017) to avoid experimenter bias in time window and electrode region selection, and to reduce the multiple comparisons problem, is to use Principal Component Analysis (PCA) to statistically determine the underlying temporal and spatial dynamics of the experimental effects in the data. We employed this method by using the factor analysis approach developed by Dien and colleagues (Dien et al., 2005; Dien et al., 2003, 2004; Spencer et al., 1999, 2001). Following published recommendations (Dien, 2012; Dien et al., 2005), we used sequential temporo-spatial PCA decomposition to identify the set of discrete or orthogonal temporal events in the voltage fluctuations, as well as discrete spatial regions of activity within each temporal event. Furthermore, as we are only interested in those brain responses that are caused by the experimental manipulations, we first reduced the data to difference waves: deviant /t/ minus standard /t/, and deviant /d/ minus standard /d/. A matrix with time samples as columns, and subjects, difference wave voltage values and electrodes as rows, was submitted to a temporal PCA, using the covariance matrix with Kaiser loading weighting. We then compared the scree plot of this full PCA (with the number of factors equal to the number of time samples) to the scree plot of a PCA of a random time point permutation of the data. The point where the two scree plots intersected (i.e., the "elbow") divides the set of factors in the PCA to those that can be interpreted as meaningful from those that can be interpreted as equivalent to the factors in a random permutation of the data and therefore not meaningful—this is the "parallel test" (Dien, 1998), which determines how many factors to retain. The parallel test resulted in 15 temporal factors. The PCA was then rerun restricted to 15 temporal factors (based on the covariance matrix, PROMAX rotation, and Kaiser factor loading), resulting in a solution that accounted for 90% of the total variance.

Using the same parameters as for the temporal PCA, each temporal factor was then submitted to a spatial ICA decomposition, which resulted in three spatial factors retained for each temporal factor. This further narrowed down the variance accounted for, such that the first spatial factor in each of the temporal factors accounted for most of the variance (for example, for TF1[4] which accounted for 28% of the variance, TF1SF1 accounted for 22%, with the rest of the spatially driven variance being allocated to TF1SF2 and TF1SF3). We therefore focused analysis on the first spatial factor of the first 5 temporal factors, and discarded from analysis temporal factors 6–15, which each accounted for <4% of the variance in the original temporal PCA. Fig. 4 illustrates the 5 main temporo-spatial underlying components in the data.

We will refer to the difference between deviant-t and standard-t as the "t-MMN," and the difference between deviant-d and standard-d as "d-MMN" in all analyses. The scores for each factor, which represent the weighted average of all time points and all channels for each subject and cell for that factor, were submitted to a repeated measures ANOVA with the two difference waves t-MMN and d-MMN as a repeated measures factor, and the two block order groups as a between-subject factor (i.e. deviant-d as first deviant vs. deviant-t as first deviant). The intercept, corresponding to the main effect of mismatch, was significant in all five temporo-spatial factors. The dependent measure was difference waves, thus a significant intercept represents a main effect of mismatch; the result shows that all five factors represent significant mismatch effects in discrete temporal events of voltage fluctuations.

However, in order to limit the scope of the analysis, we focused only on the two early effects, namely the 116 ms and the 196 ms temporal mismatch factors, which overlap with the early auditory ERP components N1 and P2. We will refer to these temporal components as the "N1 time window mismatch effect" and the "P2 time window mismatch effect".[5] These two time windows are precisely the time course where the MMN is typically observed, and the predictions based on underspecification are specifically for the MMN, which is why we focus attention on these time components. The temporally later components in the factor decomposition belong to the family of "late attention related ERPs" in oddball paradigms, and are outside the scope of the current analysis. (For example, TF1 is interpretable as the Late Negativity ERP, which indicates that a stimulus difference has risen to the level of the participant's awareness (Sussman et al., 2014).

2.2.3. Statistical analysis of factor scores

Temporal factor 4 peaks at 116 ms and overlaps with the N1 component of the Auditory Evoked Potential (AEP). The factor scores for TF4SF1 were submitted to a mixed factorial repeated measures ANOVA with phoneme as the within-subject independent variable (/t/ vs. /d/) and block order as the between-subject group variable (deviant t-first vs. deviant-d first). The dependent variable was the mean difference score for each phoneme contrast (t-MMN and d-MMN). Since factor scores represented the weighted average of all time points and all electrodes for the specific temporo-spatial factor, there are no additional time window or electrode "region of interest" factors. This ANOVA resulted in a main effect of phoneme ($F(1, 60) = 4.21$, $p < 0.05$, $\eta_p^2 = 0.06$), and a significant interaction between phoneme and block order group ($F(1, 60) = 6.9$, $p < 0.05$, $\eta_p^2 = 0.10$), such that the primacy effect difference between t-MMN and d-MMN was significantly greater for the group that heard [t] as the first deviant, than for /d/ for the group that heard [d] as the first deviant (see Fig. 5, top panel).

The t-MMN derived from deviants in the first block for the group that heard [t] as deviant first, had the overall greatest negative value. The difference between t-MMN and d-MMN in the N1 time window was

[4] We refer to temporal factor 1, 2, 3, etc. as TF1, TF2 and so on, and spatial factors as SF1, SF2, etc. TF1SF1 refers to the first spatial subfactor under temporal factor 1.

[5] Note that "MMN" in the literature often refers to a mismatch effect spanning both these time regions (Näätänen et al., 1988).

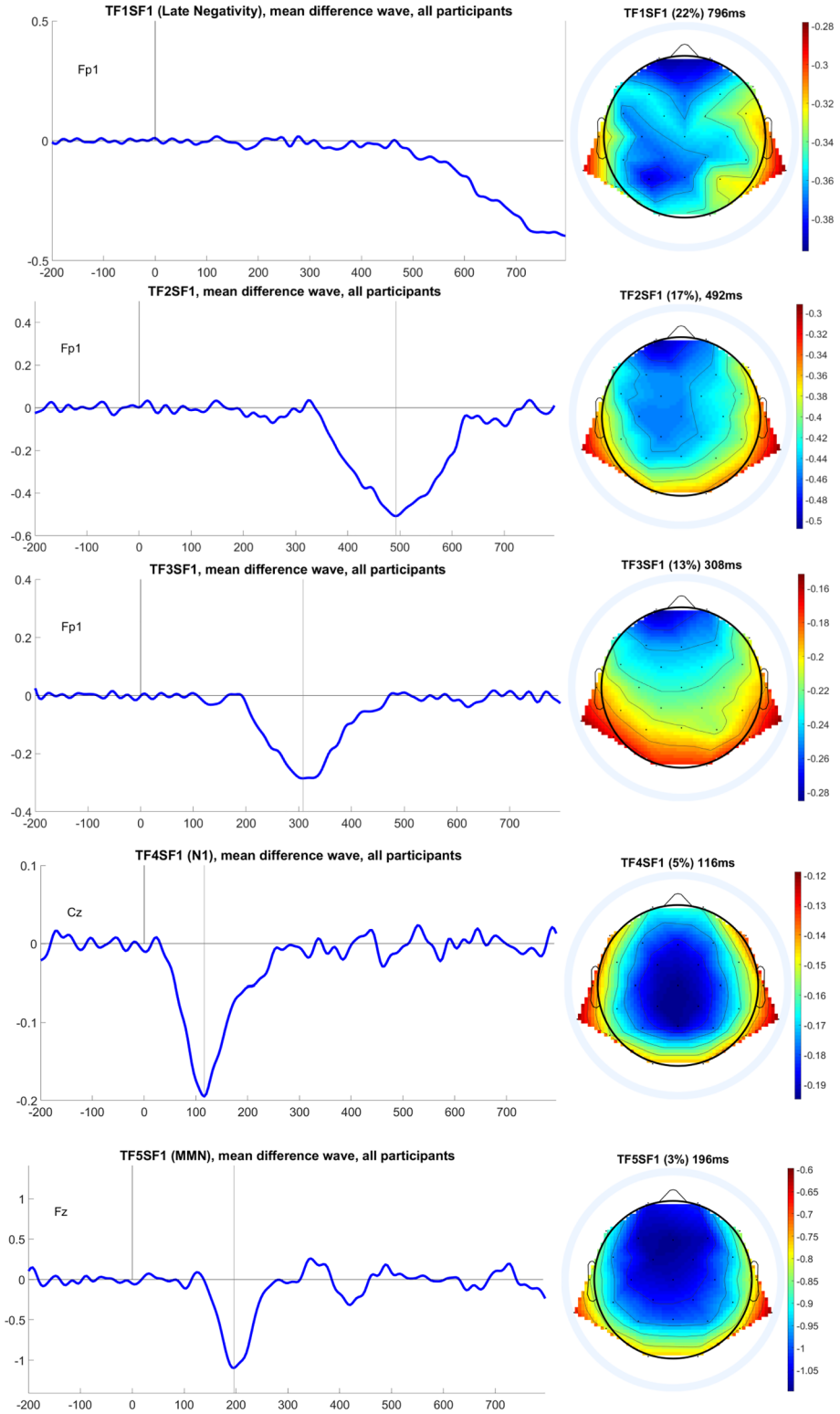

Fig. 4. Temporo-spatial factor decompositions. Each graph shows factors reconstructed as voltage on the Y-axis and time on the X-axis.

further assessed by planned orthogonal contrasts within each group. The difference between t-MMN and d-MMN was highly significant for the deviant-t first group ($t(31) = 3.331$, $p < 0.01$, effect size = 1.77 mV), but not significant for the deviant-d first group ($t(31) = 0.4$, $p > 0.05$, effect size = −0.21 mV).

Factor TF5SF1 was the next temporal component of the mismatch effect, overlapping with the P2 component in time and space. The intercept (interpretable as the main effect MMN, as the dependent measure were the difference waves), was highly significant ($F(1, 60) = 39.2$, $p < 0.0001$, $\eta_p^2 = 0.395$), which can be seen in the bottom panel of Fig. 5, because every single cell data point is below zero. The main effect of phoneme was not significant ($F(1, 60) = 0.43$, $p > 0.05$, $\eta_p^2 = 0.007$). However, the interaction between phoneme and block order was significant ($F(1, 60) = 4.85$, $p < 0.05$, $\eta_p^2 = 0.074$). Inspection of the interaction plot (Fig. 5, bottom panel) again revealed that the interaction was driven by a greater difference between the t-

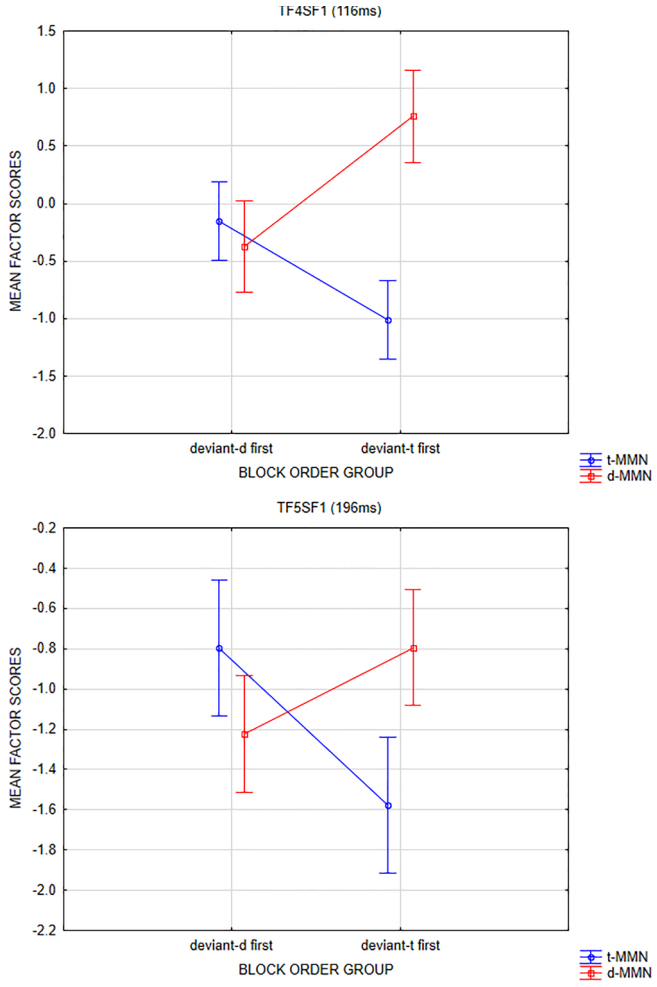

Fig. 5. Interaction plots from statistical analysis of factor scores per subject and cell for TF4SF1 and TF5SF1. Vertical bars denote ±1 standard errors.

MMN vs. d-MMN for the group that heard deviant-t first. The MMN with the greatest negativity was t-MMN for deviant-t first group (−1.6 mV). A planned orthogonal contrast analysis comparing the difference between t-MMN and d-MMN within each group revealed that the difference was significant for the deviant-t first group ($t(30) = 2.02$, $p < 0.05$, effect size = 0.78 mV), but not for the deviant-d first group ($t(30) = -1.09$, $p > 0.05$, effect size = 0.32 mV). Thus, the interaction between primacy (block order) and phoneme in both factors was driven by [t] as deviant in the first block of the t-as-first-deviant group.

2.2.4. PCA-constrained analysis of voltage

We next used the factor solution above to constrain the selection of time windows and electrode regions in the undecomposed voltage data for statistical analysis, as a way of both confirming the PCA analysis, and seeking convergence between the PCA and analysis of the mixed (in the PCA/ICA sense) surface voltage observations. First, we used the factor loadings of TF4 and TF5 to select time windows composed of those contiguous time samples with factor loadings exceeding 0.6[6] (See Fig. 6).

[6] Note that using a threshold of 0.6 is an arbitrary cut-off, used merely to ensure that a narrow enough window and electrode region is selected, to avoid having the experimental effect cancel out by including too many *unweighted* time samples and/or electrodes. We follow Dien (2019, p. 108) in using 0.6 as a default; however for the N1 time window effect, we had to increase the spatial threshold to 0.9 to avoid including the entire montage. The threshold used is simply a way to narrow down time and space to create a temporally and spatially delimited voltage effect, constrained by the PCA solution, and which

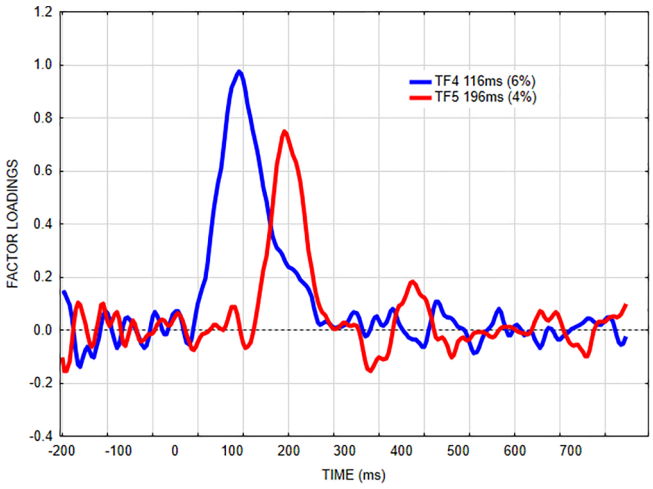

Fig. 6. Temporal factor loadings over time for TF4 and TF5; legend indicates peak latency and amount of variance accounted for.

For the early, N1 time window, this corresponded to 80–152 ms. In Fig. 7 we illustrate how the factor corresponds with and isolates a specific temporal component of the mismatch effect by overlaying the factor waveform back projected into voltage space, with the raw voltage waveforms for each group and for the primacy affected condition.

As is evident in Fig. 7, the effect of mismatch in this time region is greater for /t/ in the deviant-t first group, than it is for /d/ in the deviant-d first group.[7]

To assess the statistical significance of this for both groups and both difference waveforms, we first computed the mean voltage values for the N1-overlapping 80–152 ms time window, averaged over electrodes with factor loadings greater than 0.9 in the spatial factor (a higher threshold was used here because the effect was present in all electrodes, therefore using 0.6 would have selected all electrodes for the "region"). The mean voltage per participant and difference waveform was then submitted to a repeated-measures mixed factorial ANOVA with the within-subject factor phoneme (/t/ vs /d/), and the between-subjects factor block order group (deviant-t first vs. deviant-d first). This resulted in a marginal effect of phoneme ($F(1, 60) = 9.57$, $p < 0.06$, $\eta_p^2 = 0.057$), such that the t-MMM was greater than the d-MMN (-0.48 mV vs. 0.08 mV respectively), and a significant interaction between phoneme and block order group ($F(1, 60) = 9.05$, $p < 0.005$, $\eta_p^2 = 0.131$). The interaction was driven by the primacy effect for the t-MMN being significantly greater than the primacy effect for d-MMN. Orthogonal contrast analysis showed that in the deviant-t first group, the difference between t-MMN and d-MMN was highly significant ($t(31) = 3.47$, $p < 0.001$, effect size = 1.42 mV), whereas the corresponding contrast between d-MMN and t-MMN for the deviant-d first group was not significant ($t(31) = 0.77$, $p > 0.05$, effect size = −0.31 mV).

As a check of robustness and precision, we also conducted an independent samples *t*-test comparing only the early N1-related mismatch effect for the first deviant in each group, i.e., deviant-t for deviant-t first group vs. deviant-d for deviant-d first group. This test was not significant (t = 0.93, effect size = −0.42 mV). We attribute this to insufficient power, as this comparison only uses half the data and relies on a between-subject comparison, but it also suggests that the results should be replicated with either more power or a design that does not

(*footnote continued*)
contains an experimental effect.

[7] We omit for space reasons detailed graphical illustration of deviant-t for the deviant-d first group, and deviant-d for deviant-t first group, as the statistical analysis shows that the mismatch is greatly reduced in these conditions.

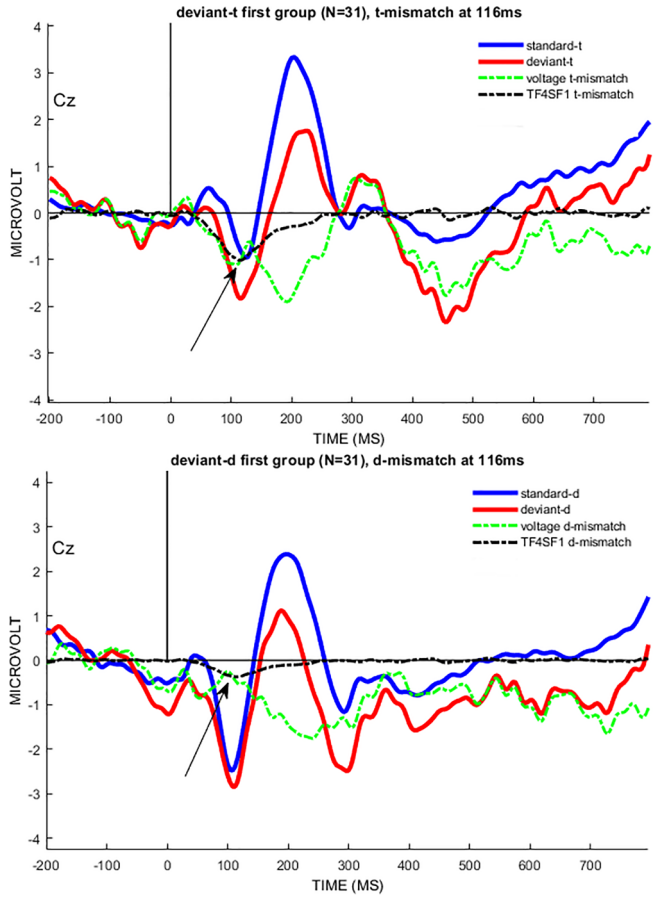

Fig. 7. Top panel: Standard-t, deviant-t and t-MMN difference waveforms for the deviant-t first group along with factor TF4SF1 waveform for t-MMN. Bottom panel: Standard-d, deviant-d and d-MMN difference waveforms for the deviant-d first group along with factor TF4SF1 waveform for d-MMN.

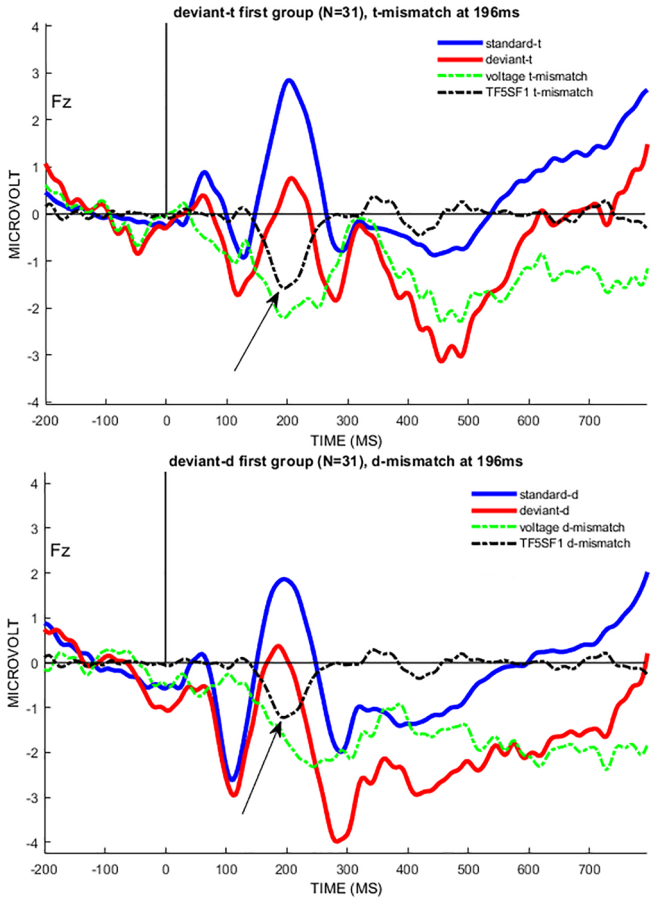

Fig. 8. Top panel: Standard-t, deviant-t and t-MMN difference waveforms for the deviant-t first group along with factor TF5SF1 waveform for t-MMN. Bottom panel: Standard-d, deviant-d and d-MMN difference waveforms for the deviant-d first group along with factor TF5SF1 waveform for d-MMN.

introduce a primacy effect.

We conducted the same analysis for the subsequent temporal factor TF5 (peak latency 196 ms), using the time samples 180–216 ms which exceeded 0.6 in their factor loading values. We again selected the electrode regions exceeding 0.9 factor loading values from the primary spatial factor in each of the two temporal factors. Fig. 8 illustrates how the temporo-spatial factor decomposition captures the second temporal voltage fluctuation caused by the experimental conditions, focusing on the primacy effected phonemes in each group.

We again computed the mean voltage per participant and cell (t-MMN and d-MMN), for the electrode region consisting of electrodes with factor loadings exceeding 0.9 in TF5SF1 (i.e., essentially the blue area in Fig. 4, bottom panel). The resulting dependent measures were submitted to ANOVA with the two difference waves as two levels of phoneme (t-MMN and d-MMN) and block order group as between-subject factor. This resulted in a significant intercept, i.e., both difference waves were significantly below zero, such that a mismatch negativity effect was observed for both phonemes ($F(1, 60) = 29.22$, $p < 0.00001$, $\eta_p^2 = 0.327$). The main effect of phoneme was not significant ($F(1,60) = 0.6$, $p < 0.5$, $\eta_p^2 = 0.005$), but we again observed a significant interaction between phoneme and block order group ($F(1, 60) = 6.69$, $p < 0.05$, $\eta_p^2 = 0.100$), such that the difference between /t/ and /d/ for the deviant-t first group was larger than the difference between /t/ and /d/ for the deviant-d first group. Orthogonal planned contrast analysis compared t-MMN to d-MMN for the deviant-t first group, which was significant ($t = 2.28$, $p < 0.05$, effect size = 1.25 mV). The contrast between d-MMN and t-MMN in the deviant-d first group was not significant ($t = 1.27$, $p > 0.05$, effect size = 0.6 mV). Again, the largest negativity in this interaction was observed for /t/ in the deviant-t first group (−1.8mv). The interaction plots for the voltage analysis for both factors are given in Fig. 9.

As a check of robustness and precision we also conducted an independent samples *t*-test comparing only the mismatch effect for the first deviant in each group, i.e., deviant-t for deviant-t first group vs. deviant-d for deviant-d first group. This difference did not reach significance ($t = 0.8$, $p > 0.05$, effect size = 0.51 mV). This also shows that the observed asymmetry is a statistically small (although meaningful) effect, and is only observed in the block order X phoneme interaction, i.e. in how the phoneme responds to the primacy effect.

3. Discussion

We observed and analyzed mismatch effects in two early time windows (determined by temporal PCA of the difference waves): A time window overlapping with the N1, 80–152 ms (peak latency 116 ms) and a time window overlapping with the P2, 180–216 ms (peak latency 196 ms). The results mirror the same pattern of statistical results that we observed in the same experiment but with English-speaking participants in Hestvik and Durvasula (2016). As we predicted from phonological analysis of underspecification in Japanese, the crucial interaction between the primacy effect and phoneme-specific mismatch effect goes in the opposite direction for Japanese-speaking participants. The combined results of Hestvik and Durvasula (2016) and the current study confirm the predictions of a phonological theory in which /t/ is underspecified in Japanese, whereas /d/ is underspecified in English, coupled with the Eulitz/Lahiri linking theory between MMN and

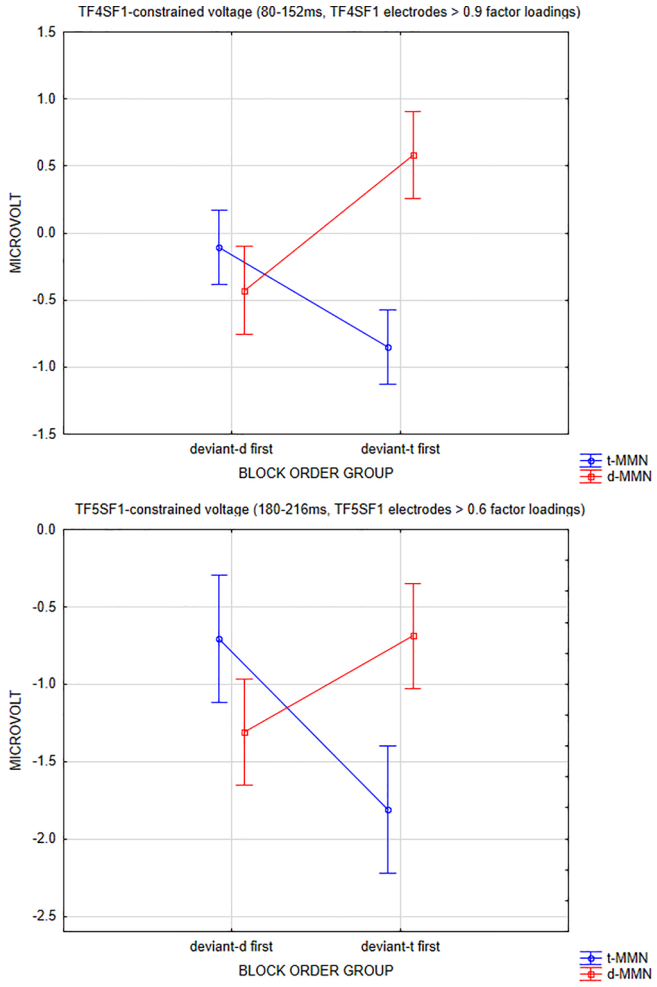

Fig. 9. Interaction plots from statistical analysis of PCA-constrained time windows and voltage regions per subject and cell for TF4SF1 and TF5SF1. Vertical bars denote ±1 standard errors.

underspecification.

The two time-windows where we observe mismatch effects overlap with the N1 and the P2 peaks of the AEP, respectively. There is a debate about whether MMN should be primarily understood as a modulation of N1, related to whether MMN results from predictive coding mechanisms vs. adaptation of neuronal responses to the standard stimuli, as argued by May and Tiitinen (2010). Note however that we computed the mismatch effects using the identity MMN paradigm, hence any difference in the N1 time window mismatch effect is independent of stimulus-specific modulation of the N1, as also argued by Näätänen et al. (1988). Kielar et al. (2008) also observed phonetic MMNs in similar time windows (80–160 ms, and 180–280 ms) as we did.

Note also that the larger mismatch effect for voiceless [t] compared to voiced [d] in the N1 time window should not be confused with the observations that long vs. short VOT values appear to correlate with amplitude differences in the N1 (Toscano et al., 2010). We conducted an analysis of the N1 responses to the standards only, and the N1 amplitude was more than 1 mV more negative for /t/ than for /d/, which is consistent with Toscano et al., (2010). However, the data analyzed here are the difference scores within-category (e.g. deviant-t minus standard-t), hence the absolute N1 modulation related to VOT is controlled for in our analysis based on difference waveforms within phoneme.

In the P2 time window, the asymmetry is only seen when the interaction between block order and phoneme is unpacked, whereas in the N1 time window, the phoneme difference survives the primacy effect in the factor score analysis and marginally so in the voltage analysis. The block order effect is a result of primacy bias as a function of the order with which a stimulus is presented in counterbalanced designs, i.e., the deviant presented in the first half of the experiment has a larger MMN than the deviant presented in the second half. This primacy bias interacted with phoneme, such that the advantage of being the first deviant was significantly greater for /t/ than for /d/. Note that the primacy effect is not a confound, but rather an experimental task effect that attenuates the MMN for deviants presented in the second part of the experiment, which has the effect of reducing the mean MMN when averaged across both blocks. If the contrasts are computed based on averaging the data across the two blocks without taking the primacy effect into account, the overall mean MMN is attenuated to the degree that the phoneme effect becomes invisible. Our conclusion is that a certain “freshness” in the MMN is required to observe the difference in MMN amplitude between /t/ and /d/, which is why we focus in the first block deviant. Our understanding is that underspecification of voicing in /t/ drives how the MMN for /t/ vs. /d/ changes in response to the primacy effect. We were aware of the primacy effect going into the study, as we observed it in our previous experiments, so we planned to take it into account in the analysis of the current data. The primacy effect has also been well documented and studied by Juanita Todd and her colleagues, and we now understand it to be an intrinsic property of “identity-MMN” designs. The only way to avoid the primacy effect is to use a different experimental design (e.g. using the random standards control condition of Horváth et al. (2007) or boost power by dramatically increasing N.

3.1. Limitations of the current study

The most direct evidence for underspecification would be a simple difference between phonemes with respect to MMN amplitude (or, if absolute waveforms were analyzed, an interaction between phoneme and stimulus condition). This simple effect was observed in the N1 time window mismatch effect by a significant two-tailed *t*-test between t-mismatch and d-mismatch in the factor score analysis, and by a marginally significant two-tailed *t*-test in the voltage analysis ($p < 0.06$), which can be interpreted as significant under our one-tailed hypothesis. A limitation of the current study is that the simple contrast was not significant in the later MMN time-window mismatch effect. Our analysis is that the asymmetry in the MMN time window is obscured by an ordering effect. In addition, an independent samples *t*-test comparing the first-block deviant from the deviant-t-first group vs. the first-block deviant in the deviant-d-first group did not reach significance, which we attribute to the lower power of this test. Thus, future studies should seek to replicate the findings using paradigms with greater power and which do not induce ordering effect, such as the “random standards” control condition version of MMN experiments (Horváth et al., 2007, 2008; Rhodes et al., 2019).

An inherent limitation in the current study is that we intentionally used the same synthetic stimuli as used in Hestvik and Durvasula for English-speaking participants, but these stimuli may not have been typically Japanese-sounding. As suggested by a reviewer, this could make it less likely that the stimuli activate native language categories, thus lowering the effect size. A future replication should be done with ecologically valid, naturally sounding Japanese stimuli.

3.2. Possible objections

A possible objection to our results is that the empirical phoneme frequencies (i.e. the total observed frequency in language use) could account for the observed asymmetry, if a deviant phoneme with a higher overall frequency in the language generates a larger MMN (Alexandrov et al., 2011; Scharinger and Samuels, 2017). In Japanese, the phoneme /t/ is in fact more frequent than the phoneme /d/, for both type and token frequencies (Tamaoka and Makioka, 2004).

However, note that the token frequency of the phoneme /t/ is higher than /d/ in English as well (Hayden, 1950; Mines et al., 1978). If this was the explanation, we would not have observed the asymmetry we did in Hestvik and Durvasula (2016). Furthermore, a search through the CMU Dictionary (Weide, 1994), a pronunciation dictionary of American English with over 134,000 pronunciation entries, also confirms that the type frequency of /t/ is larger than that of /d/, with 48,410 /t/ count (5.7%) and 32,310 /d/ count (3.8%). Thus, in both Japanese and English, /t/ is more frequent than /d/, with respect to both type and token frequencies, but the MMN asymmetry in the two languages go in opposite directions. Therefore, an account of the current results based on token phoneme frequencies seems not to be feasible.

A similar possibility suggested by a previous reviewer is that of biphone frequencies, as they have been shown to modulate the MMN (Bonte et al., 2005). For the Japanese participants, the [æ] in the stimuli is very likely perceived as /a/ (Lengeris, 2009; Strange et al., 1998), therefore the relevant biphone (type) frequencies are those of /ta/ and /da/. In Japanese, /ta/ is much more common than /da/ (/ta/ = 27,384, /da/ = 17,622) (Tamaoka and Makioka, 2004, 2009). However, again, a similar bias exists for "t" in English based on a search through the CMU Dictionary (/tæ/ = 1,138, /dæ/ = 529). Thus, as with the phoneme frequency discussion above, the differing MMN asymmetries in English and Japanese cannot be accounted for by different biphone frequencies in the two languages.

3.3. Contradictory findings

Our results differ from another recent study that tested similar cross-linguistic predictions for MMN asymmetries based on language-specific underspecification analysis. Schluter et al. (2017) examined MMN asymmetries in speakers of Arabic and Russian, which also have the opposite underspecification of voicing in stops compared to English. They used natural recordings of voiced and voiceless fricatives as well as voiced and voiceless stops in Russian, English and Arabic, and conducted a varying standards MMN experiment with Russian, English and Arabic speakers. Contrary to the predictions that we make for Japanese, and which would be the same prediction for Russian and Arabic, they observed that /d/ elicited a larger MMN than /t/ for all language groups, no matter what the phonological grammar is. They conclude that the feature [spread glottis] is universally used to encode the voiced/voiceless contrast, and that /d/ is universally underspecified for this privative feature. This result is contrary to our findings reported here, and it remains a puzzle how to consolidate these contradictory findings.

Our conclusions about abstract phoneme categories contradict many studies reporting that lexical access is modulated by gradient information available to the perceptual system (McMurray et al., 2008; McMurray et al., 2003; Mitterer, 2011; Mitterer et al., 2018; Toscano et al., 2010) . Our interpretation of these reported effects is that the experiments tap into the processes that relates phonetic representations to phonemic representations, and not the phonemes themselves. For example, stored knowledge about VOT must necessarily be used to determine whether a phonetic signal corresponds to a voiced or a voiceless consonant. Our view is that this information is utilized by the *process* that maps acoustic/phonetic information to discrete, digital phonemes representations in memory. In other words, VOT information is stored in the mapping system but not in the phonemes to which they map (see also Rhodes et al., 2019).

4. Conclusion

Based on the current study and a comparison with the results of Hestvik and Durvasula (2016), we conclude that the observed mismatch effects for /t/ vs. /d/ and how the two phonemes respond differently to the primacy effect support the theory that English /d/ is underspecified compared to /t/, whereas Japanese exemplifies the opposite, with /t/ underspecified and /d/ specified for voicing. This provides further neurobiological evidence for the abstractness of phoneme representations in long-term memory, and for cross-linguistic differences in underspecification. This result was obtained by presenting both speaker groups with identical stimuli, thus keeping the stimuli constant and forcing variance to be driven by non-stimulus properties, namely the phonological grammar of the participants and the language-specific underspecification patterns.

5. Methods and materials

5.1. Participants

Sixty-eight participants were recruited from among students at Waseda University in Tokyo, Japan. Participants signed an informed consent form before entering the study. They first participated in a behavioral pre-test experiment and then an EEG recording on separate days (see Procedures). Each participant was paid 1000 JPY for participation in the behavioral pre-test experiment and additional 2000 JPY for the EEG experiment. Six participants did not return for the EEG session and were excluded from further analysis. The remaining 62 subjects were all native speakers of Japanese, who reported no history of speech and hearing impairments using a questionnaire. In addition, all participants met criteria for being monolingual. None reported having lived outside Japan for more than two months; all had parents who were both native Japanese speakers, and all participants reported speaking only Japanese through their daily lives. Half of the participants were women. The average age was 19.7 years (SD = 1.6, range 18–24), equally distributed by sex. All participants were right-handed except one participant who was ambidextrous, as assessed by the FLANDERS handedness questionnaire (Nicholls et al., 2013).

5.2. Stimuli

We used the exact same stimuli as in Hestvik and Durvasula (2016), which in turn were created by replicating the stimulus parameters reported in Phillips et al. (2000). The stimuli were not typical sounding Japanese CV syllables. Indeed, they may have sounded artificial or foreign to the participants, due to the vowel quality after the consonant and the nature of the synthetically constructed burst. However, we intentionally opted to use the same stimuli as in Hestvik and Durvasula (2016), rather than create a more "Japanese sounding" stimulus set. The reason was two-fold: First, in order to interpret the data relative to the findings in Hestvik and Durvasula (2016), we wanted to only vary a single variable, namely the phonological grammar in the brain of the speakers, and avoid any confound introduced by using different stimuli. Second, we also wanted to address another possible alternative explanation of the Hestvik and Durvasula findings, namely that the asymmetry we observed there could be due to intrinsic acoustic differences between the stimuli. It has been shown that long vs. short-lag VOT results in distinct neural response patterns in monkeys (Steinschneider et al., 1995), and this could conceivably lead to different amplitude in the MMN, not as a result of phoneme representations but then as a result of low-level pre-linguistic neural response patterns. If this were the case, then the Japanese participants' brain responses should not be distinguishable from those of the English-speaking participants in Hestvik and Durvasula (2016). By using the exact same stimuli with Japanese participants, we can also test this prediction.

The full stimulus set consisted of 17 CV syllables that ranged from /dæ/ (with 0 ms VOT) to /tæ/ (with 80 ms VOT). Each syllable varied only in voice onset time (VOT) in 5 ms increments, created using the online version of the low-level Klatt synthesizer (Bunnell, 1999; Klatt, 1980). Sampling rate was 22,050 Hz. The average sound level of all stimuli when presented with ear inserts through both right and left ears was 72 dB. The duration of the sound stimulus was 290 ms. The full

Table 2
Experimental design matrix.

			Stimulus condition (within-subject)		Cell label/dependent measure
	Group	Block	Standard	Deviant	Deviant minus standard
Block order (between subject)	[t] as 1st deviant	1	720 /d/ standards	100 /t/ deviants	#1: t-MMN, t as first deviant
		2	720 /t/ standards	100 /d/ deviants	#2: d-MMN, t as first deviant
	[d] as 1st deviant	1	720 /t/ standards	100 /d/ deviants	#3: d-MMN, d as first deviant
		2	720 /d/ standards	100 /t/ deviants	#4: t-MMN, d as first deviant

sequence was used in behavioral pre-testing of each subject's categorical identification function, and a subset was used in the ERP experiment (see below). Based on the pre-test findings (see below), we selected as set of four stimuli on each side of the observed mean threshold of about 33 ms for use in the MMN experiment. In addition to the MMN stimuli, two "ba" and two "pa" syllables were used in an attention-controlling tracking task. For the purpose of the tracking task, two of these tracking syllables were edited to sound like a female voice, and two to sound like a male voice. MMN is an automatic response that occurs even in the absence of attention, although directing attention to the stimulus stream has been shown to enhance the MMN (Haroush et al., 2010). Therefore, we opted to use the attention-ensuring tracking task.

5.3. Design

The current study was an exact replication of Experiment 1 in Hestvik and Durvasula (2016), except the participants' first language was Japanese. The paradigm is an oddball paradigm, where one stimulus is frequently occurring in a repetitive sequence (the "standard"); interrupted infrequently by the oddball stimulus (we will refer to it as the "deviant"). The repeated standard stimuli result in a memory trace which generates predictions about the upcoming stimulus in the sequence. If the stimulus is the deviant, it does not conform to the prediction, and this generates an attenuation of the Auditory Evoked Potential (AEP) elicited by the deviant. This effect can be seen by computing a mean AEP for the standard stimuli and compare it to the mean AEP for the deviant stimulus. The subtraction waveform obtained by taking the deviant minus the standard waveform is called the Mismatch Negativity waveform.

As in Hestvik and Durvasula (2016), we utilized a counterbalanced oddball paradigm design, in order to control for the possibility that the Mismatch Negativity observed for the deviant stimuli is not due to intrinsic properties of the stimuli. This is done by running the experiment in two blocks. In the first block, stimulus A is the standard and stimulus B is the deviant. In the second block, this is reversed so that stimulus B is the standard and stimulus A is the deviant. Subsequently, the Mismatch Effect is computed by comparing, say, stimulus B as deviant in the first block to itself as a standard in the other block, and vice versa for the other stimulus. This is the "identity MMN" paradigm (Pulvermüller and Shtyrov, 2006; Pulvermüller et al., 2006). In this way, one can observe the pure modulation of the AEP by the mismatch context and control for intrinsic stimulus differences between the standard and deviant.

One drawback with this method is that it introduces a confound of its own. The auditory system adapts to the presentation of the standard stimulus in the first block, so that when the standard stimulus next serves in the role of a deviant, it has already been heard 700 times in the first block. This attenuates the MMN response to the same stimulus as deviant in the second block. This "primacy bias" effect has been observed in the previous MMN-literature, and was first reported by Todd et al. (2011), and has subsequently been examined in depth in a series of studies (Fitzgerald et al., 2018; Frost et al., 2018; Frost et al., 2016; Mullens et al., 2016; Mullens et al., 2014; Todd et al., 2014; Todd et al., 2008; Todd et al., 2016; Todd et al., 2013).

For this reason, the order of the blocks must itself be counterbalanced between subjects, introducing a between-subjects methodological factor. This is especially important in the current study, where it is hypothesized that the MMN amplitude of one stimulus should be smaller than the MMN amplitude of the other stimulus. Therefore, we must make sure that any observed asymmetry is not simply due to a "primacy bias". Consequently, Hestvik and Durvasula (2016) divided the subjects into two different block-order groups for the purpose of counterbalancing the MMN. Indeed, the key asymmetry that was observed between /d/ and /t/ in their English participants was only evident once this interaction was unpacked. For the current study with Japanese participants, we therefore employed this counterbalancing scheme as well. Participants were randomly assigned to one of two stimulus order groups with 31 participants in each order group for the EEG recording session. In one order, the stimulus /tæ/ was the deviant in the first half of the experiment (the 'deviant-t first' group); in the other, /dæ/ was the first deviant stimulus (the 'deviant-d first' group). The full design is illustrated in Table 2, showing the relationships between the within-subject independent variables stimulus condition (standard, deviant) and phoneme (/d/ vs. /t/), and the between-subject variable block order (i.e. /t/ as first deviant vs. /d/ as first deviant).

Table 2 describes the experiment and the trials in each single cell. In addition, 68 target sounds were randomly interspersed with the experimental stimuli. The standard/deviant factor can be analytically eliminated by computing difference waveforms as a dependent measure, which are described as the four MMN effects in the table. The predictions can now straightforwardly be stated with reference to the MMNs. First, we predict a main effect of stimulus condition, so that the mean of all four MMNs in Table 2 should be significantly below zero. Turning to our hypothesis, we predict a main effect of phoneme, so that t-MMN is larger than d-MMN (i.e. the mean of MMN# 1 + 4 should be greater than the mean of MMN# 2 + 3).

Next, we predict a main effect block order (i.e. the "primacy bias"), so that the mean of MMN#1 and MMN#3 should be significantly larger in amplitude than the mean of MMN#2 and MMN#4. Note that if the MMNs of /t/ and /d/ were symmetrical (i.e. equally large in amplitude), we would only expect a main effect of primacy. However, based on the findings with the same stimuli and same experimental design in Hestvik and Durvasula (2016), we predict a significant interaction between block order and phoneme-MMN, such that the primacy effect should be greater for /t/ than for /d/. I.e., we expect there to be both a primacy bias in the data as well as an underspecification effect on the MMN, and when combined, underspecification adds a component to the interaction so that primacy bias does not simply result in a crossing interaction. All these predictions will be tested with orthogonal contrast planned t-tests in conjunction with ANOVAs, using both PCA factor scores and PCA-constrained voltage means as dependent measures. Effect size will be reported in microvolt and in terms of standard deviations (Cohen's *d*).

With respect to the behavioral pre-test, we expected participants to interpret the stimuli is falling into two categories, /d/ and /t/, but we could not predict which VOT value would represent the threshold between the two categories; which was why we conducted the behavioral pre-test to discover this and indeed test whether the stimuli did fall into two categories.

5.4. Procedures

5.4.1. Behavioral pre-testing

Participants first took part in a behavioral phoneme identification task using the /dæ/-/tæ/ VOT stimulus continuum. This task was administered with E-Prime experimental control software. Participants were presented with randomly ordered stimuli from the entire VOT continuum. For each stimulus, they needed to decide whether they perceived each stimulus as /tæ/ or /dæ/ by pressing 1 or 4 on a computer keyboard. The Japanese Hiragana characters 'た' and 'だ' were displayed on a computer screen over the numbers 1 and 4 on every trial to reinforce the button press assignment, and to create an association between the stimuli and the Japanese speech sounds. The stimulus continuum of 17 stimuli were presented in 11 blocks, with random order in each block, totaling 187 trials. The response buttons for /dæ/ and /tæ/ were switched after each block to avoid button press biases. A single trial can be described as follows: First, there is a pre-stimulus wait period of 1000 ms. Then the response screen with button indicator and hiragana symbols appear for 400 ms, followed by the stimulus, lasting 290 ms. The participant had a total of 2000 ms from stimulus onset to respond with their categorization decision. When they responded, a feedback showing their reaction time would be displayed for 500 ms. The behavioral test was conducted before the EEG measure because we needed to determine the empirical VOT threshold between /d/ and /t/ before selecting stimuli for the ERP experiment.

5.4.2. EEG recording session

Minimally three days after the behavioral session, each subject returned to the lab for the EEG recording session. Subjects were tested in a sound-proof booth while seated in a comfortable reclining chair. The continuous EEG was recorded from 32 sintered Ag/AgCl passive electrodes, adhered to the subject's scalp using an EEG recording cap (Brain Products GmbH Easy Cap 40 Asian cut, Montage No. 22). One of the 32 channels was used for recording horizontal eye movement (HEOG) and placed below the outer canthus of the right eye. One channel (in the position of AFz) was used as grounding, and one channel (otherwise used as FCz) was used as a reference electrode and attached to the nose. The remaining 29 channels were mounted onto the cap, according to the 10/20-system, with the electrode adaptor A06 using high-chloride, abrasive electrolyte gel (Abralyt HiCl). The impedance level for all the electrodes was reduced to below 8kΩ. The EEG was recorded using BrainAmp (Brain Products GmbH). The analog signal was digitized at 250 Hz and recorded without online filtering.

The ERP experiment was also programmed and controlled by E-Prime software. The participants were presented with a continuous sequence of a total of 1480 standards and 200 deviants (/dæ/ and /tæ/), along with 100 randomly interspersed "target" stimuli (/bɑ/ or /pɑ/) via ER1 insert earphones (Etymotic Research). The interstimulus interval (ISI) was varied randomly between 700 ms and 890 ms in increments of 10 ms. There was a break halfway through each of the two ordering blocks, i.e. after 350 standards, 50 deviants, and 25 target stimuli. After each break (including at the start of the experiment), the stimulus sequence was always introduced by 20 additional standards, resulting in a total of 740 standards and 100 deviants in each of the two ordering blocks. In total, 1680 trials as well as 100 target stimuli had been presented by the end of the experiment. Each stimulus sequence before a break or before the block order change (where there was also a break) lasted for about 10 min. The entire EEG recording including the three breaks took about 50 min. During the EEG recording, participants were instructed to listen to the stimulus stream and identify the gender of the voice of the target stimuli (/bɑ/ or /pɑ/) presented randomly in the sequence of standards and deviants, and to press 1 for a female and 4 for a male voice using a keyboard.

CRediT authorship contribution statement

Arild Hestvik: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing - original draft, Writing - review & editing, Visualization, Funding acquisition, Project administration. **Yasuaki Shinohara:** Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing - original draft, Writing - review & editing, Visualization, Supervision, Funding acquisition, Project administration. **Karthik Durvasula:** Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Writing - original draft, Writing - review & editing. **Rinus G. Verdonschot:** Writing - original draft, Writing - review & editing. **Hiromu Sakai:** Conceptualization, Methodology, Investigation, Resources, Writing - review & editing, Supervision, Funding acquisition, Project administration. Note: The first two authors share first authorship.

Acknowledgements

This work was supported by JSPS KAKENHI Grants 19K13169,16K16884 to Yasuaki Shinohara, and JSPS KAKENHI Grant 15H01881 to Hiromu Sakai, and Waseda University Grant for Special Research Projects 2016K-202 and 2017K-221 to Yasuaki Shinohara. An initial pilot project was supported by a 2009 University of Delaware General University Research grant to Arild Hestvik. We would like to thank our research assistants, Tzu-Yin Chen, Ai Ogawa, Laura Rodrigo, Yushi Sugimoto, and Yunzhu Wang, who helped us collecting EEG experiment data, and special thanks to Yingyi Luo (Chinese Academy of Social Sciences) for help in setting up the EEG laboratory for this project.

References

Alexandrov, A.A., Boricheva, D.O., Pulvermüller, F., Shtyrov, Y., 2011. Strength of word-specific neural memory traces assessed electrophysiologically. PLoS ONE 6 (8), e22999. https://doi.org/10.1371/journal.pone.0022999.

Archangeli, D., 2008. Aspects of underspecification theory. Phonology 5 (02), 183. https://doi.org/10.1017/S0952675700002268.

Avery, P., Idsardi, W.J., 2001. Laryngeal dimensions, completion and enhancement. In: Hall, T.A. (Ed.), Distinctive Feature Theory. Walter de Gruyter, Berlin, pp. 41–70.

Beckman, J., Helgason, P., McMurray, B., Ringen, C., 2018. Rate effects on Swedish VOT: evidence for phonological overspecification. J. Phonet. https://doi.org/10.1016/j.drudis.2014.09.014.

Beckman, J., Jessen, M., Ringen, C., 2013. Empirical evidence for laryngeal features: Aspirating vs. true voice languages. J. Ling. 49 (02), 259–284. https://doi.org/10.1017/S0022226712000424.

Bonte, M.L., Mitterer, H., Zellagui, N., Poelmans, H., Blomert, L., 2005. Auditory cortical tuning to statistical regularities in phonology. Clin. Neurophysiol. 116 (12), 2765–2774. https://doi.org/10.1016/j.clinph.2005.08.012.

Bunnell, H. T. (1999). Klatt Synthesis Interface. Wilmington: Speech Research Lab, A.I. DuPont Hospital for Children and the University of Delaware.

Chao, K., Chen, L., 2008. A cross-linguistic study of voice onset time in stop consonant productions. Comput. Ling. 13 (2), 215–232.

Chomsky, N., 1986. Knowledge of Language: Its Nature, Origin, and Use. Praeger, New York.

Chomsky, N., 1995. The Minimalist Program. The MIT Press, Cambridge, Mass.

Chomsky, N., Halle, M., 1968. The Sound Pattern of English. Harper & Row, New York.

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

Cornell, S.A., Lahiri, A., Eulitz, C., 2013. Inequality across consonantal contrasts in speech perception: evidence from mismatch negativity. J. Exp. Psychol. Hum. Percept. Perform. 39 (3), 757–772. https://doi.org/10.1037/a0030862.

Davidson, L., 2016. Variability in the implementation of voicing in American English obstruents. J. Phonet. https://doi.org/10.1016/j.wocn.2015.09.003.

Dien, J., 1998. Addressing misallocation of variance in principal components analysis of event-related potentials. Brain Topogr. 11 (1), 43–55. https://doi.org/10.1023/A:1022218503558.

Dien, J., 2010. The ERP PCA Toolkit: an open source program for advanced statistical analysis of event-related potential data. J. Neurosci. Methods. https://doi.org/10.1016/j.jneumeth.2009.12.009.

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

Dien, J., 2017. ERP PCA Toolkit. Retrieved from https://sourceforge.net/projects/

erppcatoolkit/.
Dien, J., Frishkoff, G.A., 2005. Principal components analysis of ERP data. In: Handy, T. (Ed.), Event-Related Potentials: A Methods Handbook. MIT Press, Cambridge.
Dien, J., Spencer, K.M., Donchin, E., 2003. Localization of the event-related potential novelty response as defined by principal components analysis. Cognit. Brain Res. 17 (3), 637–650. https://doi.org/10.1016/S0926-6410(03)00188-5.
Dien, J., Spencer, K.M., Donchin, E., 2004. Parsing the late positive complex: mental chronometry and the ERP components that inhabit the neighborhood of the P300. Psychophysiology 41 (5), 665–678. https://doi.org/10.1111/j.1469-8986.2004.00193.x.
Docherty, G.J., 1992. The Timing of Voicing in British English Obstruents. Foris, Berlin.
Eulitz, C., Lahiri, A., 2004. Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition. J. Cognit. Neurosci. 16 (4), 577–583. https://doi.org/10.1162/089892904323057308.
Fitzgerald, K., Provost, A., Todd, J., 2018. First-impression bias effects on mismatch negativity to auditory spatial deviants. Psychophysiology 55 (4), e13013. https://doi.org/10.1111/psyp.13013.
Frost, J.D., Haasnoot, K., McDonnell, K., Winkler, I., Todd, J., 2018. The cognitive resource and foreknowledge dependence of auditory perceptual inference. Neuropsychologia. https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2018.07.005.
Frost, J.D., Winkler, I., Provost, A., 2016. Surprising sequential effects on MMN. Biol. Psychol. 116, 47–56. https://doi.org/10.1016/J.BIOPSYCHO.2015.10.005.
Garrido, M.I., Kilner, J.M., Stephan, K.E., Friston, K.J., 2009. The mismatch negativity: a review of underlying mechanisms. Clin. Neurophysiol. 120 (3), 453–463. https://doi.org/10.1016/J.CLINPH.2008.11.029.
Gow, D.W., 2001. Assimilation and anticipation in continuous spoken word recognition. J. Mem. Lang. 45 (1), 133–159. https://doi.org/10.1006/jmla.2000.2764.
Haroush, K., Hochstein, S., Deouell, L.Y., 2010. Momentary fluctuations in allocation of attention: cross-modal effects of visual task load on auditory discrimination. J. Cognit. Neurosci. 22 (7), 1440–1451. https://doi.org/10.1162/jocn.2009.21284.
Harris, J., 1994. English Sound Structure. Blackwell, Oxford, UK; Cambridge, Mass.
Hayden, R.E., 1950. The relative frequency of phonemes in general-american english. WORD 6 (3), 217–223. https://doi.org/10.1080/00437956.1950.11659381.
Hestvik, A., Durvasula, K., 2016. Neurobiological evidence for voicing underspecification in English. Brain Lang. 152, 28–43. https://doi.org/10.1016/j.bandl.2015.10.007.
Honeybone, P., 2005. Diachronic evidence in segmental phonology: the case of obstruent laryngeal specifications. In: van Oostendorp, M., van de Weije, J. (Eds.), The internal organization of phonological segments. Mouton de Gruyter, Berlin, pp. 319–354.
Horváth, J., Czigler, I., Jacobsen, T., Maess, B., Schröger, E., Winkler, I., 2007. MMN or no MMN: No magnitude of deviance effect on the MMN amplitude. Psychophysiology. https://doi.org/10.1111/j.1469-8986.2007.00599.x. 070914092401003-000.
Horváth, J., Czigler, I., Jacobsen, T., Maess, B., Schröger, E., Winkler, I., 2008. MMN or no MMN: no magnitude of deviance effect on the MMN amplitude. Psychophysiology 45 (1), 60–69. https://doi.org/10.1111/j.1469-8986.2007.00599.x.
Ito, J., Mester, A., 1986. The phonology of voicing in Japanese: theoretical consequences for morphological accessibility. Ling. Inq. 17 (1), 49–73.
Iverson, G.K., Salmons, J.C., 1995. Aspiration and laryngeal representation in germanic. Phonology 12 (3), 369–396. https://doi.org/10.1017/S0952675700002566.
Jessen, M., Ringen, C., 2003. Laryngeal features in German. Phonology 19 (02), 189–218. https://doi.org/10.1017/S0952675702004311.
Kager, R., van der Feest, S., Fikkert, P., Kerkhoff, A., & Zamuner, T. S. (2007). 2. Representations of [voice]: Evidence from acquisition (pp. 41–80). doi.org/10.1075/cilt.286.03kag.
Kawahara, S. (2018). Phonology and orthography : The orthographic characterization of rendaku and Lyman ' s Law, 224–225.
Kessinger, R.H., Blumstein, S.E., 1997. Effects of speaking rate on voice-onset time in Thai, French, and English. J. Phonet. https://doi.org/10.1006/jpho.1996.0039.
Kielar, A., Joanisse, M.F., Hare, M.L., 2008. Priming English past tense verbs: rules or statistics? J. Mem. Lang. 58 (2), 327–346. https://doi.org/10.1016/j.jml.2007.10.002.
Klatt, D.H., 1975. Vowel lengthening is syntactically determined in a connected discourse. J. Phonet. 3 (3), 129–140.
Klatt, D.H., 1980. Software for a cascade/parallel formant synthesizer. J. Acoust. Soc. Am. 67, 971–995.
Kuroda, S. Y. (2002). Rendaku. In Noriko Akatsuka & S. Strauss (Eds.), Japanese/Korean Linguistics (pp. 337–350). Stanford: CSLI.
Lahiri, A., Reetz, H., 2002. Underspecified recognition. In Laboratory Phonology 7 (pp. 637–675).
Lahiri, A., Reetz, H., 2010. Distinctive features: Phonological underspecification in representation and processing. J. Phonet. 38 (1), 44–59. https://doi.org/10.1016/j.wocn.2010.01.002.
Lengeris, A., 2009. Perceptual assimilation and L2 learning: Evidence from the perception of Southern British english vowels by native speakers of Greek and Japanese. Phonetica 66 (3), 169–187. https://doi.org/10.1159/000235659.
Luck, S.J., Gaspelin, N., 2017. How to get statistically significant effects in any ERP experiment (and why you shouldn't). Psychophysiology 54 (1), 146–157. https://doi.org/10.1111/psyp.12639.
May, P.J.C., Tiitinen, H., 2010. Mismatch negativity (MMN), the deviance-elicited auditory deflection, explained. Psychophysiology 47 (1), 66–122. https://doi.org/10.1111/j.1469-8986.2009.00856.x.
McMurray, B., Aslin, R.N., Tanenhaus, M.K., Spivey, M.J., Subik, D., 2008. Gradient sensitivity to within-category variation in words and syllables. J. Exp. Psychol. Hum. Percept. Perform. 34 (6), 1609–1631. https://doi.org/10.1037/a0011747.
McMurray, B., Tanenhaus, M.K., Aslin, R.N., Spivey, M.J., 2003. Probabilistic constraint satisfaction at the lexical/phonetic interface: evidence for gradient effects of within-category VOT on lexical access. Retrieved from. J. Psychol. Res. 32 (1), 77–97. https://doi.org/10.1023/A:1021937116271.
Mester, A., Ito, J., 1989. Feature predictability and underspecification: palatal prosody in Japanese mimetics. Language 65 (2), 258–293.
Mines, M.A., Hanson, B.F., Shoup, J.E., 1978. Frequency of occurrence of phonemes in conversational english. Lang. Speech 21 (3), 221–241. https://doi.org/10.1177/002383097802100302.
Mitterer, H., 2011. The mental lexicon is fully specified: evidence from eye-tracking. J. Exp. Psychol. Hum. Percept. Perform. 37 (2), 496–513. https://doi.org/10.1037/a0020989.
Mitterer, H., Reinisch, E., McQueen, J.M., 2018. Allophones, not phonemes in spoken-word recognition. J. Mem. Lang. 98, 77–92. https://doi.org/10.1016/j.jml.2017.09.005.
Mullens, D., Winkler, I., Damaso, K., Heathcote, A., Whitson, L., Provost, A., Todd, J., 2016. Biased relevance filtering in the auditory system: a test of confidence-weighted first-impressions. Biol. Psychol. 115, 101–111. https://doi.org/10.1016/j.biopsycho.2016.01.018.
Mullens, D., Woodley, J., Whitson, L., Provost, A., Heathcote, A., Winkler, I., Todd, J., 2014. Altering the primacy bias-How does a prior task affect mismatch negativity? Psychophysiology 51 (5), 437–445. https://doi.org/10.1111/psyp.12190.
Näätänen, R., Alho, K., 1997. Higher-order processes in auditory-change detection. Retrieved from. Trends Cognit. Sciences 1 (2), 44–45. https://doi.org/10.1016/S1364-6613(97)01013-9.
Näätänen, R., Jacobsen, T.K., Winkler, I., 2005. Memory-based or afferent processes in mismatch negativity (MMN): a review of the evidence. Psychophysiology 42 (1), 25–32. https://doi.org/10.1111/j.1469-8986.2005.00256.x.
Näätänen, R., Paavilainen, P., Rinne, T., Alho, K., 2007. The mismatch negativity (MMN) in basic research of central auditory processing: a review. Clin. Neurophysiol. 118 (12), 2544–2590. https://doi.org/10.1016/j.clinph.2007.04.026.
Näätänen, R., Picton, T.W., 1987. The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. Psychophysiology 24 (4), 375–425. https://doi.org/10.1111/j.1469-8986.1987.tb00311.x.
Näätänen, R., Sams, M., Alho, K., Paavilainen, P., Reinikainen, K., Sokolov, E.N., 1988. Frequency and location specificify of the human vertex N1 wave. Electroencephalogr. Clin. Neurophysiol. 69 (6), 523–531. https://doi.org/10.1016/0013-4694(88)90164-2.
Nicholls, M.E.R., Thomas, N.A., Loetscher, T., Grimshaw, G.M., 2013. The flinders handedness survey (FLANDERS): a brief measure of skilled hand preference. Cortex 49 (10), 2914–2926. https://doi.org/10.1016/j.cortex.2013.02.002.
Phillips, C., Pellathy, T., Marantz, A., Yellin, E., Wexler, K., Poeppel, D., Roberts, T., 2000. Auditory cortex accesses phonological categories: an MEG mismatch study. J. Cognit. Neurosci. 12 (6), 1038–1055. https://doi.org/10.1162/08989290051137567.
Picton, T.W., Bentin, S., Berg, P., Donchin, E., Hillyard, S.A., Johnson, R., Taylor, M.J., 2000. Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria. Psychophysiology 37 (2), 127–152. https://doi.org/10.1111/1469-8986.3720127.
Pulvermüller, F., Shtyrov, Y., 2006. Language outside the focus of attention: the mismatch negativity as a tool for studying higher cognitive processes. Prog. Neurobiol. 79 (1), 49–71. https://doi.org/10.1016/J.PNEUROBIO.2006.04.004.
Pulvermüller, F., Shtyrov, Y., Ilmoniemi, R.J., Marslen-Wilson, W.D., 2006. Tracking speech comprehension in space and time. NeuroImage 31 (3), 1297–1305. https://doi.org/10.1016/J.NEUROIMAGE.2006.01.030.
Rhodes, R., Han, C., Hestvik, A., 2019. Phonological memory traces do not contain phonetic information. Attent. Percept. Psychophys. 1–15. https://doi.org/10.3758/s13414-019-01728-1.
Riney, T.J., Takagi, N., Ota, K., Uchida, Y., 2007. The intermediate degree of VOT in Japanese initial voiceless stops. J. Phonet. 35 (3), 439–443. https://doi.org/10.1016/J.WOCN.2006.01.002.
Scharinger, M., 2017. Are there brain bases for phonological markedness ? In: Samuels, B. (Ed.), Beyond Markedness in Formal Phonology. John Benjamins Publishing Company.
Scharinger, M., Bendixen, A., Trujillo-Barreto, N.J., Obleser, J., 2012. A sparse neural code for some speech sounds but not for others. PLoS ONE 7 (7), e40953. https://doi.org/10.1371/journal.pone.0040953.
Scharinger, M., Merickel, J., Riley, J., Idsardi, W.J., 2011. Neuromagnetic evidence for a featural distinction of English consonants: sensor- and source-space data. Brain Lang. 116 (2), 71–82. https://doi.org/10.1016/J.BANDL.2010.11.002.
Scharinger, M., Monahan, P.J., Idsardi, W.J., 2016. Linguistic category structure influences early auditory processing: converging evidence from mismatch responses and cortical oscillations. NeuroImage 128, 293–301. https://doi.org/10.1016/j.neuroimage.2016.01.003.
Schluter, K., Politzer-Ahles, S., Al-Kaabi, M., Almeida, D., 2017. Laryngeal features are phonetically abstract. Front. Psychol. https://doi.org/10.3389/fpsyg.2017.00746.
Shimizu, K., 1977. Voicing features in the perception and production of stop consonants by Japanese speakers. Stud. Phonol. 11, 25–34.
Spencer, K.M., Dien, J., Donchin, E., 1999. A componential analysis of the ERP elicited by novel events using a dense electrode array. Retrieved from. Psychophysiology 36 (3), 409–414. https://doi.org/10.1017/S0048577299981180.
Spencer, K.M., Dien, J., Donchin, E., 2001. Spatiotemporal analysis of the late ERP responses to deviant stimuli. Psychophysiology 38 (2), 343–358. https://doi.org/10.1017/S0048577201000324.
Steinschneider, M., Liégeois-Chauvel, C., Brugge, J.F., 2011. Auditory evoked potentials and their utility in the assessment of complex sound processing. In: Winer, C., Schreiner, J.A. (Eds.), The Auditory Cortex. Springer US, pp. 535–559. https://doi.org/10.1007/978-1-4419-0074-6_25.
Steinschneider, M., Schroeder, C.E., Arezzo, J.C., Vaughan, H.G., 1995. Physiologic

correlates of the voice onset time boundary in primary auditory cortex (A1) of the awake monkey: temporal response patterns. Brain Lang. 48 (3), 326–340. https://doi.org/10.1006/brln.1995.1015.
Strange, W., Akahane-Yamada, R., Kubo, R., Trent, S.A., Nishi, K., Jenkins, J.J., 1998. Perceptual assimilation of American English vowels by Japanese listeners. J. Phonet. 26, 311–344. https://doi.org/10.1006/jpho.1998.0078.
Sussman, E.S., Chen, S., Sussman-Fort, J., Dinces, E., 2014. The five myths of MMN: redefining how to use MMN in basic and clinical research. Brain Topogr. 27 (4), 553. https://doi.org/10.1007/S10548-013-0326-6.
Takada, M., 2011. Nihongo no Goto-heisaon no Kenkyu: VOT no Kyoziteki Bunpu to Tuziteki Henka (Research on the Word-initial Stops of Japanese: Synchronic Distribution and Diachronic Change in VOT). Kuroshio, Tokyo.
Takada, M. (2012). Regional and generational variation in Japanese word-initial stops. In Papers from the First International Conference on Asian Geolinguistics (pp. 273–282).
Tamaoka, K., Makioka, S., 2004. Frequency of occurrence for units of phonemes, morae, and syllables appearing in a lexical corpus of a Japanese newspaper. Behav. Res. Methods Instrum. Comput. 36 (3), 531–547. https://doi.org/10.3758/BF03195600.
Tamaoka, K., Makioka, S., 2009. Japanese mental syllabary and effects of mora, syllable, bi-mora and word frequencies on japanese speech production. Lang. Speech 52 (1), 79–112. https://doi.org/10.1177/0023830908099884.
Tavabi, K., Elling, L., Dobel, C., Pantev, C., Zwitserlood, P., 2009. Effects of place of articulation changes on auditory neural activity: a magnetoencephalography study. PLoS ONE 4 (2), e4452. https://doi.org/10.1371/journal.pone.0004452.
Todd, J., Heathcote, A., Whitson, L.R., Mullens, D., Provost, A., Winkler, I., 2014. Mismatch negativity (MMN) to pitch change is susceptible to order-dependent bias. Front. Neurosci. 8, 180. https://doi.org/10.3389/fnins.2014.00180.
Todd, J., Michie, P.T., Schall, U., Karayanidis, F., Yabe, H., Näätänen, R., 2008. Deviant matters: duration, frequency, and intensity deviants reveal different patterns of mismatch negativity reduction in early and late schizophrenia. Biol. Psychiatry. https://doi.org/10.1016/j.biopsych.2007.02.016.
Todd, J., Provost, A., Cooper, G., 2011. Lasting first impressions: A conservative bias in automatic filters of the acoustic environment. Neuropsychologia. https://doi.org/10.1016/j.neuropsychologia.2011.08.016.
Todd, J., Provost, A., Whitson, L., Mullens, D., 2016. Initial uncertainty impacts statistical learning in sound sequence processing. J. Physiol. Paris 110 (4), 497–507. https://doi.org/10.1016/j.jphysparis.2017.01.001.
Todd, J., Provost, A., Whitson, L.R., Cooper, G., Heathcote, A., 2013. Not so primitive: context-sensitive meta-learning about unattended sound sequences. J. Neurophysiol. 109 (1), 99–105. https://doi.org/10.1152/jn.00581.2012.
Toscano, J.C., McMurray, B., Dennhardt, J., Luck, S.J., 2010. Continuous perception and graded categorization: electrophysiological evidence for a linear relationship between the acoustic signal and perceptual encoding of speech. Psychol. Sci. 21 (10), 1532–1540. https://doi.org/10.1177/0956797610384142.
Trubetskoy, N.S., 1969. Principles of Phonology. (2 Brook St., W1Y 1AA). University of California Press, Berkeley London.
Tsuchida, A., 1997. The Phonetics and Phonology of Japanese Vowel Devoicing. Cornell University.
Tsuchida, A., 2001. Japanese vowel devoicing: cases of consecutive devoicing environments. J. East Asian Ling. 10 (3), 225–245. https://doi.org/10.1023/A:1011221225072.
Weide, R.L., 1994. CMU Pronouncing Dictionary.