



Listening to different speakers: On the time-course of perceptual compensation for vocal-tract characteristics

Matthias J. Sjerps^{a,*}, Holger Mitterer^a, James M. McQueen^{a,b}

^a Max Planck Institute for Psycholinguistics, P.O. Box 310, 6500 AH Nijmegen, The Netherlands

^b Behavioural Science Institute and Donders Institute for Brain, Cognition and Behaviour, Centre for Cognition, Radboud University Nijmegen, P.O. Box 9104, 6500 HE Nijmegen, The Netherlands

ARTICLE INFO

Article history:

Received 4 February 2011

Received in revised form

22 September 2011

Accepted 27 September 2011

Available online 5 October 2011

Keywords:

Event-related potential (ERP)

Speech perception

Vowel normalization

N1

ABSTRACT

This study used an active multiple-deviant oddball design to investigate the time-course of normalization processes that help listeners deal with between-speaker variability. Electroencephalograms were recorded while Dutch listeners heard sequences of non-words (standards and occasional deviants). Deviants were [ɪpapu] or [ɛpapu], and the standard was [ɪ_εpapu], where [ɪ_ε] was a vowel that was ambiguous between [ɛ] and [ɪ]. These sequences were presented in two conditions, which differed with respect to the vocal-tract characteristics (i.e., the average 1st formant frequency) of the [papu] part, but not of the initial vowels [ɪ], [ɛ] or [ɪ_ε] (these vowels were thus identical across conditions). Listeners more often detected a shift from [ɪ_εpapu] to [ɛpapu] than from [ɪ_εpapu] to [ɪpapu] in the high F₁ context condition; the reverse was true in the low F₁ context condition. This shows that listeners' perception of vowels differs depending on the speaker's vocal-tract characteristics, as revealed in the speech surrounding those vowels. Cortical electrophysiological responses reflected this normalization process as early as about 120 ms after vowel onset, which suggests that shifts in perception precede influences due to conscious biases or decision strategies. Listeners' abilities to normalize for speaker-vocal-tract properties are for an important part the result of a process that influences representations of speech sounds early in the speech processing stream.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

In everyday life, we listen to the speech of many individuals. The current paper investigates a perceptual compensation process that helps listeners to understand speech sounds spoken by different talkers. Individuals have different vocal-tract characteristics, caused by influences such as talker sex, talker size (or vocal-tract length), speaking style, and dialect. This variance appears to challenge speech comprehension because vocal tracts can differ on the same acoustic dimensions that allow listeners to discriminate between different speech-sound categories. We ask here how early in the speech perception process listener's representations of speech sounds are changed in order to compensate for talkers' vocal-tract characteristics.

Vowels are discriminated mainly on the basis of acoustic properties that are referred to as formants. Formants are bands of increased intensity in the spectral makeup of speech sounds. For example, in English the main difference between the words "bit" and "bet" (phonemically transcribed as /bɪt/ versus /bɛt/) lies in

the frequency of the first formant (F₁). The average F₁ value for /ɛ/ is around 731 Hz while the average F₁ of /ɪ/ lies around 483 Hz, for vowels recorded from female American English speakers (F₂ value for /ɛ/: 2058; F₂ of /ɪ/: 2365). For male speakers the average F₁ value for /ɛ/ is around 580 Hz while the average F₁ of /ɪ/ lies around 427 Hz (F₂ value for /ɛ/: 1799; F₂ of /ɪ/: 2034) (Hillenbrand, Getty, Clark, & Wheeler, 1995). However, these averages do not tell the complete story. There is a large degree of overlap among different vowel categories (Hillenbrand et al., 1995; Joos, 1948). Single instances of two different vowels, spoken by two different speakers, can have very similar absolute formant values. This is not restricted to English; Dutch, the target language of the current paper, shows similar overlap in vowel categories (Adank, van Hout, & Smits, 2004; Van Nierop, Pols, & Plomp, 1973). This is especially the case when comparing speakers of different sex or age. Such variance therefore causes multiple signal-to-category mappings for a single spoken speech sound. In other words, a single sound can often be interpreted as either of two different phonemes, so that listeners may be confused whether the intended word was *bit* or *bet*.

It has been argued that listeners compensate for vocal-tract characteristics in a number of different ways (Johnson, 2005; Nearey, 1989). An important contribution may be made by a mechanism that compensates for speaker characteristics by taking into

* Corresponding author. Tel.: +31 243521334.

E-mail address: m.j.sjerps@gmail.com (M.J. Sjerps).

account the vocal-tract characteristics of the speaker as revealed in a preceding context (Ladefoged & Broadbent, 1957). Ladefoged and Broadbent (1957) found that listeners interpret a vowel that is acoustically halfway between an [ɪ] and an [ɛ] more often as /ɪ/ (which has a low F_1) when it is preceded by a sentence with a relatively high F_1 , while the same sound is more often interpreted as /ɛ/ (which has a high F_1) when preceded by a precursor sentence with a relatively low F_1 . This contrastive process therefore effectively normalizes perception for the F_1 range of a speaker and reduces potential overlap of vowel categories across speakers. Although F_1 is not the only cue to differences in vocal-tract characteristics, Ladefoged and Broadbent (1957) have shown that listeners can use F_1 characteristics to map speech sounds onto the correct phonemes.

Watkins (1991) and Watkins and Makin (1994, 1996) have argued that the bulk of this effect can be explained by a mechanism that compensates for the average spectral makeup of a precursor, whether the shape of this spectral makeup is due to vocal-tract characteristics or something else (such as room acoustics). The suggestion is that these context effects are the result of a mechanism that focuses on how a given stimulus is different from the preceding context (“Sensitivity to change”, cf. Kluender, Coady, & Kieft, 2003). This mechanism is assumed to be a general perceptual mechanism that is not specific to speech perception. In line with this claim, contrast effects similar to vowel normalization also occur for musical timbre perception (Stilp, Alexander, Kieft, & Kluender, 2010). Additional evidence stems from the finding that a precursor spoken by a female talker can influence the perception of a subsequent target sound that was produced by a male talker (Watkins, 1991), and that speech sound categorization can also be influenced by non-speech precursors (Holt, 2006). This kind of mechanism could therefore function as a means of enhancing contrast (Kluender et al., 2003; Kluender & Kieft, 2006) and displays a clear analogy to contrast effects with visual stimuli. A surface with a certain brightness will be perceived as darker when surrounded by a light surface, but as lighter when surrounded by a dark surface. Furthermore, effects of preceding context on speech sound categorization have also been observed with Japanese quail (Lotto, Kluender, & Holt, 1997). This again suggests that influences of context are in part the result of a relatively general perceptual mechanism (see Holt, 2006, for an overview of context dependent effects on categorization), but it is not clear that a single general-purpose mechanism is sufficient to explain all the results on vowel normalization. Sjerps, Mitterer, and McQueen (2011) argue, for example, that vowel normalization may primarily reflect a compensation mechanism that is based on the Long-Term Average Spectrum (LTAS) of the auditory input, but one that only operates if the input has spectro-temporal characteristics that are similar to those of speech. The primary compensation mechanism for vowel normalization thus appears to be general-purpose (based on contrast), but one which operates under some spectro-temporal constraints.

Little is known, however, about when in the processing stream this compensation mechanism has its influence. Does normalization influence low-level representations or does it influence higher-level cognitive processes? Clearly, the assumption of a general perceptual mechanism that focuses on change is more compatible with the assumption of an early locus. The present study therefore examines the temporal locus of compensation for speaker vocal-tract characteristics by tracking its neurophysiological correlates during the perception of vowels. This is a novel approach in the investigation of the extrinsic normalization of vowels.

In order to establish whether normalization influences representations early or late in the stream of processing, four different time windows were investigated that can be considered to reflect subsequent stages in the processing stream. These were the P1, the N1, the N2 and the P3 time windows. Previous neurophysiological

investigations with speech stimuli have suggested different functional interpretations for the processes/representations underlying these different waveform components. The earliest long-latency brain waves (P1 and N1 or their magnetic counterparts P1m, N1m) peaking at about 50 and 100 ms after stimulus onset respectively, seem to reflect early cortical information processing (Diesch, Eulitz, Hampson, & Ross, 1996; Makela, Alku, & Tiitinen, 2003; Obleser, Eulitz, & Lahiri, 2004). While P1 has been argued to reflect basic auditory feature extraction, N1 seems to reflect a subsequent level closer to a more abstract phonological representational stage (Tavabi, Obleser, Dobel, & Pantev, 2007). Roberts, Flagg, and Gage (2004) argue that the N1 response represents some form of abstract processing. Cortical responses that were recorded when participants listened to sounds on a vowel continuum from [u] to [a] reflected clustering of N1 peak latencies around the regions of the continuum identified as either [u] or [a] (and less clustering around the ambiguous region). Roberts et al. (2004) also find, however, that when acoustic aspects of a single stimulus are held constant while the percept is changed through a response bias induced by preceding trials, the dominant N1 latency effect is related to the physical properties of the stimulus, and not to the eventual decision. This indicates that higher-level processes like response biases do not influence N1. Furthermore, while the N1 might reflect some abstract properties, Näätänen and Winkler (1999) argue that it does not reflect a completely abstract level of processing as it does not directly reflect the consciously perceived event. Bien, Lagemann, Dobel, and Zwitserlood (2009) have also shown that there is a difference between the N1 response and conscious decisions about the signal. Furthermore, in contrast to Roberts et al. (2004), Toscano, McMurray, Dennhardt, and Luck (2010) have shown that, for the perception of the voiced versus voiceless stop consonant distinction (as in beach versus peach), there is no relation between the N1 amplitude and the categorical status of a phoneme. They found that there was a linear relation between the N1 amplitude and the step on the voiced-voiceless continuum. Thus, while it is not yet clear whether the N1 can reflect abstract aspects of speech sounds, previous results do show that the N1 reflects processes that are not influenced by response-bias or by the consciously perceived qualities of the stimulus. This shows that these processes take place early in the processing of speech information. Roberts et al. (2004) thus argue that the ultimate perception of speech sounds depends on the coding of stimulus properties that takes place during the N1 time window.

Later time windows such as the N2/MMN time window (200–300 ms after stimulus onset) do seem to reflect abstract levels of processing (Näätänen et al., 1997; Winkler et al., 1999). MMN responses, for example, are larger to deviants that are linguistically relevant for the listener (Näätänen et al., 1997; Winkler et al., 1999). Moreover, in a study measuring both N1 and MMN, Sharma and Dorman (2000) found a dissociation between measures of N1 and MMN. Both Hindi and English listeners showed similar, direct dependencies of N1 latency on a Voice Onset Time (VOT) continuum that is only relevant for listeners of Hindi (−90 to 0 ms). However, only Hindi listeners elicited a MMN effect with these stimuli. This shows that N1 and MMN reflect subsequent stages in the processing hierarchy and only the MMN response is dependent on linguistic exposure. Finally, the P3 response in response-active odd-ball designs (300–600 ms after stimulus onset) has been associated with the evaluation of deviant events with relation to subsequent behavioural action (Friedman, Cycowicz, & Gaeta, 2001). The P3 is thus likely to also reflect higher-level cognitive processes, although it is not necessarily insensitive to gradedness within speech categories (Toscano et al., 2010).

Our aim here was to investigate whether compensation for speaker vocal-tract characteristics is a process that influences representations of speech sounds at a relatively early stage of

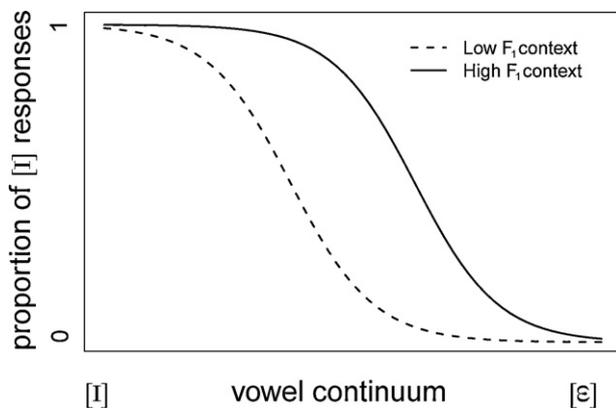


Fig. 1. Schematized response functions: mean proportions of an [ɪ] response to a range of target sounds from [ɪ] to [ɛ], presented after a precursor with an increased F_1 (high F_1 context) or decreased F_1 (low F_1 context) level.

processing (i.e., during P1 and/or N1 time windows) or at a relatively late stage of processing (i.e., during N2 and/or P3 time windows). We investigated the influence of vocal-tract characteristics on vowel perception by presenting participants with target vowels in contexts that simulate speakers with different vocal-tract characteristics. Previous findings have shown that manipulated context sentences can change the perception of subsequently presented vowels, indicated by a shift in the categorization functions for these vowels (Kiefte & Kluender, 2008; Ladefoged & Broadbent, 1957; Mitterer, 2006; Sjerps et al., 2011; Watkins, 1991; Watkins & Makin, 1994, 1996).

Consider Fig. 1, it displays two mock-up categorization functions for sounds from a vowel continuum ranging from [ɪ] to [ɛ], that represent the sort of shift in categorization function that has been found. The dotted line represents the categorization of vowels that have been presented after a precursor with a generally low F_1 , while the solid line represents categorization of the same vowel tokens, but then presented after a precursor sentence with a high F_1 . It can be observed that more sounds on the continuum are categorized as /ɪ/ (which itself has a low F_1) in the context of a precursor with a high F_1 and more often as /ɛ/ (which has a high F_1) in the context of a low F_1 precursor. The current research attempts to investigate at what level of processing the representation of speech sounds is influenced by the mechanism that leads to this shift in perception.

In the present experiment, target non-words were presented in a response-active mismatch detection design, such that listeners heard a repeating (standard) non-word that was replaced by two different (deviant) non-words on 20% of trials. The standard consisted of a non-word in which the initial vowel was manipulated to sound halfway between [ɪ] and [ɛ], from now on indicated by [ɪ_ε] (the transcription of the ambiguous sound as [ɪ_ε] should make clear that this sound does not represent an actual Dutch phoneme category). The deviant non-words started with a vowel that was an unambiguous instance of /ɪ/ or /ɛ/. The following two syllables in each non-word (/papu/) were manipulated to have a high F_1 or a low F_1 so as to induce normalization effects in different experimental blocks. The bisyllable /papu/ contains two point vowels that provide the range of a speaker's F_1 . The induced change in perception through normalization should make it harder for participants to detect a change from the ambiguous vowel [ɪ_ε] to [ɪ] than to [ɛ] in the high F_1 context, whereas listeners should find it harder to detect a change from [ɪ_ε] to [ɛ] than to [ɪ] in the low F_1 context.

Listeners thus heard the nonsense words [ɪ_εpapu] (as the standard stimulus), and [ɪpapu] and [ɛpapu] (as the deviant stimuli). In this setup the 2nd and 3rd syllables of stimulus x provided the preceding context for the next stimulus, $x+1$. This approach was chosen to be able to create an interstimulus interval (ISI) between the

context ([papu]) and the subsequent target-vowels of 750 ms (i.e., [ɪ_εpapu]–750 ms–[ɛpapu]–750 ms–[ɪ_εpapu], etc.). The contextual influence of the [papu] syllables might extend to subsequent trials (i.e., the perception of a target vowel is influenced not only by the just-preceding context), but because of the blocked presentation this influence was in the same direction within a block. The large ISI between a target vowel and its immediately preceding context is important because small ISI could lead to contextual influences that are a result of peripheral auditory influences such as the negative auditory after-image (Summerfield, Haggard, Foster, & Gray, 1984; Watkins, 1991; Wilson, 1970). Such peripheral influences can cause a compensation effect in the same direction as the more central compensation effect under investigation here and could thus obscure its effects. The contexts thus followed directly after the target vowels. This is not a problem for the interpretation of the EEG waveform, however, as the following context started 250 ms after the onset of the vowel, which leaves enough time for any early cortical signatures in response to the critical vowels to appear before any effect of following context (at least those in the P1, N1, and N2 time windows). The early components induced by the following [papu] coincide with the P3 effect induced by the target vowel. This is not a problem, however, as the P3 response is larger in amplitude than the earlier cortical responses that could be induced by the following context. It should be noted that the strength of normalization effects might decrease over repetitions (Broadbent & Ladefoged, 1960). This decrease, however, has been argued to be stronger when different context conditions are presented in a mixed fashion instead of the blocked approach that was taken here (Sjerps et al., 2011).

An additional control condition was run that had the vowel [ɔ] as the initial vowel on the standard items (i.e., [ɔpapu]), but had the same deviants as in the experimental condition ([ɪpapu] and [ɛpapu]). In this control condition the [papu] part had a neutral F_1 contour that was halfway between that of the high and low F_1 conditions. This control condition was used to test whether our design was capable of producing a clear standard-deviant mismatch effect in the cortical signatures, and when and where on the cortical topography these mismatch effects would express themselves.

The control data were analyzed by comparing the size and distribution of the effect of deviant ([ɪpapu] and [ɛpapu]) versus standard ([ɔpapu]) in the four time windows. For the experimental (i.e., non-control) stimuli, an initial analysis compared ERPs between the two standard stimuli ([ɪ_εpapu] in both the high F_1 and the low F_1 condition) versus the deviants ([ɪpapu] and [ɛpapu] in both the high F_1 and the low F_1 condition). This comparison was made to see whether and when the small auditory differences that we used in the experimental condition were able to elicit different cortical responses to deviants (note that in both sets of data the deviant vowels were the same, only the standards differed). In the final and critical analysis we tested at what point in the stages of cortical processing of speech the influence of the contexts' F_1 -properties on the detectability of a vowel change was reflected. This effect was tested by looking for an interaction between the F_1 condition and the identity of the deviant vowel, with the size of the difference response (in voltage) as the dependent variable. Note that the analysis of this critical interaction focuses on the processing of the deviants and not on the processing of the standard, despite the fact that our design hinges on the fact that the perception of the standard is changed across blocks. Normalization processes change the perceived quality of the standard and thus also the mental traces of the standard. For the critical analysis, we measured the relative strength of the cortical signature of the mental comparison of a deviant vowel to those traces of the standard. Traditionally, designs with this odd-ball paradigm focus on the difference wave between standard and deviant (cf. Näätänen & Winkler, 1999). As we were interested primarily in the interaction between deviant identity and the contexts'

F_1 properties, a comparison of the deviants themselves suffices here.

In the present study, an early influence should thus be reflected in early time windows (i.e., within the first 160 ms after vowel onset) such as those related to the P1 and/or the N1, whereas a later influence should only be able to affect cortical signatures later than about 200 ms, a time window which is related to the N2/MMN or the P3. To exemplify the expected results, imagine the analysis in the P3 time window. We expected that easy detectability of deviants would lead to a stronger positivity. The [i] deviant should be easier to detect (and thus result in a larger positivity) in the low F_1 condition than in the high F_1 condition. The difference wave for “[i] in a low F_1 context”– “[i] in a high F_1 context” should thus be positive. For [ɛ], this pattern should be reversed, and the difference wave for “[ɛ] in a low F_1 context”– “[ɛ] in a high F_1 context” should be negative. This mirror-image pattern of results should not necessarily arise only in the P3 time window; in fact, it should be observed from the point in time where the normalization processes start to take place. The question we asked was when that would be.

2. Methods

2.1. Participants

Twenty-four native speakers of Dutch from the Max Planck Institute for Psycholinguistics participant pool were tested. They received a monetary reward for their participation. None of the participants reported a hearing disorder, language impairment, or uncorrected visual impairment and all participants were right-handed.

2.2. Materials

All recordings were made by a female native speaker of Dutch. Acoustic processing of the stimuli was carried out using PRAAT software (Boersma & Weenink, 2005). The materials consisted of the three-syllable nonsense sequences “ipapu” ([ɪpapu]) and “ɛpapu” ([ɛpapu]), which are meaningless in Dutch. To create a continuum ranging from [ɪpapu] to [ɛpapu], the first vowel of a recording of [ɛpapu] was excised. This token had an F_1 frequency of 734 Hz, measured over a 40 ms window at a steady part in the middle of the vowel. A continuum was created by using a Linear Predictive Coding (LPC) procedure to generate a source and filter model of the vowel. The frequency of F_1 was decreased in the filter model, which was then recombined with the source model. Three target steps were selected ranging from [ɛ] to [i]. F_1 values ranged from 724 Hz ([ɛpapu]), through 634 Hz (the ambiguous item [ɪ_ɛpapu]), to 544 Hz ([ɪpapu]). The manipulated vowels were then low pass filtered at 1000 Hz and combined with the higher frequencies of the original vowel (1000–5000 Hz), to make them as naturally sounding as possible while retaining the perceived /ɛ/ versus /i/ identity for the two endpoints. The created vowels were adjusted so that their amplitude envelope and overall amplitude matched that of the original sound. The top left panel of Fig. 2 displays the spectra of the three vowels. The bottom left panel displays the difference between the LTAS of the endpoints [ɛ] and [i]. It shows that [ɛ] has considerably more energy in the region between ~500 and 1000 Hz than [i].

To create the [ɔpapu] item for the control condition, an instance of [ɔ] was selected from a similar segmental context (an initial /s/ followed by a /p/) and spoken by the same speaker. This sound was manipulated to have the same pitch, the same amplitude envelope and the same duration as the critical ambiguous vowel items.

The original recorded [papu] context had F_1 values of 730 Hz ([a]) and 410 Hz ([u]) measured over a 40 ms window at a steady part in the middle of the vowels. These two syllables were manipulated by the same F_1 manipulation procedure but then with either an increase of F_1 by 200 Hz, no increase, or a decrease of F_1 by 200 Hz to create the high F_1 context, the neutral context, and the low F_1 context respectively. The F_1 manipulation had some effect on the perceptual characteristics of the vowels in the [papu] sequence. Nevertheless, we will transcribe this sequence throughout as [papu], since this reflects the original utterance. The top right panel of Fig. 2 displays the LTAS of the three contexts. The bottom right panel shows the difference between the LTAS of the high F_1 and the low F_1 contexts. The high F_1 context has more energy than the low F_1 context in the region between ~500 and 1000 Hz. This pattern is similar to that of the target vowels. A normalization mechanism that operates through compensation for LTAS could thus invoke a perceptual shift with these stimuli.

The context bisyllables were recombined with the different steps of the manipulated onset vowels to create the experimental items. This resulted in the following items: for the experimental high F_1 and low F_1 conditions, instances of [ɪpapu] (deviant), [ɪ_ɛpapu] (standard) and [ɛpapu] (deviant); for the control condition, with the neutral context, instances of [ɪpapu] (deviant),

[ɔpapu] (standard) and [ɛpapu] (deviant). The stimuli can be found online at <http://www.holgermitterer.eu/SjMcQueenMit.html>.

2.3. Procedure

Deviant trials were presented randomly mixed between the standards, and presented in the same block (i.e., this was a multi-deviant design). The ratio of presentation was standard 80% and deviants 20% (each deviant 10%). When a deviant occurred it was always followed by at least two standards. The three conditions (high F_1 , low F_1 and neutral control conditions) were presented in separate blocks, each one once in the first half of the experiment and once in the second half of the experiment (i.e., each condition was presented twice). Every block consisted of 400 trials (320 standards, 80 deviants). This resulted in a total of 2400 trials. Presentation order of the different blocks was balanced over participants using a Latin-square design. Participants were instructed to press a button with their right hand whenever they heard a deviant trial. They were instructed that there could be different deviants, and that the difference would always be on the first vowel of the non-words. Participants were also told that a deviant trial would not necessarily sound like a shift in vowel category, but could also be the same vowel that was just pronounced somewhat differently.

Participants were tested individually in a single session in a soundproof, electrically shielded room. They were seated in a comfortable chair at a distance of approximately 60 cm from a computer screen and instructed to relax and avoiding excessive blinking and movements. The instructions were presented on the screen in written form. After each block, participants were allowed to take a break for as long as they wanted. The session included half an hour of electrode application and instruction and one hour stimulus presentation during which the EEG data was recorded.

Button-press responses were measured and analyzed to investigate whether the effect of context resulted in behaviorally measurable differences in detection ability between the different vowels.

2.4. EEG recording and analysis

The EEG was recorded from 36 active Ag/AgCl electrodes, of which 32 were mounted in a cap (actiCap), referenced to a nose electrode. Recording and analyses were carried out with Brain Vision Analyzer (version 1.05.0005). Two separate electrodes were placed at the left and right mastoids. Blinks were monitored through an electrode on the infraorbital ridge below the left eye. Horizontal eye movements were monitored through two electrodes in the cap (LEOG and REOG), placed approximately at each outer canthus. The ground electrode was placed on the forehead. Electrode impedance was kept below 20 k Ω , which is a sufficiently low impedance when using active electrodes. EEG and EOG recordings were amplified through BrainAmp DC amplifiers using a bandpass filter of 0.016–100 Hz, digitized on-line with a sampling frequency of 500 Hz, and stored for off-line analysis.

Bipolar vertical EOG was computed as the difference between the electrode on the infraorbital ridge of the left eye and the electrode situated right above the left eye. Bipolar horizontal EOG was computed as the difference between the LEOG and REOG electrodes. Data was corrected for the electrooculogram using the Gratton and Coles method in Brain Vision Analyser. EEG was band filtered between 1 and 30 Hz (24 dB/oct). The signals were segmented into epochs of 900 ms (with a 100 ms pre-stimulus baseline, time-locked to the vowel onset). Epochs with an amplitude outside the range of -70 to $70 \mu\text{V}$ were automatically excluded (on average 93.8% of the epochs were kept). Average ERPs were then computed across trials per participant for each type of initial vowel in each type of context.

Wave amplitudes were measured and analyzed for four time windows comprising the time windows P1: 30–80 ms (Tavabi, Elling, Dobel, Pantev, & Zwitserlood, 2009), N1: 80–160 ms (Cacace & McFarland, 2003), N2: 200–300 ms (Celsis et al., 1999; Saarinen, Paavilainen, Schoger, Tervaniemi, & Näätänen, 1992) and P3: 300–600 ms (Snyder, Hillyard, & Galambos, 1980). Analyses were conducted on a subset of 15 electrodes (F3, F4, F7, F8, C3, C4, T7, T8, P3, P4, P7, P8, Fz, Cz, Pz). Analyses were conducted separately for the midline (Fz, Cz, Pz) and the lateral (F3, F4, F7, F8, C3, C4, T7, T8, P3, P4, P7, P8) electrodes (see below for how these electrodes were grouped for the analyses).

A first set of analyses was conducted to determine whether an effect of deviant was visible in the ERP. These analyses compared the cortical signatures of the control condition with the neutral context for the standard ([ɔpapu]) versus the average of the two deviants ([ɪpapu] and [ɛpapu]).

A second set of analyses was conducted to compare the average of the two standard stimuli ([ɪ_ɛpapu] in both the high F_1 and the low F_1 condition) versus the averaged ERPs for the deviants ([ɪpapu] and [ɛpapu] in both the high F_1 and the low F_1 condition).

The third set of analyses comprised comparisons between the two deviant vowels in the two critical conditions ([ɪpapu] and [ɛpapu] in the high F_1 context condition and [ɪpapu] and [ɛpapu] in the low F_1 context condition). An effect of context on the detectability (reflected in amplitude of the cortical signatures) of the deviant vowels should result in an interaction between the factors Vowel ([i] versus [ɛ]) and Context (high F_1 versus low F_1).

Analyses on the electrophysiological data were run using SPSS software. Trials were included in the analysis irrespective of whether they were detected as a deviant

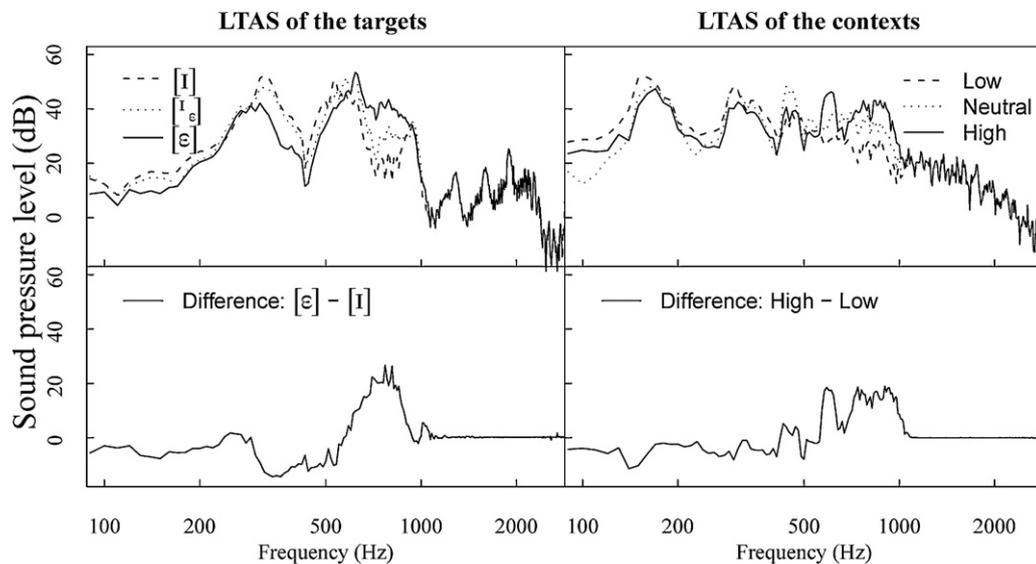


Fig. 2. Long-Term Average Spectra (LTAS) plots of the stimuli. Upper left panel: LTAS for the deviant target vowels [I] (dashed line) and [ε] (solid line), and the ambiguous standard vowel (dotted line). Left bottom panel: the difference spectrum for the deviant target vowels ([ε]–[I]). Upper right panel: LTAS for the [papu] contexts that were manipulated to have a low (dashed line), a neutral (dotted line), or a high F_1 (solid line). Right bottom panel: the difference spectrum for the [papu] contexts (high–low). The x-axis is displayed on a logarithmic scale because this scale more accurately reflects the response properties of the auditory system.

or not. We did not differentiate between these so as to maximize statistical power and to keep the number of contributing trials similar across conditions (note in particular that in the critical comparisons we predicted differences across F_1 conditions in the number of deviants that would be detected). The focus of this research was on the time-course of normalization, and thus not on potential differences between cortical responses to detected versus non-detected deviants, although this question could be the focus of future research. Following the convention for the data analysis of electrophysiological recordings, the Greenhouse–Geisser correction was applied. The analyses investigated the size and location of the effects of deviants in the control and experimental conditions. The lateral analysis (all but the three midline electrodes) included the factors AntPost (with the three levels Anterior, Medial, Posterior), Hemisphere (with the two levels Left and Right), MedLat (with the two levels Medial and Lateral) and Deviant (with the two levels Deviant and Standard). The analysis investigating the size and location of the effect of context on the perception of the vowels (the normalization effect) did not include the factor Deviant but instead included the factors Vowel (with two levels [I] and [ε]) and Context (with two levels high F_1 and low F_1). Appendices report the outcomes of the analyses. Effects are only reported if they include the factor Deviant or the interaction between Context and Vowel (for the final, critical, analysis). If interactions were found they were broken up in their constituent (sets of) electrodes. Only highest level interactions were broken down in this way.

3. Results

3.1. Behavioural data

In the control condition participants detected 98.4% of the deviants. In the experimental conditions participants detected on average 52.5% of the deviants. The behavioural data were analyzed using linear mixed-effects models in R (version 2.6.2, R development core team, 2008, with the lmer function from the lme4 package of Bates & Sarkar, 2007). Detection responses were modelled using the logit-linking function (Dixon, 2008). Hits were coded as 1 and misses as 0. Different models were tested in a deductive way, starting from a complete model including the factors Context (levels –1 (low F_1 context) and 1 (high F_1 context)), Vowel (levels –1 (initial [I]) and 1 (initial [ε])) and the interaction between Context and Vowel. All factors were centered around zero to make the interpretation of effects more straightforward (Barr, 2008). Non-significant predictors were taken out of the analysis in a stepwise fashion, starting from the highest order interaction, until no predictors could be removed without significant loss of fit. Modelling settled on a main effect for the factor Vowel ($b=0.080$, $p=0.002$) which indicates that [εpapu] deviants were more easily detected

than [Ipapu] deviants (54.1% versus 50.9% respectively). An interaction was found between Vowel and Context ($b=-0.586$, $p<0.001$) which reflects the fact that in the high F_1 condition [εpapu] deviants were more easily detected than [Ipapu] deviants ([εpapu] = 65.7%, [Ipapu] = 37.9%) whereas this effect was in the opposite direction for the low F_1 condition ([εpapu] = 42.6%, [Ipapu] = 63.9%). The critical normalization effect was thus observed in the behavioural responses.

3.2. EEG data: control condition

Fig. 3A displays the grand averages of the standards (dashed) and the deviants (thick solid line) for all 15 electrodes that were included in the analysis, along with the difference waves (thin solid line). Fig. 3B displays the average scalp distribution of the difference wave for the 4 different time windows that are analyzed here (a larger set of 28 electrodes were included for the creation of these maps). In general the data show a clear effect of deviant detection, expressed over a large part of the analysis window. Appendix A displays the effects that include the factor Deviant, along with the broken-down constituents for the highest order interactions that include this factor. The four specific time windows will be discussed separately, first describing the analysis of the lateral electrodes and then the midline analysis.

3.2.1. P1 time window (30–80 ms)

In the lateral analysis (using all but the three midline electrodes) it was found that there was a negativity for deviants relative to the standard that was significantly expressed over all four posterior electrodes (P7, P3, P4, P8), and only two non-posterior electrodes (C3 and F4). The effect sizes (partial η^2) were largest on electrodes close to the midline and on posterior electrodes. A similar pattern was found for the midline analysis (Pz, Cz, Fz), which revealed a main effect for deviant vowels along with a two-way interaction with the anterior to posterior dimension. When broken down the latter reflected a significant negativity on Pz, a small effect on Cz, and no significant effect on Fz. This is in line with the posterior distribution of the deviant effects observable from Fig. 3B.

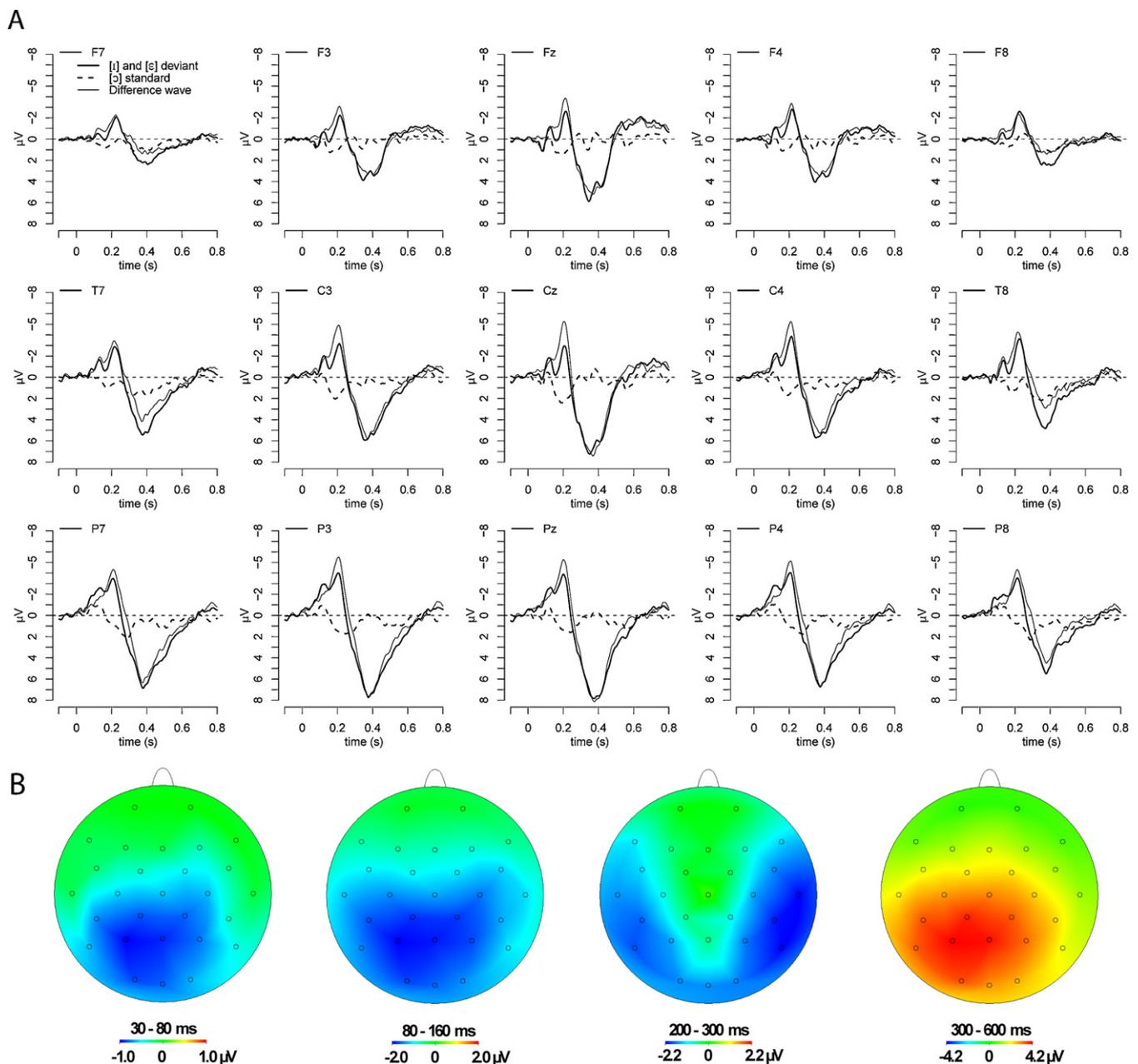


Fig. 3. (A) Grand averages of the ERPs measured in the control condition for the 15 analyzed electrodes. The dashed line represents the standard stimulus, the thick solid line represents the average over the two deviant stimuli, the thin solid line represents the difference wave between the two (deviant–standard). (B) Topographic distributions of the difference wave (average deviant–standard) in the control condition for the 4 selected time-domains P1 (30–80 ms); N1 (80–160 ms); N2 (200–300 ms); P3 (300–600 ms). 28 electrodes are displayed. Note that the scales for the separate time windows differ.

3.2.2. N1 time window (80–160 ms)

In the lateral analysis, deviant vowels gave rise to negativity effects that were largest towards the midline for posterior and central electrodes, but towards the lateral electrodes for the frontal electrodes. Small hemispheric asymmetries showed that for the right hemisphere negativities were stronger towards midline electrodes while the left hemisphere showed more reliable negativities on the lateral electrodes. The midline analysis revealed a main effect for deviant and an interaction with the anterior to posterior dimension. There were more reliable negativities for posterior electrodes.

3.2.3. N2 time window (200–300 ms)

The lateral analysis showed that deviants elicited negativities that were more reliable over lateral electrode sites and towards

posterior electrodes. The midline analysis revealed an interaction between the anterior to posterior dimension with the factor Deviant, but when broken down none of the electrodes showed significant effects of Deviant.

3.2.4. P3 time window (300–600 ms)

The lateral analysis showed a small hemispheric difference in positivity which showed that for the right hemisphere effects were stronger towards midline electrodes while for the left hemisphere effect were stronger on lateral electrodes. Furthermore, for anterior electrodes effects were stronger over the right hemisphere while for medial electrodes effects were stronger for the left hemisphere. For posterior electrodes effects were of a similar size for the left and right hemisphere. The midline analysis revealed a main effect of Deviant and more reliable positivities towards posterior electrodes.

3.2.5. Summary

Effects of presenting a deviant non-word were found as a negativity in the earliest time window that was analyzed (30–80 ms). The effects in this time window were in a negative direction though, indicating that the detection of deviants did not result in a stronger P1. Effects of deviant detection also resulted in strong negativities in the N1 and N2 time windows. Positive enhancement was found in the P3 time domain. In general, these effects were more reliable over posterior sites. These findings show that our design is capable of producing cortical effects during all time windows reflecting the detection of the deviants [ϵ papu] and [ɪpapu].

3.3. EEG data: experimental standard versus deviant analyses

The second analysis compared the average experimental deviant (average over [ϵ papu] and [ɪpapu] in both the high and the low F_1 context conditions) to the average experimental standard ([$^l_\epsilon$ papu] in both the high and the low F_1 context condition), following the same protocol as the analysis for the above control condition. Fig. 4A displays the grand averages of the standards (dashed) and the deviants (solid) for all analyzed electrodes, along with the difference wave (thin solid line). Fig. 4B displays the scalp distribution of the difference wave over the 4 different time windows. Appendix B displays the effects that include the factor Deviant, along with the broken-down constituents for the highest order interactions that include this factor. From Fig. 4A, a first observation can be made that the deviant effects are in general much smaller than the effects observed in the control comparison. This was expected as the acoustic difference between the ambiguous sound [$^l_\epsilon$] and the deviants [ϵ] and [ɪ] that were used here was much smaller than the acoustic difference between the standards and deviants used in the control condition. Analyses for the separate time windows investigated at what points in time the effect of deviant detection was observed.

3.3.1. P1 time window (30–80 ms)

In the analysis of the lateral electrodes (12 electrodes: all but the three midline electrodes) stronger effects of deviants were found towards central and posterior electrodes. The analysis of the midline electrodes (Pz, Cz, Fz) also revealed a main effect and an interaction with the anterior to posterior dimension. When this interaction was broken down effects were stronger towards posterior electrodes, with the strongest effect for Pz.

3.3.2. N1 time window (80–160 ms)

The analysis of the lateral electrodes revealed that all but one set of electrodes (F7, F8) revealed significant effects of deviants. The strongest effects were located on posterior electrodes towards the midline (P3, P4). The midline analysis revealed a similar pattern as a main effect of Deviant and an interaction of Deviant with the anterior to posterior dimension was found. When broken down this also showed that the largest negativities were located towards the posterior electrodes (Cz, Pz).

3.3.3. N2 time window (200–300 ms)

The analysis of the lateral electrodes revealed that deviants elicited negativities that were spread over all of the analyzed sets of electrodes. The strongest effects were observed, however, over central and posterior electrodes towards the midline (C3, C4, P3, P4). The midline analysis revealed a similar pattern. A main effect of Deviant and an interaction of Deviant with the anterior-posterior dimension was again found. When this was broken down the strongest effects resided on the medial and posterior electrodes (Cz and Pz).

3.3.4. P3 time window (300–600 ms)

The analysis of the lateral electrodes revealed effects of Deviant that, when broken down, were stronger over left than over right electrodes, and generally stronger towards the midline. The midline analysis revealed a main effect of Deviant and a small interaction of Deviant with the anterior to posterior dimension. The effect of deviant was strong over the whole midline but slightly stronger towards anterior electrodes.

3.3.5. Summary

These analyses showed that an effect of the detection of a deviant can also be observed with the relatively small vowel change of the ambiguous sound towards [ɪ] or [ϵ] which were only detected as deviant on 52.5% of the trials. Effects of deviant detection were again observed mainly over posterior electrodes and during all four tested time windows. We can now test which of these effects were influenced by context.

3.4. EEG data: context effects

Fig. 5A displays the difference waves for the vowels presented in the different context conditions (solid = ([ɪ] low F_1) – ([ɪ] high F_1); dashed = ([ϵ] low F_1) – ([ϵ] high F_1)), along with the difference wave between the two (thin solid line). The latter reflects the numerical interaction effect. Fig. 5B displays an enlarged version of the panel for electrode P7 (posterior left lateral electrode). Fig. 5C displays the scalp distribution of the interaction effect over the 4 different time windows. The behavioural data showed that listeners found it hard to detect a shift from the ambiguous sound [$^l_\epsilon$] to [ɪ] in the high F_1 context condition, and from [$^l_\epsilon$] to [ϵ] in the low F_1 context condition. If the compensation effect were due to a late and high-level process, the interaction should only be reflected in a late cortical signature such as the P3. If it were due to an early process then it should be observed in an earlier time window such as P1 or N1.

The effect of context should be expressed as follows, taking the P3 response as an example: listeners found it hard to detect a shift from the ambiguous sound [$^l_\epsilon$] to [ɪ] in the high F_1 context condition. This should thus result in a small cortical deviant effect in the P3 for that condition. In the low F_1 context condition the detection of a shift from the ambiguous sound [$^l_\epsilon$] to [ɪ] should be easier, however, leading to a larger effect on the P3. The difference between these (([ɪ] low F_1) – ([ɪ] high F_1)) was thus expected to result in a positive P3 effect. For the [ϵ] deviants this pattern was expected to be in the opposite direction. The difference line for the [ϵ] deviant (([ϵ] low F_1) – ([ϵ] high F_1)) should therefore result in a negative P3 effect. It was thus expected that the difference line for [ɪ] and the difference line for [ϵ] would show mirror-image patterns around zero. Critically, this relation should hold from the point where the context effect exerts its influence (i.e., not only in the P3 window). This approach therefore shows at what point in time the interaction effect is visible. Fig. 5A (and B for electrode P7) displays the ERPs and shows the mirror-image pattern in the P3 window, especially on central and posterior electrodes, but also in earlier time-windows. Appendix C reports the significant effects that include the interaction between Context and Vowel (instead of the factor “Deviant” as in the two previous analyses), along with the broken-down highest interactions.

3.4.1. P1 time window (30–80 ms)

In the lateral analysis (12 electrodes: all but the three midline electrodes), a small four-way interaction that included the Context by Vowel interaction was found. When broken down, however,

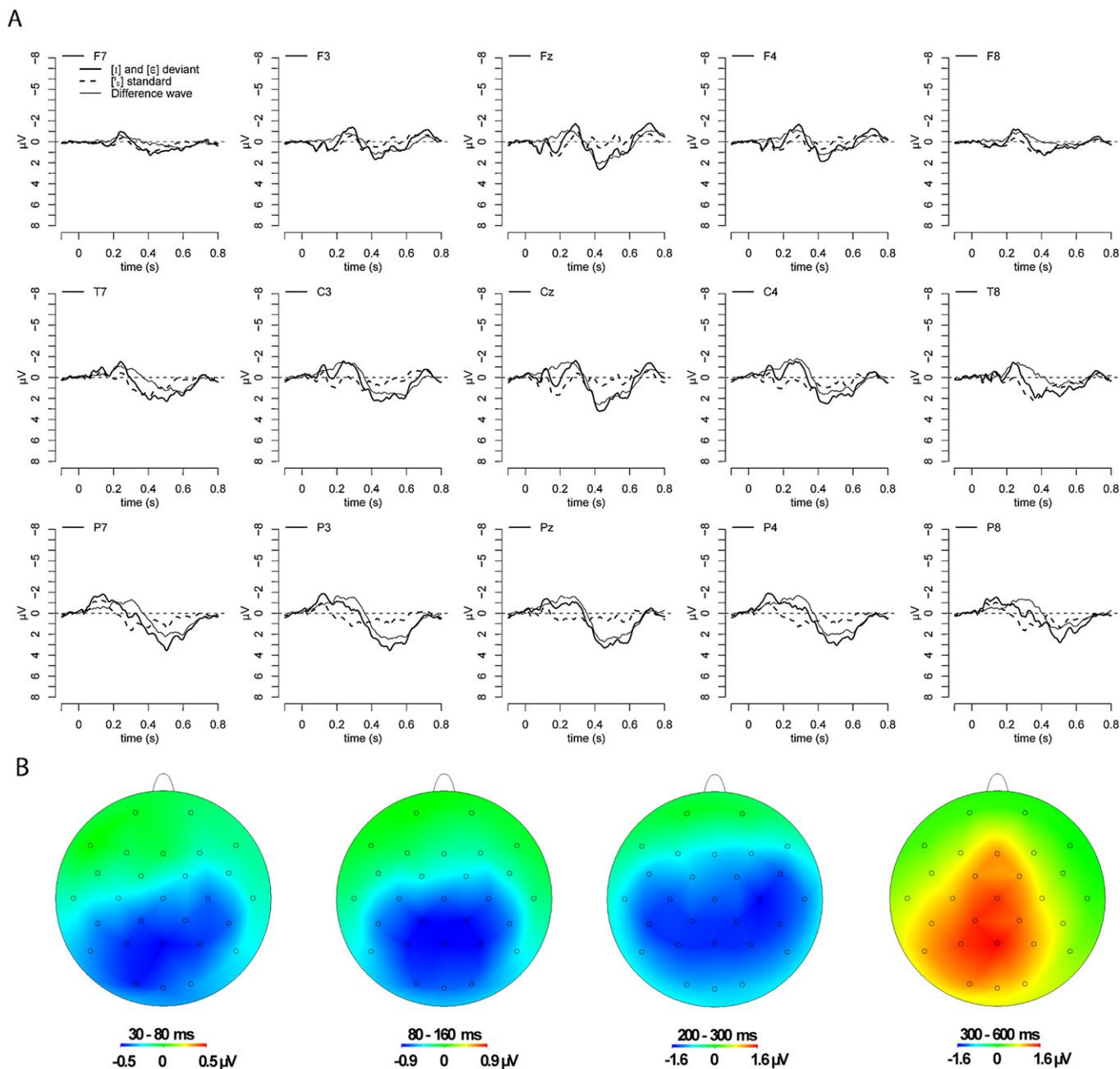


Fig. 4. (A) Grand averages of the ERPs measured in the experimental condition for the 15 analyzed electrodes. The dashed line represents the standard stimulus (averaged over the two context conditions), the thick solid line represents the average over the two deviant stimuli (averaged over the two context conditions), the thin solid line represents the difference wave between the two (deviant–standard). (B) Topographic distributions of the difference wave (average deviant–standard) in the control condition for the 4 selected time-domains P1 (30–80 ms); N1 (80–160 ms); N2 (200–300 ms); P3 (300–600 ms). 28 electrodes are displayed. Note that the scales for the separate time windows differ.

none of the separate electrode pairs revealed a significant effect of Context by Vowel. No effect was found in the midline analysis (Pz, Cz, Fz).

3.4.2. N1 time window (80–160 ms)

In the lateral analysis two three-way and one four-way interactions that included the critical Context by Vowel interaction were found. When broken down these analyses revealed effects of Context by Vowel only on the left lateral electrodes (F7, T7, P7), and a trend towards an effect on the posterior electrodes (P7, P3, P4, P8; $p=0.063$, partial $\eta^2=0.134$). No effect was found for the midline electrodes.

3.4.3. N2 time window (200–300 ms)

The analysis of the lateral electrodes revealed a two-way interaction between Context and Vowel and a three-way interaction of Context by Vowel with the anterior to posterior distribution. When broken down, the strongest effects of Context by Vowel were found over the posterior electrodes (P7, P3, P4, P8). The midline analysis revealed a small Context by Vowel interaction.

3.4.4. P3 time window (300–600 ms)

The lateral analysis revealed a three-way and a four-way interaction. When broken down none of the pairs of electrodes reached significance. The midline analysis revealed an interaction of Context by Vowel.

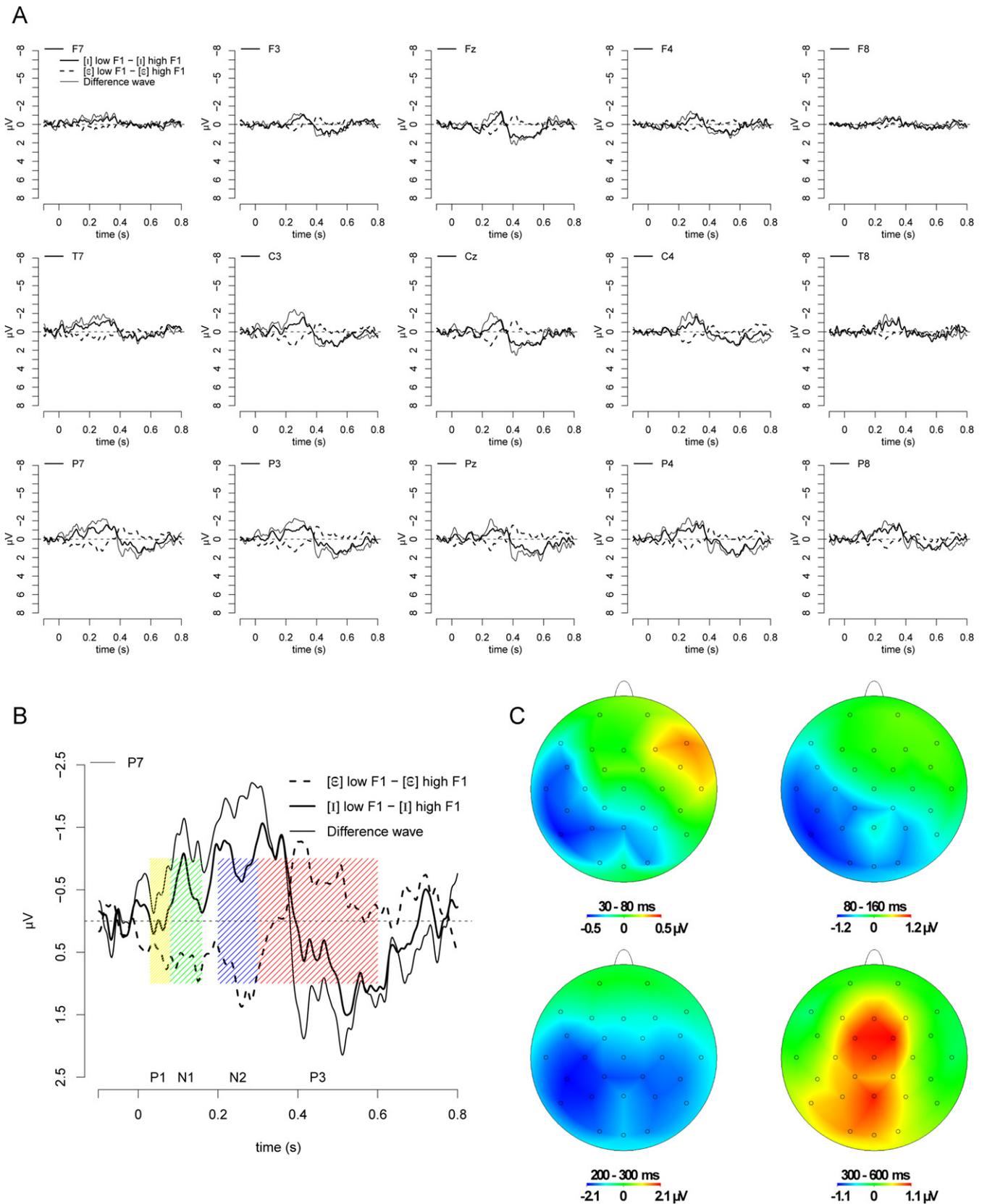


Fig. 5. (A) Difference waves of the ERP for the 15 analyzed electrodes in the experimental condition, investigating the critical vowel by context interaction. The dashed line represents the difference wave for [ɛpapu] (low F_1 –high F_1), the thick solid line represents the difference wave for [ɪpapu] (low F_1 –high F_1), and the thin solid line represents the difference between the two (i.e., the difference between the context effects for each deviant), and thus reflects the interaction effect. (B) An enlarged version of electrode P7. The lines represent the same data as in (A). Additionally, the four analyzed time windows are indicated along with color coding: P1, yellow (30–80 ms); N1, green (80–160 ms); N2, blue (200–300 ms); P3, red (300–600 ms). (C) Topographic distributions of the difference wave representing the numerical interaction effect (“low F_1 [ɛpapu]”–“high F_1 [ɛpapu]”)–(“low F_1 [ɪpapu]”–“high F_1 [ɪpapu]”) in the experimental condition for the 4 selected time-domains. Note that the scales for the separate time windows differ. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

3.4.5. Summary

These critical analyses revealed cortical signatures of the interaction between the F_1 properties of the /papu/ context and the identity of the target vowel. The earliest reliable effect was observed in the N1 time window (there was a small effect in the P1 time-window, but when broken down none of the electrode sites revealed significant effects). The N1 effect had a left lateralized distribution. The analyses of the N2 and P3 time windows also revealed reflections of the interaction between the F_1 properties of the /papu/ context and vowel identity.

4. General discussion

The current paper investigated the level of processing at which compensation for vocal-tract characteristics in speech perception has its influence. In a two-deviant active oddball design, listeners were asked to detect the vowel deviants [ɛ] and [ɪ], embedded in a stream of standards that consisted of an ambiguous sound [ɪ_ε] (a sound that was acoustically halfway between [ɛ] and [ɪ]). These standard and deviant vowels were prepended to a non-word context consisting of two syllables (/papu/) that were spliced onto the vowels. Critically, in separate blocks, the /papu/ part was manipulated to mimic either a speaker with a generally high F_1 or a speaker with a generally low F_1 (the standard and deviant vowels were identical across these conditions). These conditions tested compensation for vocal-tract characteristics, as the conditions that mimicked different vocal tracts were expected to influence the detectability of the deviant vowels (represented by both the behavioural and cortical responses) in different ways. In an additional control block the deviant vowels were the same versions of [ɛ] and [ɪ]. However, the standard vowel was [ɔ]. Moreover, in this control condition the /papu/ part had a neutral F_1 range. The cortical responses to these stimuli were recorded and analyzed over 4 different time windows. These were the P1 (30–80 ms), the N1 (80–160 ms), N2 (200–300) and the P3 (300–600 ms) time windows.

The critical analysis investigated compensation for the F_1 properties of the /papu/ context. The behavioural detection results showed that listeners found it harder to detect a shift from [ɪ_ε] to [ɪ] in the high F_1 context condition (compared to detecting a shift from [ɪ_ε] to [ɛ]), while the opposite was true in the low F_1 context condition. As has been shown before, this means that the same acoustic token of a vowel, presented in two different contexts, can be perceived differently (Ladefoged & Broadbent, 1957).

The cortical responses to deviants that were recorded during this task showed that the effect of compensation for F_1 properties in the context was not yet observed in the P1 time-domain. A marginally significant 4-way interaction hinted at an effect, but when this interaction was broken down, data from none of the separate electrode sites revealed significant effects. At this level of processing the comparison of the deviant vowels to the representation of the standards did thus not induce reliable differences in the cortical response. During the N1 time window, however, an effect of compensation for context properties was observed. This suggests that the compensation process influences an early level of processing. The negativity in the N1 time window was observed during the initial vowel and could thus not have been a direct reflection of the exogenous influence of the subsequent context signal (which followed 250 ms after vowel onset).

The fact that we observed an effect in the N1 time domain suggests that the normalization process has a relatively early cognitive locus. As mentioned in the introduction, Roberts et al. (2004) report that a decision bias induced by the identity of a preceding trial did not lead to changes in the N1 of a subsequent trial. We observed the consequences of compensation in the N1 domain and not only

during N2/P3 time windows. This shows that it is unlikely that compensation effects are the result of decision strategies, biases or the listener's conscious interpretation of speech sounds, occurring during the processing of either the standard or the deviant vowel. Instead, normalization appears to change the perception of speech sounds at a very early level of processing.

The control analyses investigated the cortical responses that were recorded during the control condition (standard [ɔ] and deviants [ɛ] and [ɪ], all with a neutral [papu] context) and the cortical responses to the experimental materials (standard: [ɪ_ε], averaged over both the high F_1 and the low F_1 context condition; deviants: average over [ɛ] and [ɪ] over both the high F_1 and the low F_1 context condition). These analyses established that our design was capable of eliciting cortical deviant responses, and in which direction (positive or negative) the effects in the different time windows were expected. In both analyses, the deviants resulted in reliable differences in the cortical signature during the P1 time window, observed as a less strong positivity, especially over central posterior electrodes. During the N1 and the N2 time windows, the deviants resulted in stronger negative deflections. During the P1 and N1 time windows the cortical responses to the detection of an oddball were mainly expressed over posterior electrodes. During the N2 time window the distribution seemed to be spreading also towards lateral and frontal areas. This broad negativity had a topographical distribution that was unlike the more frontocentral distribution that is found in classical MMN experiments (Näätänen & Winkler, 1999). This suggests that the negativity should not be readily interpreted as an MMN. This conclusion is supported by the observation that there was no positive deflection at the mastoids during this time window (or the P1 and N1 time windows).

The positivities to deviants in the P3 time domain were similar to other reports of mismatch detection in a response active oddball experimental design. The mainly posterior distribution of the positivity in these conditions suggests that this component should be interpreted as the P3b that is observed after deviants that are infrequent but not unexpected (Friedman et al., 2001). From a comparison of the deviant waveforms between the midline electrodes (mainly Cz and Pz) in the control condition (Fig. 3A) and the experimental standard versus deviant effects (Fig. 4A) it can be observed that the P3 signal is longer in the experimental condition than in the control condition. Lengthening of P3 as a function of increasing task difficulty or complexity of the stimulus evaluation is a known characteristic of the P3 wave (McCarthy & Donchin, 1981). Although the deviants [ɛ] and [ɪ] were identical in these conditions the difference between the standard and the deviants was much larger in the control condition (where the standard was the vowel [ɔ]) than in the experimental condition where the standard was the ambiguous sound [ɪ_ε] which was acoustically halfway between the deviants [ɛ] and [ɪ]. The two control analyses showed that effects could be observed during all four time windows and indicated whether positivities or negativities should be expected in the different time windows. These analyses validated our analysis of the critical interaction.

The fact that the analyses of the control conditions revealed reliable effects in all time windows allows us to address two questions about the vowel normalization mechanism in the experimental condition: at what point in time does the effects of the compensation mechanism reveal its influence, and at what point in time is that influence finished? With respect to the first question, unlike the analysis of the average standards versus the average deviants, no reliable effect was observed during P1 in the critical experimental analysis. It is important to note here that the critical analysis was based on the analysis of the same (number of) deviant trials as the deviant trials that were used for the analysis of the average standards versus the average deviants. The lack of a reliable effect in the P1 domain can thus be interpreted as evidence that compensation

effects affect processes that take place during the N1 time window, and no earlier. However, Fig. 5B and C shows that during the P1 time window there was already a trend towards the effect that became significant during the N1 time window (a left lateralized negativity). There are two possible interpretations of this observation. First, this trend could reflect changes in perception that have their origin earlier in the processing stream, but which are not significant due to a lack of experimental power. This would mean that the interaction effect visible during the N1 time window is a mere consequence of the fact that compensation has influenced levels of representation even earlier in the processing stream (possibly in the periphery of the auditory system). It has been shown that brain-stem level processes can influence hair-cell activity in the cochlea in a suppressive manner, and this mechanism has been argued to improve perception in noisy environments (and with extreme intensities as a protective mechanism against acoustic trauma, Kirk & Smith, 2003; May, Budelis, & Niparko, 2004). It is possible that these projections can operate in a sufficiently sophisticated way to induce context effects (Stilp et al., 2010). Such a mechanism would be a very peripheral implementation of normalization effects.

A second possible interpretation of the initial trend of an effect on P1 is that it is a predecessor of the process that is strong in N1. The processes that become dominant in the N1 time window might already have been partly active during the P1 time window. The first waves of activity after auditory stimulation reach the primary auditory cortex after 10–15 ms (Liegeois-Chauvel, Musolino, & Chauvel, 1991). This makes it possible that a change-sensitive process that produces its peak activity during the N1 time domain is already becoming active in earlier time windows.

Some aspects of our data make the second interpretation more likely. It is possible that deviant detection in the control and experimental standard versus deviant conditions led to early deviant effects in subcortical areas. The fact that the earliest time window that was analyzed (P1) showed robust deviant detection effects in the control condition supports this interpretation. This would mean, however, that if compensation processes also influenced early subcortical representations, then, at the level of the cortex, compensation should have been resolved and should only lead to increased detectability. In the critical analysis, the context by vowel interaction in the N1 time-domain showed a distinctively left-sided distribution, mainly on the lateral electrodes. The slowly rising negativity that was visible during the detection of deviants in both the control condition and the overall comparison of experimental deviants versus standards had a distinctly central posterior distribution. The interaction effect visible in the critical context analysis in the P1 and N1 time domains therefore seems to have a different scalp distribution (one that is more strongly left lateralized) than the effects in the same window for the control and the experimental standard-deviant effects (compare the distributions in the N1 time window of Figs. 3B, 4B and 5C). These observations suggest that the interaction wave in the critical comparison is not just a reflection of increased detectability as a result of compensation for context instantiated in the peripheral system. The current data therefore suggest that the compensation mechanism under investigation here starts to influence processing just before or during the N1 time window, and certainly no later than that. Future research could focus on earlier components of the ERP signal to further inform the discussion of what the earliest point in time is where compensation mechanisms influences perception.

With respect to the second question, concerning the point in time when the compensation mechanism has no further influence, the current data also provide information. The effects in the later two time windows (N2 and P3) in the context condition were very similar with respect to their topographies to those observed during the control and experimental conditions. This indicates that once compensation has had its influence, the increase in detectability

that is the result of this compensation mechanism just adds to the overall detectability of the deviants. The compensation mechanism does not seem to influence any higher representational levels as the cortical response reflecting compensation no longer displays a topography that is different from regular deviant detection. The current findings thus suggest that at the start of the N2 time window (200 ms after target onset) the compensation mechanism has already exerted its influence, and has little additional influence on processing after that point.

A final question about normalization concerns the way in which it is implemented. It has been proposed that an important part of compensation for vocal-tract characteristics in a preceding carrier sentence is based on compensation for average spectral distributions in context signals (Watkins & Makin, 1994, 1996). Listeners continuously build up a representation of the LTAS of preceding sounds, and interpret subsequent signals relative to that LTAS. Watkins and Makin have argued that the result is that listeners perceive target sounds as if they were inversely filtered for the precursor signal by decreasing the perceptual impact of those frequency regions that were very pronounced in that precursor. Such an operation would thus make listeners sensitive to changing acoustic properties (Kluender et al., 2003; Kluender & Kieft, 2006; Stilp et al., 2010). Compensation for LTAS cannot account for all normalization findings (see, for instance, Johnson, Strand, & D'Imperio, 1999, for phoneme categorization-shifts induced by visual context, and Sjerps et al., 2011, for the demonstration that normalization did not always occur for non-speech stimuli, even though those stimuli had the similar LTAS relations as matched speech stimuli that did elicit normalization). Given the spectral relations between precursors and targets in the experiment presented here, however, compensation for spectral characteristics is likely to play an important role. The N1 has been shown to be sensitive to differences of F_1 (Diesch et al., 1996; Roberts et al., 2004), or F_1 and F_2 (Obleser, Elbert, Lahiri, & Eulitz, 2003; Poeppel et al., 1997; Tiitinen, Makela, Makinen, May, & Alku, 2005), and, in an extensive review, Näätänen and Winkler (1999) suggest that the processes underlying N1 retain information on individual static stimulus features. The proposal by Watkins and Makin (1994, 1996) necessarily assumes some form of storage of the LTAS of preceding input. The suggestion that the processes that underlie the N1 retain information on static stimulus features makes them a good candidate for being responsible for LTAS-based compensation. The current results support this interpretation.

Additionally, it has also been suggested that N1 might be sensitive to the F_1/F_3 ratio instead of absolute formant values (Monahan & Idsardi, 2010). Similar to the rationale in the current paper, that proposal was made in the light of a mechanism that helps listeners deal with between-speaker variation in formant frequencies, but then in a vowel-intrinsic manner (i.e., relying only on information within the target vowel). In contrast to Monahan and Idsardi (2010) and the other papers mentioned above, the current paper focused on vowel-extrinsic influences on perception, as the set of target vowels were identical across the two context conditions. In speech perception, vowel intrinsic normalization and normalization for vowel-extrinsic acoustic information probably operate in tandem (Johnson, 2005), and extrinsic normalization procedures have the potential to play a large role in overcoming between-speaker variation that is the result of anatomical/physiological differences between speakers (Adank, Smits, & van Hout, 2004). The findings reported here, in combination with previous reports in the literature, suggest that the processes underlying the N1 play an important role in normalizing perceptual input to reduce within-category variability.

The method used in this paper presents a novel approach to research on normalization of vowels for the spectral properties of a context speakers' voice. The results demonstrated that

compensation for speaker vocal-tract characteristics can be observed through cortical measures and as such provide information about the level of representation that this process influences. The consequences of compensation for speaker characteristics were observed as soon as 120 ms after vowel onset. This makes it unlikely that the context-induced shifts in perception as they have been observed by others (Holt, 2005; Ladefoged & Broadbent, 1957; Sjerps et al., 2011; Watkins, 1991; Watkins & Makin, 1994, 1996) were due to strategic effects or biases based on conscious percepts. Instead, compensation for speaker vocal-tract characteristics is for

an important part the result of a mechanism that influences an early stage of processing in speech perception.

Acknowledgments

We would like to thank Marcel Bastiaansen for suggestions while designing the study, Katja Poelmann and Ellie van Setten for assistance with electrode application, and two anonymous reviewers for constructive commentary.

Appendix A. Control condition

Component	Analysis	Factor		F (df)	p	Partial η^2				
P1	Lateral	Deviant		10.527 (1, 23)	0.004	0.314				
		AntPost \times Hemisphere \times MedLat \times Deviant		3.798 (2, 46)	0.038	0.142				
		Ant	Left	Lat	Deviant (F7)	1.923 (1, 23)	0.179	0.077		
				Med	Deviant (F3)	3.125 (1, 23)	0.090	0.120		
			Right	Med	Deviant (F4)	5.966 (1, 23)	0.023	0.206		
				Lat	Deviant (F8)	0.027 (1, 23)	0.871	0.001		
		Med	Left	Lat	Deviant (T7)	2.653 (1, 23)	0.117	0.103		
				Med	Deviant (C3)	12.059 (1, 23)	0.002	0.344		
			Right	Med	Deviant (C4)	0.538 (1, 23)	0.471	0.023		
				Lat	Deviant (T8)	0.500 (1, 23)	0.487	0.021		
		Post	Left	Lat	Deviant (P7)	8.950 (1, 23)	0.007	0.280		
				Med	Deviant (P3)	16.522 (1, 23)	<0.001	0.418		
			Right	Med	Deviant (P4)	13.545 (1, 23)	0.001	0.371		
				Lat	Deviant (P8)	7.690 (1, 23)	0.011	0.251		
		Midline		Deviant		5.162 (1, 23)	0.033	0.183		
				AntPost \times Deviant		19.979 (2, 46)	<0.001	0.465		
				Ant	Deviant (Fz)	0.416 (1, 23)	0.525	0.018		
				Med	Deviant (Cz)	5.302 (1, 23)	0.031	0.187		
				Post	Deviant (Pz)	11.740 (1, 23)	0.002	0.338		
		N1	Lateral	Deviant		43.426 (1, 23)	<0.001	0.654		
MedLat \times Deviant				9.835 (1, 23)	0.005	0.300				
AntPost \times Deviant				25.320 (2, 46)	<0.001	0.524				
Hemisphere \times MedLat \times Deviant	Left			Lat	Deviant (F7, T7, P7)	4.516 (1, 23)	0.045	0.164		
				Med	Deviant (F7, T7, P7)	39.291 (1, 23)	<0.001	0.631		
	Right			Med	Deviant (F3, C3, P3)	38.760 (1, 23)	<0.001	0.628		
				Lat	Deviant (F4, C4, P4)	39.057 (1, 23)	<0.001	0.629		
AntPost \times MedLat \times Deviant	Left			Lat	Deviant (F8, T8, P8)	30.370 (1, 23)	<0.001	0.569		
				Med		8.727 (2, 46)	0.001	0.275		
	Right			Lat	Deviant (F7, F8)	35.396 (1, 23)	<0.001	0.606		
				Med	Deviant (F3, F4)	20.425 (1, 23)	<0.001	0.470		
Med	Left			Lat	Deviant (T7, T8)	38.118 (1, 23)	<0.001	0.624		
				Med	Deviant (C3, C4)	41.862 (1, 23)	<0.001	0.645		
	Right			Lat	Deviant (P7, P8)	31.470 (1, 23)	<0.001	0.578		
				Med	Deviant (P3, P4)	45.479 (1, 23)	<0.001	0.664		
AntPost \times Hemisphere \times Deviant	Left			Lat		3.634 (2, 46)	0.035	0.136		
				Med	Deviant (F7, F3)	28.014 (1, 23)	<0.001	0.549		
	Right			Lat	Deviant (F4, F8)	24.248 (1, 23)	<0.001	0.513		
				Med	Deviant (T7, C3)	38.230 (1, 23)	<0.001	0.624		
Med	Left			Lat	Deviant (C4, T8)	38.205 (1, 23)	<0.001	0.624		
				Med	Deviant (P7, P3)	39.447 (1, 23)	<0.001	0.632		
	Right			Lat	Deviant (P4, P8)	37.116 (1, 23)	<0.001	0.617		
				Med						
Midline				Deviant		16.755 (1, 23)	<0.001	0.421		
				AntPost \times Deviant		18.476 (2, 46)	<0.001	0.445		
				Ant	Deviant (Fz)	6.704 (1, 23)	0.016	0.226		
				Med	Deviant (Cz)	15.461 (1, 23)	0.001	0.402		
				Post	Deviant (Pz)	23.652 (1, 23)	<0.001	0.507		
N2	Lateral			Deviant		15.866 (1, 23)	0.001	0.408		
				MedLat \times Deviant		4.611 (1, 23)	0.043	0.167		
				Med	Deviant (F3, F4, C3, C4, P3, P4)	Lat	Deviant (F7, F8, T7, T8, P7, P8)	6.955 (1, 23)	0.015	0.232
						Med		33.340 (1, 23)	<0.001	0.592
				AntPost \times Deviant	Left	Lat	Deviant (F7, F3, F4, F8)	9.197 (2, 46)	0.001	0.286
						Med	Deviant (T7, C3, C4, T8)	10.328 (1, 23)	0.004	0.310
					Right	Lat	Deviant (T7, C3, C4, T8)	15.109 (1, 23)	0.001	0.396
						Med	Deviant (P7, P3, P4, P8)	17.445 (1, 23)	<0.001	0.431

Component	Analysis	Factor	<i>F</i> (df)	<i>p</i>	Partial η^2		
P3	Midline	AntPost \times Deviant		3.828 (2, 46)	0.035	0.143	
		Ant	Deviant (Fz)	0.149 (1, 23)	0.704	0.006	
		Med	Deviant (Cz)	0.026 (1, 23)	0.873	0.001	
		Post	Deviant (Pz)	1.343 (1, 23)	0.258	0.055	
	Lateral	Deviant		81.817 (1, 23)	<0.001	0.781	
		MedLat \times Deviant		22.720 (1, 23)	<0.001	0.497	
		Hemisphere \times Deviant		23.159 (1, 23)	<0.001	0.502	
		AntPost \times Deviant		60.290 (2, 46)	<0.001	0.724	
		Hemisphere \times MedLat \times Deviant		6.244 (1, 23)	0.020	0.214	
		Left	Lat	Deviant (F7, T7, P7)	89.169 (1, 23)	<0.001	0.795
			Med	Deviant (F3, C3, P3)	58.140 (1, 23)	<0.001	0.717
		Right	Med	Deviant (F4, C4, P4)	79.908 (1, 23)	<0.001	0.776
			Lat	Deviant (F8, T8, P8)	61.902 (1, 23)	<0.001	0.729
		AntPost \times MedLat \times Deviant		4.214 (2, 46)	0.026	0.155	
		Ant	Lat	Deviant (F7, F8)	41.473 (1, 23)	<0.001	0.643
			Med	Deviant (F3, F4)	39.373 (1, 23)	<0.001	0.631
		Med	Lat	Deviant (T7, T8)	89.685 (1, 23)	<0.001	0.796
			Med	Deviant (C3, C4)	58.754 (1, 23)	<0.001	0.719
		Post	Lat	Deviant (P7, P8)	73.945 (1, 23)	<0.001	0.763
			Med	Deviant (P3, P4)	81.729 (1, 23)	<0.001	0.780
		AntPost \times Hemisphere \times Deviant		16.433 (2, 46)	<0.001	0.417	
		Ant	Left	Deviant (F7, F3)	35.124 (1, 23)	<0.001	0.604
			Right	Deviant (F4, F8)	48.262 (1, 23)	<0.001	0.677
		Med	Left	Deviant (T7, C3)	70.884 (1, 23)	<0.001	0.755
			Right	Deviant (C4, T8)	64.078 (1, 23)	<0.001	0.736
		Post	Left	Deviant (P7, P3)	80.429 (1, 23)	<0.001	0.778
			Right	Deviant (P4, P8)	80.459 (1, 23)	<0.001	0.778
		Midline	Deviant		47.682 (1, 23)	<0.001	0.675
			AntPost \times Deviant		19.404 (2, 46)	<0.001	0.458
			Ant	Deviant (Fz)	18.315 (1, 23)	<0.001	0.443
			Med	Deviant (Cz)	36.848 (1, 23)	<0.001	0.616
			Post	Deviant (Pz)	66.104 (1, 23)	<0.001	0.742

Appendix B. Deviant–standard comparison

Component	Analysis	Factor	<i>F</i> (df)	<i>p</i>	Partial η^2		
P1	Lateral	Deviant		11.954 (1, 23)	0.002	0.342	
		AntPost \times Deviant		4.818 (2, 46)	0.035	0.173	
		Ant	Deviant (F7, F3, F4, F8)	4.601 (1, 23)	0.043	0.167	
		Med	Deviant (T7, C3, C4, T8)	11.404 (1, 23)	0.003	0.331	
		Post	Deviant (P7, P3, P4, P8)	9.870 (1, 23)	0.005	0.300	
	Midline	Deviant		7.237 (1, 23)	0.013	0.239	
		AntPost \times Deviant		5.968 (2, 46)	0.014	0.206	
		Ant	Deviant (Fz)	0.486 (1, 23)	0.493	0.021	
		Med	Deviant (Cz)	7.291 (1, 23)	0.013	0.241	
		Post	Deviant (Pz)	12.535 (1, 23)	0.002	0.353	
	N1	Lateral	Deviant		31.778 (1, 23)	<0.001	0.580
MedLat \times Deviant			27.983 (1, 23)	<0.001	0.549		
AntPost \times Deviant			17.830 (2, 46)	<0.001	0.437		
AntPost \times MedLat \times Deviant			10.811 (2, 46)	<0.001	0.320		
Ant			Lat	Deviant (F7, F8)	0.920 (1, 23)	0.347	0.038
			Med	Deviant (F3, F4)	6.802 (1, 23)	0.016	0.228
Med			Lat	Deviant (T7, T8)	5.897 (1, 23)	0.023	0.204
			Med	Deviant (C3, C4)	49.336 (1, 23)	<0.001	0.682
Post			Lat	Deviant (P7, P8)	18.125 (1, 23)	<0.001	0.441
			Med	Deviant (P3, P4)	66.861 (1, 23)	<0.001	0.744
Midline		Deviant		40.171 (1, 23)	<0.001	0.636	
		AntPost \times Deviant		11.154 (2, 46)	0.002	0.327	
		Ant	Deviant (Fz)	3.678 (1, 23)	0.068	0.138	
		Med	Deviant (Cz)	39.084 (1, 23)	<0.001	0.630	
		Post	Deviant (Pz)	58.571 (1, 23)	<0.001	0.718	
		N2	Lateral	Deviant		81.526 (1, 23)	<0.001
MedLat \times Deviant				15.270 (1, 23)	0.001	0.399	
AntPost \times Deviant				18.705 (2, 46)	<0.001	0.449	

Component	Analysis	Factor	F (df)	p	Partial η^2		
P3	Midline	AntPost \times MedLat \times Deviant		3.297 (2, 46)	0.050	0.125	
		Ant	Lat	Deviant (F7, F8)	28.098 (1, 23)	<0.001	0.550
			Med	Deviant (F3, F4)	43.469 (1, 23)	<0.001	0.654
			Med	Lat	Deviant (T7, T8)	52.219 (1, 23)	<0.001
		Post	Med	Deviant (C3, C4)	95.766 (1, 23)	<0.001	0.806
			Lat	Deviant (P7, P8)	50.073 (1, 23)	<0.001	0.685
			Med	Deviant (P3, P4)	75.792 (1, 23)	<0.001	0.767
		AntPost \times Hemisphere \times Deviant			3.850 (2, 46)	0.030	0.143
		Ant	Left	Deviant (F7, F3)	25.562 (1, 23)	<0.001	0.526
			Right	Deviant (F4, F8)	39.471 (1, 23)	<0.001	0.632
			Med	Left	Deviant (T7, C3)	70.675 (1, 23)	<0.001
		Post	Right	Deviant (C4, T8)	58.769 (1, 23)	<0.001	0.719
			Left	Deviant (P7, P3)	66.837 (1, 23)	<0.001	0.744
			Right	Deviant (P4, P8)	53.862 (1, 23)	<0.001	0.701
		Deviant			73.052 (1, 23)	<0.001	0.761
	AntPost \times Deviant			9.705 (2, 46)	0.002	0.279	
	Ant	Deviant (Fz)		30.336 (1, 23)	<0.001	0.569	
		Deviant (Cz)		64.379 (1, 23)	<0.001	0.737	
		Deviant (Pz)		83.282 (1, 23)	<0.001	0.784	
	Lateral	Deviant		17.046 (1, 23)	<0.001	0.426	
		MedLat \times Deviant		21.343 (1, 23)	<0.001	0.481	
		Hemisphere \times Deviant		22.806 (1, 23)	<0.001	0.498	
		AntPost \times Deviant		15.343 (2, 46)	<0.001	0.400	
		Hemisphere \times MedLat \times Deviant		7.264 (1, 23)	0.013	0.240	
		Left	Lat	Deviant (F7, T7, P7)	19.616 (1, 23)	<0.001	0.460
			Med	Deviant (F3, C3, P3)	25.318 (1, 23)	<0.001	0.524
			Right	Med	Deviant (F4, C4, P4)	15.586 (1, 23)	0.001
		Right	Lat	Deviant (F8, T8, P8)	1.823 (1, 23)	0.190	0.073
			AntPost \times Hemisphere \times Deviant		4.407 (2, 46)	0.018	0.161
			Ant	Left	Deviant (F7, F3)	14.532 (1, 23)	0.001
Right		Deviant (F4, F8)		5.488 (1, 23)	0.028	0.193	
Med		Left		Deviant (T7, C3)	23.610 (1, 23)	<0.001	0.507
Post		Right	Deviant (C4, T8)	6.190 (1, 23)	0.021	0.212	
		Left	Deviant (P7, P3)	25.934 (1, 23)	<0.001	0.530	
		Right	Deviant (P4, P8)	11.616 (1, 23)	0.002	0.336	
Midline		Deviant		21.120 (1, 23)	<0.001	0.479	
		AntPost \times Deviant		3.862 (2, 46)	0.047	0.144	
		Ant	Deviant (Fz)		21.350 (1, 23)	<0.001	0.481
			Deviant (Cz)		19.720 (1, 23)	<0.001	0.462
		Deviant (Pz)		18.875 (1, 23)	<0.001	0.451	

Appendix C. Context effects

Component	Analysis	Factor	F (df)	p	Partial η^2		
P1	Lateral	Hemisphere \times MedLat \times Context \times Vowel		5.942 (1, 23)	0.023	0.205	
		Left	Lat	Con \times Vow (F7, T7, P7)	1.054 (1, 23)	0.315	0.044
			Med	Con \times Vow (F3, C3, P3)	0.446 (1, 23)	0.511	0.019
		Right	Med	Con \times Vow (F4, C4, P4)	0.010 (1, 23)	0.921	<0.001
			Lat	Con \times Vow (F8, T8, P8)	0.532 (1, 23)	0.473	0.023
N1	Lateral	Hemisphere \times Context \times Vowel		9.791 (1, 23)	0.005	0.299	
		Hemisphere \times MedLat \times Context \times Vowel		12.934 (1, 23)	0.002	0.360	
		Left	Lat	Con \times Vow (F7, T7, P7)	9.118 (1, 23)	0.006	0.284
			Med	Con \times Vow (F3, C3, P3)	3.033 (1, 23)	0.095	0.117
		Right	Med	Con \times Vow (F4, C4, P4)	0.234 (1, 23)	0.633	0.010
			Lat	Con \times Vow (F8, T8, P8)	0.011 (1, 23)	0.916	<0.001
		AntPost \times Context \times Vowel			7.944 (2, 46)	0.005	0.257
		Ant	Con \times Vow (F7, F3, F4, F8)		0.004 (1, 23)	0.952	<0.001
			Con \times Vow (T7, C3, C4, T8)		0.445 (1, 23)	0.511	0.019
			Con \times Vow (P7, P3, P4, P8)		3.831 (1, 23)	0.063	0.143
N2	Lateral	Context \times Vowel		10.294 (1, 23)	0.004	0.309	
		AntPost \times Context \times Vowel		11.901 (2, 46)	0.001	0.341	
		Ant	Con \times Vow (F7, F3, F4, F8)		4.079 (1, 23)	0.055	0.151
			Con \times Vow (T7, C3, C4, T8)		11.447 (1, 23)	0.003	0.332
		Con \times Vow (P7, P3, P4, P8)		13.756 (1, 23)	0.001	0.371	

Component	Analysis	Factor	F (df)	p	Partial η^2	
P3	Midline	Context × Vowel	4.478 (1, 23)	0.045	0.163	
		Lateral				
	MedLat × Context × Vowel	AntPost × MedLat × Context × Vowel	10.366 (1, 23)	0.004	0.311	
		Ant	5955 (2, 46)	0.009	0.206	
		Med	Med Con × Vow (F3, F4)	3.301 (1, 23)	0.082	0.126
			Lat Con × Vow (F7, F8)	0.563 (1, 23)	0.460	0.024
		Post	Med Con × Vow (C3, C4)	2.045 (1, 23)	0.166	0.082
			Lat Con × Vow (T7, T8)	0.030 (1, 23)	0.864	0.001
		Midline	Med Con × Vow (P3, P4)	2.938 (1, 23)	0.100	0.113
			Lat Con × Vow (P7, P8)	1.452 (1, 23)	0.240	0.059
			Context × Vowel	7.454 (1, 23)	0.012	0.245

References

- Adank, P., Smits, R. & van Hout, R. (2004). A comparison of vowel normalization procedures for language variation research. *Journal of the Acoustical Society of America*, 116(5), 3099–3107.
- Adank, P., van Hout, R. & Smits, R. (2004). An acoustic description of the vowels of Northern and Southern Standard Dutch. *The Journal of the Acoustical Society of America*, 116(3), 1729–1738.
- Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, 59(4), 457–474.
- Bates, D. M. & Sarkar, D. (2007). *lme4: Linear mixed-effects models using Eigen and Eigen4* (version 0.999375-27) [software application].
- Bien, H., Lagemann, L., Döbel, C. & Zwitserlood, P. (2009). Implicit and explicit categorization of speech sounds – Dissociating behavioural and neurophysiological data. *European Journal of Neuroscience*, 30(2), 339–346.
- Boersma, P. & Weenink, D. (2005). *Praat: Doing phonetics by computer*.
- Broadbent, D. E. & Ladefoged, P. (1960). Vowel judgements and adaptation level. *Proceedings of the Royal Society of London Series B-Biological Sciences*, 151(944), 384–399.
- Cacace, A. T. & McFarland, D. J. (2003). Quantifying signal-to-noise ratio of mismatch negativity in humans. *Neuroscience Letters*, 341(3), 251–255.
- Celsis, P., Doyon, B., Boulanouar, K., Pastor, J., Demonet, J. F. & Nespoulous, J. L. (1999). ERP correlates of phoneme perception in speech and sound contexts. *Neuroreport*, 10(7), 1523–1527.
- Diesch, E., Eulitz, C., Hampson, S. & Ross, B. (1996). The neurotopography of vowels as mirrored by evoked magnetic field measurements. *Brain and Language*, 53(2), 143–168.
- Dixon, P. (2008). Models of accuracy in repeated-measures designs. *Journal of Memory and Language*, 59(4), 447–456.
- Friedman, D., Cycowicz, Y. M. & Gaeta, H. (2001). The novelty P3: An event-related brain potential (ERP) sign of the brain's evaluation of novelty. *Neuroscience and Biobehavioral Reviews*, 25(4), 355–373.
- Hillenbrand, J., Getty, L. A., Clark, M. J. & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*, 97, 3099–3111.
- Holt, L. L. (2005). Temporally nonadjacent nonlinguistic sounds affect speech categorization. *Psychological Science*, 16(4), 305–312.
- Holt, L. L. (2006). The mean matters: Effects of statistically defined nonspeech spectral distributions on speech categorization. *Journal of the Acoustical Society of America*, 120(5), 2801–2817.
- Johnson, K. (2005). Speaker normalization in speech perception. In D. B. Pisoni, & R. E. Remez (Eds.), *The handbook of speech perception* (pp. 363–389). Oxford: Blackwell.
- Johnson, K., Strand, E. A. & D'Imperio, M. (1999). Auditory-visual integration of talker gender in vowel perception. *Journal of Phonetics*, 27(4), 359–384.
- Joos, M. (1948). Acoustic phonetics. *Language*, 24(2), 5–136.
- Kieft, M. & Kluender, K. R. (2008). Absorption of reliable spectral characteristics in auditory perception. *Journal of the Acoustical Society of America*, 123(1), 366–376.
- Kirk, E. C. & Smith, D. W. (2003). Protection from acoustic trauma is not a primary function of the medial olivocochlear efferent system. *Jaro-Journal of the Association for Research in Otolaryngology*, 4(4), 445–465.
- Kluender, K. R., Coody, J. A. & Kieft, M. (2003). Sensitivity to change in perception of speech. *Speech Communication*, 41(1), 59–69.
- Kluender, K. R. & Kieft, M. J. (2006). Speech perception within a biologically realistic information-theoretic framework. In M. A. Gernsbacher, & M. Traxler (Eds.), *Handbook of psycholinguistics* (2nd ed., pp. 153–199). London: Elsevier.
- Ladefoged, P. & Broadbent, D. E. (1957). Information conveyed by vowels. *Journal of the Acoustical Society of America*, 29, 98–104.
- Liegeois-Chauvel, C., Musolino, A. & Chauvel, P. (1991). Localization of the primary auditory area in man. *Brain*, 114, 139–153.
- Lotto, A. J., Kluender, K. R. & Holt, L. L. (1997). Perceptual compensation for coarticulation by Japanese quail (*Coturnix coturnix japonica*). *Journal of the Acoustical Society of America*, 102(2), 1134–1140.
- Makela, A. M., Alku, P. & Tiitinen, H. (2003). The auditory N1m reveals the left-hemispheric representation of vowel identity in humans. *Neuroscience Letters*, 353(2), 111–114.
- May, B. J., Budelis, J. & Niparko, J. K. (2004). Behavioral studies of the olivocochlear efferent system – Learning to listen in noise. *Archives of Otolaryngology-Head & Neck Surgery*, 130(5), 660–664.
- McCarthy, G. & Donchin, E. (1981). A metric for thought – A comparison of P300 latency and reaction time. *Science*, 211(4477), 77–80.
- Mitterer, H. (2006). Is vowel normalization independent of lexical processing? *Phonetica*, 63(4), 209–229.
- Monahan, P. J. & Idsardi, W. J. (2010). Auditory sensitivity to formant ratios: Toward an account of vowel normalisation. *Language and Cognitive Processes*, 25(6), 808–839.
- Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., et al. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, 385(6615), 432–434.
- Näätänen, R. & Winkler, I. (1999). The concept of auditory stimulus representation in cognitive neuroscience. *Psychological Bulletin*, 125(6), 826–859.
- Nearey, T. M. (1989). Static, dynamic, and relational properties in vowel perception. *Journal of the Acoustical Society of America*, 85(5), 2088–2113.
- Oblers, J., Elbert, T., Lahiri, A. & Eulitz, C. (2003). Cortical representation of vowels reflects acoustic dissimilarity determined by formant frequencies. *Cognitive Brain Research*, 15(3), 207–213.
- Oblers, J., Eulitz, C. & Lahiri, A. (2004). Magnetic brain response mirrors extraction of phonological features from spoken vowels. *Journal of Cognitive Neuroscience*, 16(1), 31–39.
- Poeppel, D., Phillips, C., Yellin, E., Rowley, H. A., Roberts, T. P. L. & Marantz, A. (1997). Processing of vowels in supratemporal auditory cortex. *Neuroscience Letters*, 221(2–3), 145–148.
- Roberts, T. P. L., Flagg, E. J. & Gage, N. M. (2004). Vowel categorization induces departure of M100 latency from acoustic prediction. *Neuroreport*, 15(10), 1679–1682.
- Saarienen, J., Paavilainen, P., Schoger, E., Tervaniemi, M. & Näätänen, R. (1992). Representation of abstract attributes of auditory stimuli in the human brain. *Neuroreport*, 3(12), 1149–1151.
- Sharma, A. & Dorman, M. F. (2000). Neurophysiologic correlates of cross-language phonetic perception. *Journal of the Acoustical Society of America*, 107(5), 2697–2703.
- Sjerps, M. J., Mitterer, H. & McQueen, J. M. (2011). Constraints on the processes responsible for the extrinsic normalization of vowels. *Attention, Perception, & Psychophysics*, 73, 1195–1215.
- Snyder, E., Hillyard, S. A. & Galambos, R. (1980). Similarities and differences among the P3 waves to detected signals in 3 modalities. *Psychophysiology*, 17(2), 112–122.
- Stilp, C. E., Alexander, J. M., Kieft, M. J. & Kluender, K. R. (2010). Auditory color constancy: Calibration to reliable spectral properties across nonspeech context and targets. *Attention, Perception, & Psychophysics*, 72, 470–480.
- Summerfield, Q., Haggard, M., Foster, J. & Gray, S. (1984). Perceiving vowels from uniform spectra: Phonetic exploration of an auditory aftereffect. *Perception & Psychophysics*, 35(3), 203–213.
- Tavabi, K., Elling, L., Döbel, C., Pantev, C. & Zwitserlood, P. (2009). Effects of place of articulation changes on auditory neural activity: A magnetoencephalography study. *PLoS One*, 4(2), e4452.
- Tavabi, K., Oblers, J., Döbel, C. & Pantev, C. (2007). Auditory evoked fields differentially encode speech features: An MEG investigation of the P50m and N100m time courses during syllable processing. *European Journal of Neuroscience*, 25(10), 3155–3162.
- Tiitinen, H., Makela, A., Makinen, V., May, P. & Alku, P. (2005). Disentangling the effects of phonation and articulation: Hemispheric asymmetries in the auditory N1m response of the human brain. *BMC Neuroscience*, 6(1), 62.
- 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. *Psychological Science*, 21(10), 1532–1540.
- Van Nierop, D. J. P. J., Pols, L. C. W. & Plomp, R. (1973). Frequency analysis of Dutch vowels from 25 female speakers. *Acustica*, 29(2), 110–118.

- Watkins, A. J. (1991). Central, auditory mechanisms of perceptual compensation for spectral-envelope distortion. *Journal of the Acoustical Society of America*, *90*(6), 2942–2955.
- Watkins, A. J. & Makin, S. J. (1994). Perceptual compensation for speaker differences and for spectral-envelope distortion. *Journal of the Acoustical Society of America*, *96*(3), 1263–1282.
- Watkins, A. J. & Makin, S. J. (1996). Effects of spectral contrast on perceptual compensation for spectral-envelope distortion. *Journal of the Acoustical Society of America*, *99*(6), 3749–3757.
- Wilson, J. P. (1970). An auditory after-image. In R. P.omp, & G. F. Smoorenburg (Eds.), *Frequency analysis and periodicity detection in hearing* (pp. 303–318). Leiden: Sijthoff.
- Winkler, I., Lehtokoski, A., Alku, P., Vainio, M., Czigler, I., Csepe, V., et al. (1999). Pre-attentive detection of vowel contrasts utilizes both phonetic and auditory memory representations. *Cognitive Brain Research*, *7*(3), 357–369.