



Research paper

Auditory event-related responses to diphthongs in different attention conditions

David J. Morris^{a,*}, Kurt Steinmetzger^b, John Tøndering^a^a Department of Nordic Studies and Linguistics, University of Copenhagen, Copenhagen-S, Denmark^b Department of Speech, Hearing and Phonetic Sciences, University College London, UK

HIGHLIGHTS

- The direction of a diphthong yields differences in the P1-N1-P2 responses.
- N1-P2 responses to diphthongs occur even when attention is on a visual task.
- P2 vertex amplitude evoked with a diphthong varies with attention condition.

ARTICLE INFO

Article history:

Received 8 December 2015

Received in revised form 4 May 2016

Accepted 5 May 2016

Available online 6 May 2016

Keywords:

ERP

Attention

N1

P2

Acoustic change complex

Vowel perception

ABSTRACT

The modulation of auditory event-related potentials (ERP) by attention generally results in larger amplitudes when stimuli are attended. We measured the P1-N1-P2 acoustic change complex elicited with synthetic overt (second formant, $F2\Delta = 1000$ Hz) and subtle ($F2\Delta = 100$ Hz) diphthongs, while subjects (i) attended to the auditory stimuli, (ii) ignored the auditory stimuli and watched a film, and (iii) diverted their attention to a visual discrimination task. Responses elicited by diphthongs where $F2$ values rose and fell were found to be different and this precluded their combined analysis. Multivariate analysis of ERP components from the rising $F2$ changes showed main effects of attention on P2 amplitude and latency, and N1-P2 amplitude. P2 amplitude decreased by 40% between the attend and ignore conditions, and by 60% between the attend and divert conditions. The effect of diphthong magnitude was significant for components from a broader temporal window which included P1 latency and N1 amplitude. N1 latency did not vary between attention conditions, a finding that may be related to stimulation with a continuous vowel. These data show that a discernible P1-N1-P2 response can be observed to subtle vowel quality transitions, even when the attention of a subject is diverted to an unrelated visual task.

© 2016 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

As Simons and Chabris' invisible gorilla convincingly demonstrates [23], when we direct our attention to a specific occurrence, other events can go unnoticed. Functional descriptions of the modulation of perception by attention attribute a central role to salience filters that monitor perceptual input, and bias signals that regulate sensitivity control by priming neural populations in sensory cortices, see [13] for a review. Sensitivity and neural responsiveness is upregulated when a perceptual decision must be rendered,

and this may co-occur with downregulation of sensitivity for neural populations tuned to non-task relevant stimuli [10,27]. This complimentary co-regulation of top-down sensitivity is evident in cross-modal electrophysiology studies, where the introduction of visual stimuli typically results in a decrease in the amplitude of late auditory event-related potentials (ERPs). For example, the effect of simultaneously-performed tasks on auditory ERPs was investigated by Lavoie et al. [14] in presentation conditions which were: eyes open; eyes closed; reading a book; watching a movie, and; adding 3 digit numbers. N1-P2 amplitudes to click stimuli from 10 normal hearing listeners were reported to be lowest when arithmetic was performed concurrent with the presentation of auditory stimulus, and highest when subjects performed no task and had their eyes closed. Oray et al. [20] presented loud white noise tone bursts alone and also asynchronously paired with visual stimuli which were black and white vertical bars. They found mean decreases in amplitude between auditory alone and auditory-visual conditions

Abbreviations: dB HL, deciBel hearing level; EEG, electroencephalography; ERP, event-related potential; $F2$, second formant; MMN, mismatch negativity; RT, response time.

* Corresponding author at: Department of Nordic Studies and Linguistics, University of Copenhagen, Njalsgade 120, Copenhagen-S 2300, Denmark.

E-mail address: dmorris@hum.ku.dk (D.J. Morris).

that, when measured at the vertex, were 40% for N1 and 25% for P2. They attributed this attention-related reduction to a suppressive mechanism that benefited auditory perception when presented in the absence of visual stimuli.

The advantage that selective attention provides auditory perception has also been attributed to a temporal boost that promotes the neural processing of attended stimuli, an increase that may be indexed by N1 latency [21,26]. Folyi et al. [6] investigated this N1 latency acceleration, with ERP measurements to near-threshold tonebursts, in conditions where participants ignored the stimuli and watched a silent movie, or responded to stimulus onset. They found that N1 latencies decreased and amplitudes increased in the attend condition relative to the ignore condition. Effects of attention on auditory ERPs have also been investigated in mismatch negativity (MMN) studies. While the auditory MMN, elicited with fine frequency changes between standard and deviant tones, was found to decrease in amplitude when subjects sat passively or read a book [18], it was not affected by the level of difficulty of a simultaneously performed visual discrimination task [19].

In the present study we examined the P1-N1-P2 complex recorded in response to synthesized diphthongs, defined here as spectral changes that alter the quality of the vowel. It has been demonstrated that when recorded with speech stimuli, the P1-N1-P2 is stable over test sessions [28] and it can be elicited with synthetic speech [1,32], although component peaks can have shorter latencies relative to elicitation with natural speech stimuli [25]. The synthetic vowels that we used were presented to subjects in a stimulus paradigm akin to the ‘alternating strategy’ used in Martin et al. [17]. It has been suggested that measurement paradigms involving continuous acoustic stimulation with alternating changes in a-b-a-b form have the potential to benefit the electrical signal-noise ratio of the response, as the number of samples in the average ERP may be doubled by combining a-b and b-a transitions. We also included a subtle diphthong, in order to examine how the P1-N1-P2 response to fine-grained vowel quality transitions were affected by different attention conditions. In an auditory-visual presentation condition we instructed subjects to divert their attention from the concurrently presented diphthongs, to a cognitive task which was visual discrimination of Japanese symbols. The visual discrimination task required subjects to attend to the spatial detail of the symbols, in this way competing for attentional resources with the vowel formant transitions, a competition that we hypothesised may override ERP markers of subtle diphthongs.

2. Material and methods

2.1. Subjects

Sixteen healthy subjects who were university students and staff participated in this study. Data from four subjects were eliminated from analysis due to excessive noise in the responses. The remaining 12 subjects (6 female) had a mean age of 29 years 6 mos (SD 9 years 7 mos). All subjects reported that their right hand was dominant and they used this hand to provide responses during the test session. They all had normal vision and normal hearing as revealed by audiometric screening (pure tone thresholds less than 25 dB HL at 250–4000 Hz, in both ears) and all provided informed consent prior to the experiment. All subjects were native Danish speakers and none of the subjects had any prior knowledge of Japanese Kanji orthography, which was used in the visual task.

2.2. Stimuli

The auditory stimuli were two one-second and two two-second audio files that were vowels synthesized from a Klatt grid in PRAAT [3]. Overt diphthongs, where second formant (F2) values increased



Fig 1. Example of stimuli from the visual task. The three Japanese Kanji ideograms were presented on a computer screen where two were the same and one was different.

by 1000 Hz and subtle diphthongs, where F2 increased by 100 Hz occurred in the middle portion of the audio files, while all other acoustic parameters remained constant. The onset and offset F2 value was 1000 Hz which corresponded approximately to cardinal vowel 15 as described in [8, pp. 12–13] or [ɜ:]. The vowel quality after the subtle transition was slightly fronted, while after the overt transition, it corresponded approximately to cardinal vowel two. This vowel quality is closest to an [ɛ:] in Danish. A compressed version of the one second stimuli are available supplemental to this article as audio files 1 and 2, for stimuli with F2 transitions to 1100 and 2000 Hz, respectively. Fig. 2 shows the spectral details of the vowel and the transitions, which were 40 ms in duration. The four audio files were presented randomly so that they formed a continuous triphthongal chain (interstimulus interval = 0) within which vowel quality transitions were followed by steady-state portions of 460 and 960 ms. The sound field presentation level of the continuous stimuli was 70 dB(A) and this varied by less than 1 dB across the 3 vowel qualities. In total, 1440 diphthongs were presented to each participant and presentation conditions were counterbalanced across subjects.

2.3. Presentation conditions

Subjects underwent EEG testing in three conditions: attend, ignore and divert. During the attend condition they were instructed to respond to the auditory stimuli when F2 increased by 1000 Hz, that is, when the transition was from an [ɜ:] to an [ɛ:]. These rising overt diphthongs occurred irregularly due to the randomization of the one and two second stimulus files. In the ignore condition subjects watched a silent film. During the divert condition subjects performed a visual discrimination task involving three Japanese symbols, where two were the same and one was different (see Fig. 1). This was similar to the visual search task that was used in [2], where subjects were required to identify words from a grid of letters. However, we were hesitant about the possibility of interference between lexical items in combination with synthetic speech, and therefore created the stimuli for the visual task from five pairs of Kanji ideograms. Further motivation for the selection of this task in the divert condition is based on the assumption that it would require subjects to devote cognitive, particularly visuo-spatial, processing resources to the features of the Kanji symbols, but would only minimally task the subjects’ episodic working memory. Presentation of the visual trials was temporally independent of the diphthongal change, as we were interested in unmixing the crosstalk between the auditory ERP and the response from the visual task, an undertaking that would not be able to be accomplished if we had used sequential presentation of auditory and visual stimuli as, for instance, employed in [15].

During the divert condition subjects were instructed to devote their attention to the visual task and select as quickly and accurately as possible the symbol that was different by pressing the corresponding button on a response box. Prior to testing, subjects completed a training round with 12 visual trials in the absence of auditory stimuli, during which onscreen feedback to responses was provided as correct or incorrect. Feedback was not provided dur-

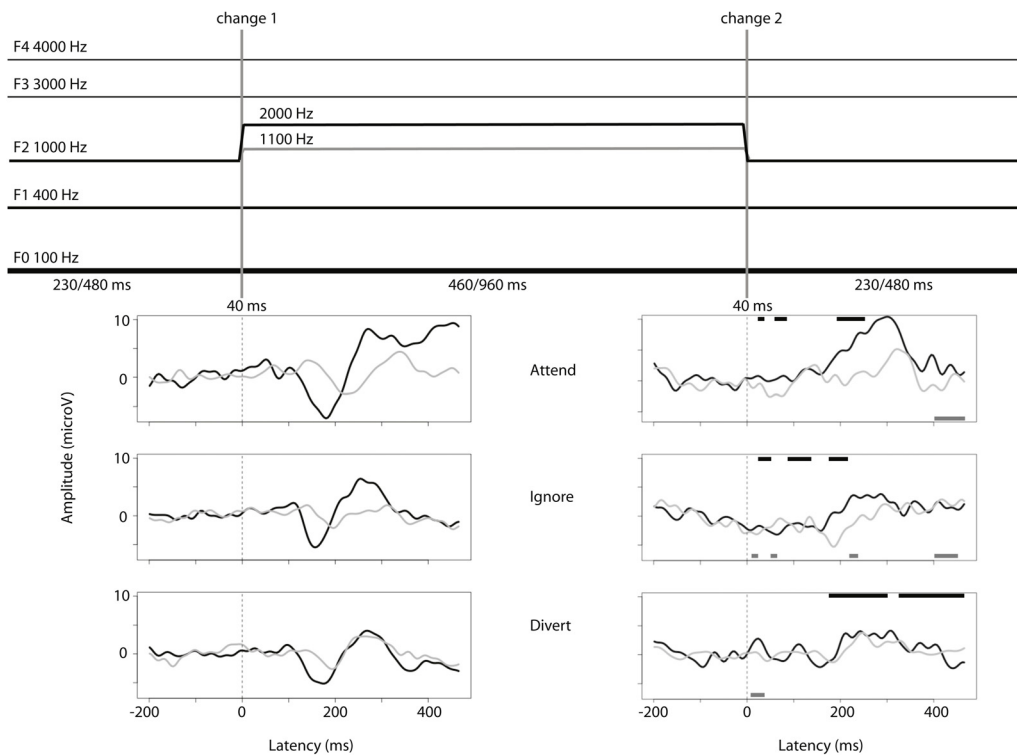


Fig. 2. Spectral and temporal details of the synthesized vowel stimuli (upper panel) aligned with group vertex (Cz) ERPs from the three attention conditions (lower panels). Bars (lower right panels) indicate latencies in which the first-rising F2 and second-falling F2 responses differed ($p < 0.05$), black overt ($F2\Delta = 1000$ Hz) and grey subtle ($F2\Delta = 100$ Hz) diphthongs.

ing the 360 randomized presentations that occurred during the ERP test session, where the intertrial interval was 0 and response times (RT) were measured.

2.4. EEG recording

Recordings were made in an electrically shielded and sound-treated room. Subjects were seated in an armchair with a speaker directly in front of them. To the right of the speaker was a laptop which was used to present the silent film in the ignore condition and the visual stimuli in the divert condition. Sintered Ag-AgCl electrodes were positioned at 16 scalp locations (O1, Oz, O2, P3, Pz, P4, C3, Cz, C4, F3, Fz, F4, Fp1, Fp2, T7, and T8) according to the extended international 10–20 system. Four additional electrodes were positioned around the eyes to monitor horizontal and vertical eye movements, and two electrodes were fixed to the left and right mastoids (M1 and M2). The Biosemi Active Two system was used to make EEG recordings at a sampling rate of 2048 Hz.

2.5. Data processing and statistical analysis

The continuous EEG data from all three conditions underwent identical processing. This involved down-sampling to 256 Hz and filtering between 0.1–30 Hz. The mean of the two mastoids was used as reference as these sites are below the Sylvian fissure and could therefore provide robust P1–N1–P2 responses [17,29]. Data were inspected visually and sections showing strong atypical artefacts were removed. Noisy channels were interpolated with the signal from surrounding electrodes and the continuous EEG data was submitted to Independent Component Analysis. Independent components, with topographic maps that were indicative of either eye movements or noisy channels, were removed from the data of each subject. The mean number of components removed was 1.6, SEM ± 0.2 . Epochs of -200 – 470 ms around the vowel changes were

then extracted and rejected if the electrical amplitude exceeded ± 50 μ V, or if linear trends exceeded 100 μ V within a 500 ms sliding window.

Permutation statistics were used to compare vertex responses in the full latency window to the first and second vowel quality changes. It is appropriate to use permutation statistics as this method preserves the pairing of data points, in our case from change 1 and 2, in the pooled data from which surrogate copies were drawn and from which the underlying distributions were generated. The comparison was based on 2000 surrogate data copies and all reported p values have been corrected for multiple comparisons with the false detection rate as implemented in EEGLAB [4]. After this initial pass we conducted parametric analysis of non-specific component data including P1, N1 and P2, from the vertex recording site. This was in order to observe the effects of attention and vowel quality transition at relevant points on the ERP time course and to facilitate comparison with previous studies including [6,17,24]. All processing and statistical analyses were performed with EEGLAB 13.4.4b, SPSS v. 22 and R [22].

3. Results

3.1. First and second vowel quality changes

Comparison of the responses from the first and second vowel quality changes in the three presentation conditions showed that there were significant differences and these occurred in both the subtle and the overt diphthongs. The latency windows in which the first and second changes were significantly different ($p < 0.05$), are shown in Fig. 2. A P3 (P300) peak can also be seen in the ERP from the first overt diphthong in the attend condition, as subjects had been instructed to respond to this transition. Due to these differences we did not collapse the ERP data from the first and second vowel

Table 1

First vowel quality transition: ANOVA results based on component data from the ERPs according to effect (df).

Component metric	Condition (2,60)		Diphthong (1,60)		Condition × Diphthong (2,60)	
	F	p	F	p	F	p
P1 amplitude	1.74	0.18	0.03	0.86	0.55	0.55
P1 latency	0.21	0.81	15.04	<0.01*	0.01	0.99
N1 amplitude	0.12	0.88	9.34	0.01*	0.12	0.12
N1 latency	0.03	0.97	0.54	0.46	0.82	0.44
P2 amplitude	5.89	<0.01*	9.89	<0.01*	2.43	0.09
P2 latency	4.49	0.01*	7.19	<0.01*	0.42	0.66
N1-P2 amplitude	4.43	0.01*	27.49	<0.01*	2.19	0.12
P1-N1 amplitude	0.57	0.57	13.02	<0.01*	0.52	0.59

Table 2

Group mean differences in ERP component metrics for the first change between the three conditions, with values expressed as percentages of *attend and #ignore responses.

	Attend-Ignore*		Attend-Divert*		Ignore-Divert#	
	1000Hz	100Hz	1000Hz	100Hz	1000Hz	100Hz
F2 Δ						
P1 amplitude	7.6	10.6	56.7	33.6	53.1	25.8
N1 amplitude	14.6	-11.5	1.3	5.6	-15.5	15.4
P2 amplitude	39.7	54.9	59.6	-15.0	32.9	-88.3
P1-N1	11.9	0.9	22.3	21.3	11.8	20.6
N1-P2	30.6	33.9	38.5	12.1	11.3	-33.0
P1 latency	0.8	2.4	-3.4	9.5	-4.2	7.3
N1 latency	8.1	-6.5	7.0	2.8	-1.1	8.8
P2 latency	11.1	6.4	9.5	11.4	-1.8	5.4

quality changes and subsequent analysis is based on the first change where there is an identifiable P1.

In order to examine the effects of both attention state and vowel formant transition on the first change, we quantified the peak amplitude and latency of ERP components from all subjects. Peak amplitudes were defined as the maximum value from baseline within the 50–150 ms and 200–350 ms post change latency windows for P1 and P2, respectively. The high upper limit of the latency window for P2 was chosen in order to include peaks from responses to the subtle vowel transition which, as shown in the group traces (Fig. 2), occurred later for the attend and ignore conditions. N1 was defined as the minimum value of the ERPs in the 100–200 ms post change latency window. P1-N1 and N1-P2 values were derived by subtracting the amplitudes of P1 from N1 and P2 from N1, respectively. These data were submitted to multivariate analysis with presentation condition, diphthong and their interaction as effects. Table 1 shows that the significant main effect of condition was concentrated at P2 (P2 amplitude and latency and N1-P2 amplitude). A significant main effect of diphthong, was found at P2 component metrics and also P1 latency, N1 amplitude and P1-N1 amplitude.

Bonferroni corrected post hoc analysis of the amplitudes of P2 and N1-P2 showed that there was a significant difference between the attend and ignore conditions (P2, $\mu = 2.25$, $p = 0.01$; N1-P2, $\mu = 2.44$, $p = 0.03$), and the attend and the divert conditions (P2, $\mu = 2.35$, $p = 0.01$; N1-P2, $\mu = 2.17$, $p = 0.05$), but not the ignore and divert conditions (P2, $\mu = 0.1$, $p = 1$; N1-P2, $\mu = -2.7$, $p = 1$). This analysis also showed that mean P2 latency was significantly longer in the attend than in the divert condition ($\mu = 31.61$, $p = 0.02$), but there were no significant differences between the attend and ignore ($\mu = 25.92$, $p = 0.07$) or the divert and ignore conditions ($\mu = -5.68$, $p = 1$).

As we were interested in the net effect of presentation condition on overt and particularly subtle vowel quality transitions, we calculated the percentage changes in peak amplitude and latency based on the first vowel quality change (see Table 2). These showed

considerable reductions in P2 and N1-P2 amplitudes between the attend and the other two conditions. Most ERP component amplitudes also decreased between the ignore and the divert conditions. However, P2 amplitude in response to subtle diphthongs increased from 1.5 μV in the ignore condition to 2.8 μV in the divert conditions, and over the same conditions N1-P2 amplitude increased from 3.2 μV to 4.3 μV .

3.2. Visual task responses

One subject completed all of the visual discrimination items in the course of the test session and ERP responses recorded from this subject in the period when the task was not performed were discarded from analysis. Analysis of the behavioural responses from the visual task showed that there were 143 trials with RTs lower than 200 ms. These were approximately 2.5% of the trials and they were removed from the data as this is shorter than visual uptake and response. Mean accuracy for all participants on the visual discrimination was 97.7% (SD 15) and the mean RT was 948 ms (SD 346 ms).

To examine the possibility that the RT data reflected the attention that subjects devoted to the task, we compared mean RTs from each subject to the individual difference in ERP amplitudes between the attend and the divert conditions across all vowel formant transitions. No significant correlations were found for the differences in mean P1 amplitude (Spearman's $\rho = 0.19$, $p = 0.5$), N1 amplitude (Spearman's $\rho = 0.3$, $p = 0.3$), P2 amplitude (Spearman's $\rho = 0.2$, $p = 0.5$), or N1-P2 amplitude ($\rho = 0.02$, $p = 0.9$).

4. Discussion

The present study replicates the effect whereby selective attention in a condition where subjects respond to stimuli, enhances the amplitude of auditory ERPs, while when subjects ignore the stimuli or divert their attention to a visual task, the amplitude is diminished. Vertex P2 response amplitude decreased by approximately 40% in the ignore condition, and by approximately 60% in the divert condition, relative to the attend condition. Furthermore, P2 latency was significantly longer in the attend condition relative to the divert condition, but there was no significant difference between the divert and ignore conditions for P2 amplitude and latency. In comparison to the effect of attention, the effect of diphthong was found to be significant for components that occur across a broader post change latency window, which included P1 and P2 latency, and N1, P2, N1-P2 and P1-N1 amplitude. The main effect of diphthong on components over a broader post change latency window may reflect the temporal jitter that occurs with near-threshold stimuli [26], such as a subtle change in vowel quality. However, even for the subtle diphthongs, landmark components of the late auditory ERP were not completely eliminated by simultaneous performance of the visual task in the divert condition. This indicates that in everyday listening scenarios, where attention may be focussed on visual tasks, like driving or operating a mobile telephone, there is a discernible neural change response to even fine-grained acoustic changes.

Significant differences in vertex N1 latency were not found in the ERP from the different attention conditions employed in the present study. Studies investigating the effect of attention on auditory neural measures have reported that attend conditions result in N1 amplitude enhancement [9,15] and N1 latency decreases in many [6,21,26] but not all [16] investigations. These previous reports indicate that N1 is an early and distinct marker of attention in the vertex ERP. Given the robust attentional effect on P2 amplitude that we observed, along with the absence of any effect on N1, our results suggest that this may not be so straightforward. It is

unclear why a significant attentional N1-latency decrease was not present in our data, however a factor that may have contributed to this was the continuous vowel stimuli which may not have allowed for sufficient refraction in neural populations responsible for N1 generation. Moreover, post hoc analyses of the P2 component data from the present study extend the findings of [19] and support the position that differences in the complexity of a simultaneously performed visual task do not have a significant influence on vertex components.

Presentation of the three Japanese symbols involved in the visual task in the divert condition occurred immediately after response selection from the preceding trial. In choosing a visual task with no interval between stimulus presentations, our intention was to constantly provide the subject with visual stimuli to which their attention would be diverted. Nonetheless, it may have been possible for subjects to switch their attention between auditory and visual stimuli. Accounts of the time course of attentional switching vary from 50 ms [31] to 500 ms [5]. Therefore, in the window between the presentation of visual stimulus and subsequent response, where the mean RT was approximately 1 s, subjects could conceivably have accomplished an attentional switch. However, as the continuous auditory stimuli was of constant amplitude and offered little novelty for the subject, it may be assumed that attentional switching was unlikely. The lack of correlation between response times on the visual task and decreases in component amplitudes between the attend and divert conditions does not support a capacity model of attention [11], where decreases in attention to the auditory stimuli would have been reflected in behavioural measures.

We report significant differences between ERPs from the first and second vowel quality changes in all conditions. The waveform morphology from the vertex electrode (see Fig. 2) shows that second vowel changes are generally smaller in amplitude and that they lack a distinct P1 component. A seemingly plausible explanation for this asymmetry is that one direction of diphthongal change may have been more familiar to the subjects. However, this is not likely because the Danish vowel inventory includes diphthongs in which the place of articulation changes between front and back in both directions [7]. That means that the differences in the ERP that we observed between the first and second changes are unlikely to be attributable to listener familiarity with one direction of the diphthong.

The asymmetry in the ERP to the first and second changes may instead be attributable to several, possibly related factors. Firstly, auditory-cortical responses may be preferentially sensitive to rising spectral change. This directional explanation is supported by the finding that N1-P2 responses are present when pure tone stimuli transition from a steady state to an upward glide, but are absent when the transition is from a downward glide to a steady state [30]. Similarly, late auditory evoked potentials to the onset and offset of noise have been shown to differ, with offset responses typically showing a bifid N1 component [12]. In our study, the first and second changes where the value of F2 increased and decreased, respectively, may be processed in a way that is similar to onset and offset responses. Secondly, if auditory-cortical changes are preferentially sensitive to rising diphthongs, then these, and only these, may have elicited a processing negativity that could have overlapped and interfered with the N1-P2 responses from the downward diphthongs, or the baseline of these. While this explanation is plausible given that the duration over which processing negativity extends may be up to a second, it does not explain the morphology of the waveform to the second changes during the ignore and divert conditions, where little or no negativity would be expected. Thirdly, the acoustic characteristics of the steady-state portion of the vowel may have played a role, as Agung MMN, mismatch negativity et al. [1] reported significant differences in the

peak latency of N1 recorded in response to naturally produced [i] and [u] vowels, which both had similar formant values to our overt diphthong. Regardless of the underlying cause, the asymmetry that we found in the responses to different directions of the diphthong is relevant to the adoption of speech stimuli in clinical settings. Specifically, the directional asymmetry in the P1-N1-P2 response can prevent the combined analysis of ERPs from both directions of a speech contrast, thus reducing efficiency in continuous acoustic stimulation paradigms.

Acknowledgement

David is supported by the Innovation Fund Denmark, J. Nr. 162-2013-6

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.neulet.2016.05.002>.

References

- [1] K. Agung, S.C. Purdy, C.M. McMahon, P. Newall, The use of cortical auditory evoked potentials to evaluate neural encoding of speech sounds in adults, *J. Am. Acad. Audiol.* 17 (2006) 559–572.
- [2] S.M. Berman, R. Heilweil, W. Ritter, J. Rosen, Channel probability and Nd: an event-related potential sign of attention strategies, *Biol. Psychol.* 29 (1989) 107–124.
- [3] P. Boersma, D. Weenink, Praat: Doing Phonetics by Computer [Computer Program]. Version 5.4.21, 2015, Retrieved from <http://www.praat.org/>.
- [4] A. Delorme, S. Makeig, EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis, *J. Neurosci. Methods* 134 (2004) 9–21.
- [5] J. Duncan, R. Ward, K. Shapiro, Direct measurement of attentional dwell time in human vision, *Nature* 369 (1994) 313–315.
- [6] T. Folyi, B. Fehér, J. Horváth, Stimulus-focused attention speeds up auditory processing, *Int. J. Psychophysiol.* 84 (2012) 155–163.
- [7] N. Grønnum, Fonetik og fonologi: almen og dansk [Phonetics and Phonology: General and Danish, Akad. Forl. (2005).
- [8] Handbook of the International Phonetic Association, A Guide to the Use of the International, International Phonetic Alphabet, CUP, Cambridge, 2003.
- [9] S.A. Hillyard, R.F. Hink, V.L. Schwent, T.W. Picton, Electrical signs of selective attention in the human brain, *Science* 182 (1973) 177–180.
- [10] J.A. Johnson, R.J. Zatorre, Neural substrates for dividing and focusing attention between simultaneous auditory and visual events, *Neuroimage* 31 (2006) 1673–1681.
- [11] D. Kahneman, Attention and effort, Prentice-Hall, N.J, 1973.
- [12] E.S. Kappenman, S.J. Luck, H. Pratt, Sensory ERP Components, OUP, Oxford, 2011.
- [13] E.I. Knudsen, Fundamental components of attention, *Annu. Rev. Neurosci.* 30 (2007) 57–78.
- [14] B.A. Lavoie, J.E. Hine, R.D. Thornton, The choice of distracting task can affect the quality of auditory evoked potentials recorded for clinical assessment, *Int. J. Audiol.* 47 (2008) 439–444.
- [15] Y.J. Luo, J.H. Wei, Cross-modal selective attention to visual and auditory stimuli modulates endogenous ERP components, *Brain Res.* 842 (1999) 30–38.
- [16] V. Mäkinen, P. May, H. Tiitinen, Transient brain responses predict the temporal dynamics of sound detection in humans, *Neuroimage* 21 (2004) 701–706.
- [17] B.A. Martin, A. Boothroyd, D. Ali, T. Leach-Berth, Stimulus presentation strategies for eliciting the acoustic change complex: increasing efficiency, *Ear Hear.* 31 (2010) 356.
- [18] A. Müller-Gass, R.M. Stelmack, K.B. Campbell, "...and were instructed to read a self-selected book while ignoring the auditory stimuli": the effects of task demands on the mismatch negativity, *Clin. Neurophysiol.* 116 (2005) 2142–2152.
- [19] A. Müller-Gass, R.M. Stelmack, K.B. Campbell, The effect of visual task difficulty and attentional direction on the detection of acoustic change as indexed by the mismatch negativity, *Brain Res.* 1078 (2006) 112–130.
- [20] S. Oray, Z.-L. Lu, M.E. Dawson, Modification of sudden onset auditory ERP by involuntary attention to visual stimuli, *Int. J. Psychophysiol.* 43 (2002) 213–224.
- [21] T.W. Picton, S.A. Hillyard, Human auditory evoked potentials. II. Effects of attention, *Electroencephalogr. Clin. Neurophysiol.* 36 (1974) 191–199.
- [22] R Development Core Team, R: A Language and Environment for Statistical Computing [Computer Software Manual], Vienna, Austria, 2014 (Retrieved from <http://www.r-project.org/>).

- [23] D.J. Simons, C. Chabris, Gorillas in our midst: sustained inattentive blindness for dynamic events, *Perception* 28 (1999) 1059–1074.
- [24] J.S. Snyder, C. Alain, T.W. Picton, Effects of attention on neuroelectric correlates of auditory stream segregation, *J. Cogn. Neurosci.* 18 (2006) 1–13.
- [25] S. Swink, A. Stuart, Auditory long latency responses to tonal and speech stimuli, *J. Speech Lang. Hear. Res.* 55 (2012) 447–459.
- [26] A.R.D. Thornton, M. Harmer, B.A. Lavoie, Selective attention increases the temporal precision of the auditory N 100 event-related potential, *Hear. Res.* 230 (2007) 73–79.
- [27] S. Treue, J.C.M. Trujillo, Feature-based attention influences motion processing gain in macaque visual cortex, *Nature* 399 (1999) 575–579.
- [28] K.L. Tremblay, L. Friesen, B.A. Martin, R. Wright, Test-retest reliability of cortical evoked potentials using naturally produced speech sounds, *Ear Hear.* 24 (2003) 225–232.
- [29] H.G. Vaughan Jr., W. Ritter, The sources of auditory evoked responses recorded from the human scalp, *Electroencephalogr. Clin. Neurophysiol.* 28 (1970) 360–367.
- [30] A. Weise, E. Schroger, B. Feher, T. Folyi, J. Horvath, Auditory event-related potentials reflect dedicated change detection activity for higher-order acoustic transitions, *Biol. Psychol.* 91 (2012) 142–149.
- [31] J.M. Wolfe, Guided search 2.0 a revised model of visual search, *Psychon. Bull. Rev.* 1 (1994) 202–238.
- [32] D.L. Woods, R. Elmasian, The habituation of event-related potentials to speech sounds and tones, *Electroencephalogr. and Clin. Neurophysiol./Evoked Potentials Sect.* 65 (1986) 447–459.