

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/316354693>

Neural indices of phonemic discrimination and sentence-level speech intelligibility in quiet and noise: A P3 study

Article in *Hearing research* · April 2017

DOI: 10.1016/j.heares.2017.04.009

CITATIONS

0

READS

121

5 authors, including:



Tess K. Koerner

University of Minnesota Twin Cities

15 PUBLICATIONS 56 CITATIONS

SEE PROFILE



Yang Zhang

University of Minnesota Twin Cities

70 PUBLICATIONS 712 CITATIONS

SEE PROFILE



Peggy Nelson

University of Minnesota Twin Cities

53 PUBLICATIONS 584 CITATIONS

SEE PROFILE



Boxiang Wang

University of Minnesota Twin Cities

6 PUBLICATIONS 11 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Cracking the speech code: A cross-linguistic neurobehavioral approach to language learning in typical and atypical populations. [View project](#)



Investigating neural mechanisms for speech and voice processing deficits and plasticity in children with autism [View project](#)



Research Paper

Neural indices of phonemic discrimination and sentence-level speech intelligibility in quiet and noise: A P3 study

Tess K. Koerner ^a, Yang Zhang ^{a, b, c, *}, Peggy B. Nelson ^{a, c}, Boxiang Wang ^d, Hui Zou ^d^a Department of Speech-Language-Hearing Sciences, University of Minnesota, Minneapolis, MN 55455, USA^b Center for Neurobehavioral Development, University of Minnesota, Minneapolis, MN 55455, USA^c Center for Applied Translational Sensory Science, University of Minnesota, Minneapolis, MN 55455, USA^d School of Statistics, University of Minnesota, Minneapolis, MN 55455, USA

ARTICLE INFO

Article history:

Received 7 December 2016

Received in revised form

12 April 2017

Accepted 16 April 2017

Available online 18 April 2017

Keywords:

Speech perception

Event-related potential

P3

Inter-trial phase coherence (ITPC)

Event-related spectral perturbation (ERSP)

ABSTRACT

This study examined how speech babble noise differentially affected the auditory P3 responses and the associated neural oscillatory activities for consonant and vowel discrimination in relation to segmental- and sentence-level speech perception in noise. The data were collected from 16 normal-hearing participants in a double-oddball paradigm that contained a consonant (/ba/ to /da/) and vowel (/ba/ to /bu/) change in quiet and noise (speech-babble background at a -3 dB signal-to-noise ratio) conditions. Time-frequency analysis was applied to obtain inter-trial phase coherence (ITPC) and event-related spectral perturbation (ERSP) measures in delta, theta, and alpha frequency bands for the P3 response. Behavioral measures included percent correct phoneme detection and reaction time as well as percent correct IEEE sentence recognition in quiet and in noise. Linear mixed-effects models were applied to determine possible brain-behavior correlates. A significant noise-induced reduction in P3 amplitude was found, accompanied by significantly longer P3 latency and decreases in ITPC across all frequency bands of interest. There was a differential effect of noise on consonant discrimination and vowel discrimination in both ERP and behavioral measures, such that noise impacted the detection of the consonant change more than the vowel change. The P3 amplitude and some of the ITPC and ERSP measures were significant predictors of speech perception at segmental- and sentence-levels across listening conditions and stimuli. These data demonstrate that the P3 response with its associated cortical oscillations represents a potential neurophysiological marker for speech perception in noise.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Measures of brain electrical activity have been important in investigating mechanisms that allow listeners to extract target signals from interfering background noise for successful speech communication. Previous auditory event-related potential (AERP) studies have demonstrated the detrimental effects of background noise on the timing and strength of neural responses to speech and non-speech stimuli (Billings et al., 2011; Koerner and Zhang, 2015; Parbery-Clark et al., 2011; Russo et al., 2009). Furthermore, the noise-induced changes in different AERP components have been shown to predict behavioral measures of perceptual and cognitive

abilities (Anderson et al., 2011; Anderson et al., 2010b; Billings et al., 2013; Billings et al., 2015; Koerner et al., 2016; Song et al., 2011). The present AERP study represents a sequel to our previous work (Koerner et al., 2016) to determine neural correlates of speech-in-noise perception at the syllable and sentence levels.

Of particular interest to the current report is the auditory P3 (or P300) response, which is thought to be sensitive to attentional and cognitive processes involved in auditory and speech perception (Picton, 1992; Polich, 2004). The P3 is typically elicited using an active-listening oddball paradigm, in which the subjects are instructed to respond when they detect an infrequent deviant stimulus (e.g., /ba/) within a string of repeated standard stimuli (e.g., /da/). Several studies have examined the effects of noise on the P3 response for segmental speech processing, including phonemic contrasts such as /ba/ vs /da/ and how various signal-to-noise ratios (SNRs) impacted the P3 response for speech discrimination (Kaplan-Neeman et al., 2006; Whiting et al., 1998). Two general

* Corresponding author. 115 Shevlin Hall, Department of Speech-Language-Hearing Sciences, 164 Pillsbury Drive SE, University of Minnesota, Minneapolis, MN 55455, USA.

E-mail address: zhanglab@umn.edu (Y. Zhang).

findings emerged from these studies. First, listening in noise resulted in significant increases in P3 latency and reductions in P3 amplitude for behaviorally discriminable speech stimuli, which were accompanied with increases in behavioral reaction time and reductions in accuracy. Second, Pearson correlation tests showed a negative correlation between sentence intelligibility scores and the P3 peak latency measures for detecting the phonemic (/ba/ vs. /da/) change as well as a positive correlation between behavioral reaction time measures of phonemic discrimination and the corresponding P3 latencies in the normal-hearing adult listeners (Bennett et al., 2012). Thus the P3 response for phonemic discrimination appears to be a potential neurophysiological marker for speech-in-noise perception at both segmental and sentential levels. However, there was a limitation in the experimental design of the previous studies as they all focused on consonantal change detection alone. It remains to be determined how the P3 responses to consonant discrimination differ from those to vowel discrimination and how well the P3 for vowel contrasts can predict speech perception in noise.

Our work was motivated by the fact that vowels and consonants in a spoken language elicit different patterns of behavioral and neural responses. At the subcortical level, frequency following response (FFR) measures showed larger noise-induced effects on consonant encoding than vowel encoding in the CV syllable context (Anderson et al., 2010a, 2010b; Russo et al., 2004; Song et al., 2011), which is consistent with the behavioral finding of larger noise-induced impact on the detection of consonants compared to vowels (Korczak and Stapells, 2010; Miller and Nicely, 1955; Parikh and Loizou, 2005; Phatak and Allen, 2007; Pickett, 1957). At the cortical level, however, recent studies showed that the neural coding of vowel contrasts as reflected in the mismatch negativity (MMN) response was more susceptible to the presence of background noise compared to consonants (Koerner et al., 2016; Niemitalo-Haapola et al., 2015). Unlike the P3 response that requires an overt behavioral response for detecting a stimulus change and peaks at a later time point than the MMN, the MMN is thought to index pre-attentive automatic change detection independent of focused attention (Näätänen et al., 2007). In our previous MMN study (Koerner et al., 2016), we argued that the differential effects of noise on the MMNs for consonant discrimination and vowel discrimination reflected their different contributions to speech intelligibility in noise as evidenced in multiple behavioral studies (Cole et al., 1996; Fogerty and Humes, 2012, 2010; Fogerty and Kewley-port, 2009; Kewley-Port et al., 2007; Owren and Cardillo, 2006). On this point, there is an abundance of neuroimaging data indicating that distinct brain mechanisms are involved in processing consonants and vowels (Caramazza et al., 2000; Carreiras et al., 2009; Carreiras and Price, 2008).

To date, previous speech-in-noise research studies have not directly compared cortical processing of consonant discrimination and vowel discrimination in the presence or absence of background noise using the attention-driven P3 response. The current P3 study followed up our previous MMN work with a double-oddball paradigm (Koerner et al., 2016) in which two deviants consisting of either a consonant change (from /ba/ to /da/) or a vowel change (from /ba/ to /bu/) were presented within the same listening session. We investigated the role of attention in the neural processing of consonant and vowel stimuli by examining the differential effects of background noise on neural responses as listeners actively detected each of these stimulus contrasts. We were particularly interested in examining how the segmental-level P3 responses for consonant and vowel stimuli jointly or separately contributed to sentence-level performance.

In addition to conventional analysis on the latency and amplitude of AERP components, researchers have also begun to use time-

frequency analysis techniques to determine how experimental stimulus and task factors impact induced and evoked cortical oscillations within the ongoing EEG signal. The oscillations are thought to play a key role in enabling sensory and cognitive processing across and within cortical networks (Başar et al., 1999; Klimesch et al., 2007; Koerner and Zhang, 2015; Makeig et al., 2004; Sauseng et al., 2007; Zhang et al., 2011). Specifically, oscillations in the delta (1–4 Hz), theta (4–8 Hz), and alpha (8–12 Hz) frequency bands have been found to be associated with the cortical P3 response, which may represent underlying cognitive demands related to different processes of signal processing and attentional engagement (Demiralp et al., 2001; Intriligator and Polich, 1994, 1995; Kolev et al., 1997; Polich, 1997; Spencer and Polich, 1999; Yordanova and Kolev, 1998). For example, Basar-Eroglu et al. (1992) examined the functional significance of delta and theta oscillations using two auditory P3 paradigms: an omitted stimulus paradigm, which required attention to every third signal, and an oddball paradigm, which required additional signal matching and decision making processes to respond to rare, randomly presented target stimuli. From their results, it was suggested that theta oscillations, which were altered across both tasks, are functionally related to focused attention and signal detection processes. On the other hand, delta band activity only increased during the oddball paradigm, suggesting delta oscillations are more specifically related to signal matching and decision making processes. Furthermore, event-related oscillations in the alpha band have been shown to reflect top-down processing resources that are important for inhibiting task irrelevant maskers, such as during speech perception in complex listening environments (Klimesch, 2012; Straub et al., 2014; Wilsch et al., 2014). Together, the existing literature shows the importance of time-frequency analysis to specify delta, theta, and alpha oscillatory activities underlying the auditory P3 responses for consonant discrimination and vowel discrimination that are differentially affected by the background noise. For the current study, we will examine event-related cortical oscillations for the P3 responses in terms of neural synchrony and EEG spectral power across trials.

We hypothesized that background noise would significantly impact the P3 response as well as oscillatory activities in delta, theta, and alpha frequency bands in response to both consonant and vowel contrasts. Based on the behavioral data from our previous study (Koerner et al., 2016), we expected that background noise would have a larger effect on the detection of the consonant change compared to the vowel change in the attentive listening condition. Furthermore, we were particularly interested in investigating whether the AERP and time-frequency measures would be significant predictors of behavioral performance at both the segmental and sentence levels. The results from the current P3 study with an active listening condition would complement those from our previous MMN work that did not require focused attention to detect the consonant and vowel contrasts in the double-oddball paradigm (Koerner et al., 2016).

2. Methods

2.1. Subjects

The participants were 16 individuals (mean age = 22.5 years, age range = 19–32 years, 6 males, 10 females) with normal hearing (as shown in standard audiological assessment with hearing thresholds < 25 dB HL for pure tones from 0.25 to 8 kHz). All participants were right handed native speakers of American English, and had no history for speech, language, or cognitive disorders or difficulties. The Human Research Protection Program at the University of Minnesota approved the research protocol, and all

participants provided informed consent prior to beginning the study.

2.2. Stimuli

Three consonant-vowel (CV) syllables, /ba/, /da/, and /bu/, were synthesized using a 10 kHz sampling rate in the Hlsyn software program (Sensimetrics Corporation, USA) (Koerner and Zhang, 2015). Each syllable was 170 ms in duration with a steady fundamental frequency of 100 Hz and a steady F4 of 3300 Hz. For the /ba/ sound, the Hlsyn software generated formant transitions in the first 50 ms of the CV syllables with onset frequencies at 328 Hz, 1071 Hz, and 2298 Hz respectively for F1, F2, and F3. The F1, F2, and F3 onset frequencies were set at 362 Hz, 1832 Hz, and 2540 Hz for /da/ and 230 Hz, 900 Hz, and 2480 Hz for /bu/ respectively. For the vowel portion (50–170 ms) of the /ba/ and /da/ syllables, the steady center F1, F2, and F3 frequencies were 674 Hz, 1140 Hz, and 2350 Hz. The steady center F1, F2, and F3 frequencies were 320 Hz, 860 Hz, and 2620 Hz for the vowel portion of /bu/. IEEE sentences (IEEE, 1969) were also used to obtain sentence recognition scores. We used a four-talker speech babble background noise that was adopted from the Quick Speech In Noise Test (Quick-SIN) (Niquette et al., 2001). The speech and noise stimuli were resampled at 44.1 kHz and were normalized to create a –3 dB SNR using Sony SoundForge 9.0 (Sony Creative Software, USA).

2.3. Procedure

The reported EEG data were taken from a larger scale study with a 2-h recording session including both passive and active listening conditions (Koerner et al., 2013). All ERP and behavioral test sessions were conducted in an electrically and acoustically treated booth (ETS-Lindgren Acoustic Systems). The CV syllables were presented via bilateral Etymotic ER-2 insert headphones using Evoke software (ANT Inc., Netherlands). The speech signals were presented at 60 dB SL relative to the individual listeners' hearing threshold at 1 kHz (Koerner et al., 2016; Koerner and Zhang, 2015; Nie et al., 2014; Rao et al., 2017). Participants were presented with two listening conditions in both the EEG and behavioral tests: signals in quiet and signals in a four-talker speech babble noise.

As described in our previous MMN study (Koerner et al., 2016), the double-oddball paradigm included two speech contrasts representing a vowel change (from /ba/ to /bu/) and a consonant change (from /ba/ to /da/) within the same recording session. Unlike the passive listening MMN protocol, the listeners in the current P3 study were asked to press a keyboard response button each time they heard a deviant stimulus representing either a consonant or vowel change. The back vowels /a/ and /u/ differ primarily in the steady vowel F1 while the /b/ and /d/ consonants represent transient differences in place of articulation and are cued by second and third formant frequency transitions, which are in a frequency range commonly affected by hearing impairment (Ladefoged, 2006; Miller and Nicely, 1955). The order of stimulus presentation was pseudo-randomized in the double-oddball paradigm so that no blocks began with a deviant stimulus and no two deviants were presented in succession. The standard stimulus /ba/ had a probability of occurrence of 0.75 and both /da/ and /bu/ each had a probability of occurrence of 0.125. Each listening condition consisted of 10 blocks for a total of 832 trials for standard stimuli and 104 trials for each deviant stimulus. Participants were given 10–15 s breaks between blocks. The interstimulus interval (ISI) was 1000 ms with a 100 ms randomization from trial to trial. The quiet and noise listening conditions were counter-balanced across subjects to reduce potential test order effects. During the recording

session, participants were instructed to relax and minimize excessive movements.

Sentence recognition performance was recorded in an additional 30-min behavioral session using randomized lists of IEEE sentences presented through TDH-39 headphones. The IEEE lists consisted of 10 low context sentences with 5 key words in each sentence. For both the quiet and noise conditions, two sentence lists of 50 key words were presented from one female and one male talker for a total of 100 key words per listening condition (IEEE, 1969). Participants were instructed to repeat the IEEE sentences they heard out loud, as best as they could, while a certified audiologist evaluated the word-by-word responses for recognition accuracy.

2.4. Data analysis

2.4.1. ERP measures

Continuous EEG data were recorded using the Advanced Neuro Technology EEG System (Advanced Source Analysis version 4.7) and a 64-channel Ag AgCl electrode WaveGuard cap with a REFA-72 amplifier (TMS International BV) (bandwidth = 0.016–200 Hz, sampling rate = 512 Hz). The average impedance of electrodes was below 5 kOhms. ERP waveform analysis was completed offline in BESA (Version 6.0, MEGIS Software GmbH, Germany). The EEG data were bandpassed at 0.5–40 Hz. The ERP epoch length consisted of a 100 ms prestimulus baseline and a 700 ms poststimulus interval. Automatic artifact rejection criteria were set at $\pm 50 \mu\text{V}$. The P3 was analyzed with an averaged mastoid reference at the Pz electrode. The time window for assessing P3 peak latency was 250–680 ms, which was based on the grand average waveforms in the quiet and noise conditions relative to the pre-stimulus interval. Computation of the P3 amplitude used an integration window of 40 ms centered at peak. Similar time windows for P3 quantification were used in our previous P3 studies (Nie et al., 2014; Rao et al., 2017).

Two event-related time frequency analysis measures were computed to evaluate trial-by-trial cortical oscillations in delta, theta, and alpha frequency bands: inter-trial phase coherence (ITPC) and event-related spectral perturbation (ERSP). Inter-trial phase coherence evaluates the synchronization of trial-by-trial oscillations as a function of time and frequency (David et al., 2006; Makeig et al., 2004; Nash-Kille and Sharma, 2014):

$$ITPC_{(t,f)} = \frac{1}{n} \sum_{k=1}^n \frac{F_k(t,f)}{|F_k(t,f)|},$$

where F stands for the Fourier transform, t stands for time, f is frequency, k is the trial number, n is the total number of accepted trials, and $||$ is the complex norm (Delorme and Makeig, 2004). Resulting ITPC values in a given frequency band can range from 0, which represents no synchronization across trials, to 1, which represents perfect synchronization across trials. Event-related spectral perturbation measures trial-by-trial change in evoked power (in dB) from pre-stimulus baseline as a function of frequency and time (Fuentemilla et al., 2006; Makeig, 1993):

$$ERSP_{(f,t)} = \frac{1}{n} \sum_{k=1}^n |F_k(f,t)|^2,$$

where F stands for the Fourier transform, t is time, f is frequency, k is the trial number, and n is the total number of accepted trials (Delorme and Makeig, 2004).

Analysis of ITPC and ERSP at electrode Pz was completed using the *newtimef* function in EEGLAB (Delorme and Makeig, 2004). A short-term Fourier Transform (STFT) with Hanning window tapering (Koerner and Zhang, 2015; Koerner et al., 2016) was adopted to extract the ITPC and ERSP values for the delta, theta, and alpha frequency bands, which is recommended for the analysis of low frequency activities. In order to overcome restrictions from the use of fixed windows in conventional analysis, the modified STFT method used overlapping sliding windows adapted to the target

frequency bins, and zero-padding was applied for short epochs that did not have sufficient sample points for the Fourier transform. Estimated frequencies were from 0.5 to 40 Hz with a step interval of 0.5 Hz. The analysis time windows for both deviant stimuli were 300–400 ms for the quiet listening condition and 400–500 ms for the listening condition with background noise. The ITPC and ERSP values represented the maximum in the defined time windows, which were chosen based on peak latency data from individual P3 waveforms.

The statistical analyses from both the AERP and behavioral portions of the study were analyzed in R (R Core Team, 2014). A repeated measures analysis of variance (RM-ANOVA), with $\alpha = 0.05$, was conducted to examine the statistical significance of stimulus type (/da/ or /bu/), listening condition (quiet or noise), and any potential interactions (stimulus x condition) on P3 latency, amplitude, ITPC, and ERSP at electrode Pz. Where significant interaction effects were observed, tests of simple main effects and post hoc two-tailed t-tests for selected factors of interest were also conducted to evaluate how consonant- and vowel-changes were processed differently in quiet and in noise.

2.4.2. Behavioral measures

Percent correct change discrimination scores and reaction time for the detection of consonant- and vowel-changes in the double-oddball paradigm were obtained from the button-press responses recorded during the quiet and noise conditions. A RM-ANOVA was completed to investigate the statistical significance of stimulus type (/da/ or /bu/), listening condition (quiet or noise), and any potential interactions (stimulus x condition) on behavioral discrimination accuracy and reaction time. A RM-ANOVA was carried out to examine the significance of listening condition (quiet vs. noise) on IEEG sentence recognition.

2.4.3. Analysis of brain-behavior relationships

Linear mixed-effects (LME) regression models were developed in R (R Core Team) using the *nlme* package (Pinheiro et al., 2016) to examine whether neural measures were predictive of behavioral performance at the segmental and sentence levels across participants, listening conditions, and stimuli (Koerner et al., 2016; Koerner and Zhang, 2017). Similar statistical techniques have been used previously to examine links between neurophysiological and behavioral measures of speech perception (Billings et al., 2013, 2015). These models allow for an examination of multiple neural measures as predictor variables, or fixed effects, on a particular outcome variable while taking into account repeated measures across participants (Baayen et al., 2008; Bagiella et al., 2000; Magezi, 2015).

Linear mixed-effects models were created with by-participant intercept as a random effect. Listening condition (quiet vs. noise) and stimulus (/bu/ vs. /da/) were added as blocking variables. P3 latency and amplitude as well as ITPC and ERSP in delta, theta, and alpha bands were added as fixed effects. Outcome variables included phoneme change detection performance, reaction time, and sentence-level recognition. Final reduced models with fewer terms were chosen with the Akaike's Information Criterion (AIC) using stepwise linear regression with bidirectional elimination on main effects. The models we used to examine percent correct sentence recognition did not contain stimulus type as a fixed effect because sentence recognition performance only varied by listening condition. Data transformations included re-scaling P3 latency and behavioral reaction time values as well as log-transforming percent correct phoneme detection and sentence recognition scores to account for skewness in the data. For each LME model used in this study, the significance of each variable in predicting behavioral performance was assessed with $\alpha = 0.05$.

3. Results

Analysis of phoneme- and sentence-level behavioral data (Table 1 and Fig. 4 in supplemental materials) showed that background noise had a significant impact on percent correct phoneme change detection and reaction time as well as percent correct sentence recognition. Background noise also impacted P3 latency and amplitude (Fig. 1, Table 2, and Fig. 5 in supplemental materials) as well as trial-by-trial neural synchrony (Fig. 2, Table 3, and Fig. 6 in supplemental materials). Several significant neural markers of behavioral speech perception at both the phoneme- and sentence-levels were revealed using stepwise linear-mixed effects regression models (Table 4).

3.1. Percent correct phoneme discrimination and sentence recognition

As expected, RM-ANOVA revealed a significant detrimental

Table 1

Means (standard error) for behavioral percent correct detection of vowel (/bu/) and consonant (/da/) changes, behavioral reaction time for vowel (/bu/) and consonant (/da/) changes, as well as percent correct behavioral sentence recognition performance.

Behavioral measure	Listening condition	
	Quiet	Noise
/bu/ detection (%)	97.77 (0.71)	91.35 (2.12)
/da/ detection (%)	91.57 (1.89)	33.35 (3.84)
Sentence recognition (%)	99.31 (0.35)	67.25 (3.53)
/bu/ response time (ms)	468.94 (9.41)	553.06 (15.82)
/da/ response time (ms)	513.01 (11.97)	622.91 (11.36)

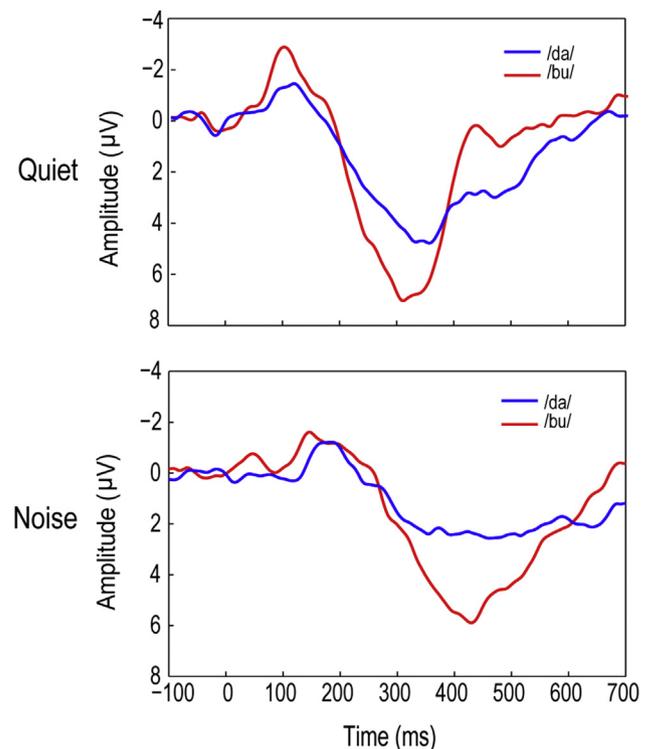


Fig. 1. Grand mean P3 waveforms at the Pz electrode depicting responses averaged across participants to the consonant-vowel syllable /bu/ (red) and /da/ (blue) in the quiet (top row) and speech-babble (bottom row) listening conditions. An analysis window of 250–680 ms was used to extract individual P3 latencies. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2

Mean (standard error) values for P3 latency (ms) and amplitude (μV) in response to the CV syllable vowel change (/bu/) and consonant change (/da/) at electrode Pz in quiet and in noise.

	Latency (ms)		Amplitude (μV)	
	Quiet	Noise	Quiet	Noise
/bu/	362.16 (27.34)	459.06 (20.6)	7.64 (0.69)	6.29 (0.77)
/da/	399.29 (27.33)	480.54 (31.61)	5.97 (0.6)	3.31 (0.41)

effect of background noise on behavioral reaction time ($F(1,15) = 47.12, p < 0.001$) and percent correct phoneme detection ($F(1,15) = 94.41, p < 0.001$) (Table 1). More specifically, behavioral reaction time was significantly prolonged and percent correct phoneme detection was significantly reduced in background noise compared to the quiet listening condition. RM-ANOVA also revealed a significant effect of stimulus on behavioral reaction time ($F(1,15) = 28.77, p < 0.001$) and percent correct phoneme detection ($F(1,15) = 147.40, p < 0.001$), such that reaction time was significantly longer and percent correct phoneme detection significantly poorer during behavioral discrimination of the consonant change (from /ba/ to /da/) compared to the vowel change (from /ba/ to /bu/). There was also a significant effect of background noise on sentence-level recognition ($F(1,15) = 83.44, p < 0.001$).

Additionally, RM-ANOVA revealed a significant interaction between listening condition and stimulus for percent correct phoneme detection ($F(1,15) = 20.94, p < 0.001$). Post-hoc t-tests

indicated that the presence of background noise significantly decreased percent correct detection of the consonant change ($t(1,15) = 16.35, p < 0.001$) and had a smaller, yet still significant effect on detection of the vowel change ($t(1,15) = 3.61, p < 0.01$).

3.2. Averaged P3 amplitude and latency measures

As predicted, RM-ANOVA revealed significant effects of background noise on P3 latency ($F(1,15) = 15.85, p < 0.01$) and amplitude ($F(1,15) = 23.5, p < 0.001$) (Table 2 and Fig. 1). The neural responses to the two deviant speech stimuli in quiet tended to have shorter latencies and larger amplitudes than responses in background noise. There was also a significant effect of stimulus on P3 amplitude across the quiet and noise listening conditions ($F(1,15) = 21.7, p < 0.001$), such that P3 amplitude was smaller in response to /da/ compared to /bu/. However, there was no significant difference in P3 latency between responses to /da/ and /bu/ ($F(1,15) = 2.60, p = 0.128$) across listening conditions. Additionally, there were no significant interactions between listening condition and stimulus for P3 latency ($F(1,15) = 0.22, p = 0.64$) or amplitude ($F(1,15) = 1.48, p = 0.24$).

3.3. ITPC and ERSP

RM-ANOVA indicated a significant effect of noise on ITPC in the delta ($F(1,15) = 7.68, p < 0.05$), theta ($F(1,15) = 12.02, p < 0.01$), and alpha ($F(1,15) = 11.03, p < 0.01$) frequency bands across the two

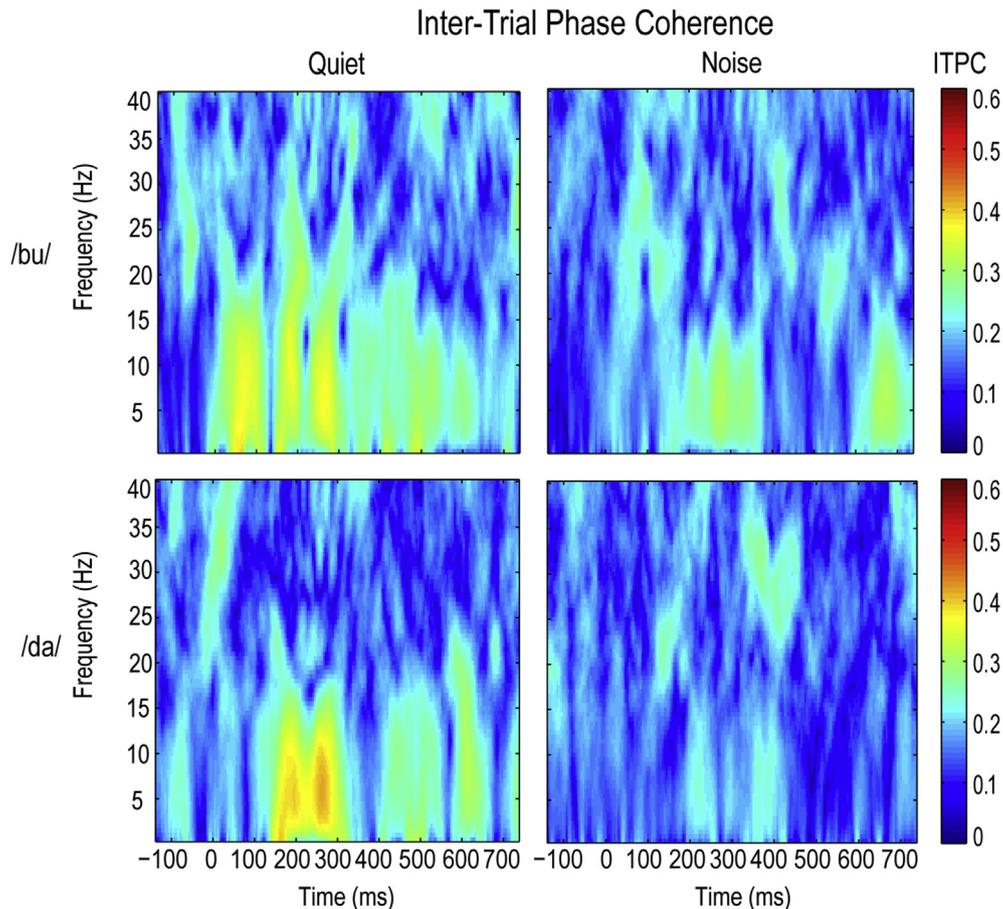


Fig. 2. Grand mean inter-trial phase coherence (ITPC) data in both quiet (left column) and babble noise (right column) listening conditions. ITPC values range from 0 to 1 for change-detection responses to the CV syllables /bu/ (top row) and /da/ (bottom row). ITPC was calculated using an analysis window of 300–400 ms for the quiet listening condition and a window of 400–500 ms for the listening condition with background noise.

Table 3

Mean (standard error) ITPC and ERSP values in response to the CV syllable vowel change (/bu/) and consonant change (/da/) at electrode Pz in quiet and in noise in delta, theta, and alpha frequency bands.

	Delta		Theta		Alpha	
	Quiet	Noise	Quiet	Noise	Quiet	Noise
ITPC						
/bu/	0.48 (0.02)	0.4 (0.02)	0.47 (0.02)	0.38 (0.02)	0.47 (0.02)	0.37 (0.02)
/da/	0.46 (0.02)	0.4 (0.01)	0.44 (0.02)	0.39 (0.02)	0.44 (0.02)	0.39 (0.02)
ERSP						
/bu/	0.63 (0.51)	0.45 (0.52)	0.68 (0.57)	0.39 (0.5)	0.53 (0.53)	0.24 (0.47)
/da/	0.79 (0.49)	0.74 (0.40)	0.73 (0.51)	0.56 (0.51)	0.58 (0.49)	0.41 (0.37)

Table 4

F-statistics for fixed effects (P3 latency, amplitude, ITPC and ERSP in delta, theta, and alpha bands) included in final reduced linear mixed-effects regression models for each behavioral measure.

Variable	Percent correct phoneme detection	Phoneme detection reaction time	Percent correct sentence recognition (/bu/)	Percent correct sentence recognition (/da/)
Intercept	269.97***	5439.97***	295.46***	398.83***
Condition	115.72***	92.89***	667.05***	398.74***
Stimulus	104.55***	32.02***	–	–
Latency	–	–	–	–
Amplitude	7.01*	11.03**	–	–
Delta ITPC	–	–	2.06	3.06
Theta ITPC	2.93.	–	9.69**	–
Alpha ITPC	–	–	4.36.	–
Delta ERSP	–	–	–	1.96
Theta ERSP	0.34	0.87	–	–
Alpha ERSP	8.29**	4.86*	–	3.53

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

deviant syllables, /da/ and /bu/ (Table 3 and Fig. 2). In contrast, there was no significant effect of background noise on ERSP in the delta ($F(1,15) = 0.11$, $p = 0.75$), theta ($F(1,15) = 0.53$, $p = 0.48$), or alpha ($F(1,15) = 0.61$, $p = 0.45$) frequency bands due to the existence of very large inter-subject variability of across-trial EEG spectral power for the P3 component relative to the baseline (Table 3 and Fig. 3). There were also no significant effects of stimulus on ITPC for delta ($F(1,15) = 0.14$, $p = 0.71$), theta ($F(1,15) = 0.21$, $p = 0.65$), or alpha ($F(1,15) = 0.06$, $p = 0.81$) frequency bands or on ERSP in delta ($F(1,15) = 0.51$, $p = 0.48$), theta ($F(1,15) = 0.13$, $p = 0.72$), or alpha ($F(1,15) = 0.15$, $p = 0.70$) frequency bands.

3.4. Linear mixed-effects regression models

Stepwise linear regression using bidirectional elimination by AIC was used to obtain reduced linear mixed-effects regression models with fewer terms. Table 4 shows a summary of reduced model outputs for segmental-level and sentence-level behavioral measures. The linear mixed-effects regression models for predicting segmental-level performance revealed that P3 amplitude was a significant predictor of percent correct segmental-level change detection ($F(1,42) = 7.01$, $p < 0.05$) and reaction time ($F(1,43) = 11.03$, $p < 0.01$) across listening conditions and stimuli. Additionally, ERSP in the alpha band was a significant predictor of percent correct segmental-level change detection ($F(1,42) = 8.29$, $p < 0.01$) and reaction time ($F(1,43) = 4.86$, $p < 0.05$) across listening conditions and stimuli. ITPC in the theta band showed a similar but insignificant trend as a predictor of percent correct segmental-level change detection ($F(1,12) = 2.93$, $p < 0.10$). The models for predicting sentence-level performance revealed that ITPC in the theta frequency band ($F(1,12) = 9.69$, $p < 0.01$) in response to /bu/ was a significant predictor of percent correct sentence-level performance across listening conditions. Although ITPC in the alpha band showed a similar trend, it did not reach statistical significance ($F(1,12) = 4.36$, $p < 0.10$) for being a predictor of sentence-level performance. There was also a similar but

insignificant trend in ERSP of the alpha frequency band in response to /da/ as a predictor of percent correct sentence-level performance across listening conditions ($F(1,12) = 3.53$, $p < 0.10$). The residual plots from all linear mixed-effects models appeared normally distributed and did not show any signs of heteroscedastic variance or significant trends.

4. Discussion

This ERP study was designed to examine the effects of background noise on P3 latency, amplitude, ITPC, and ERSP in response to a consonant and vowel change and also aimed to determine whether these neural measures could be significant neurophysiological predictors of segmental- and sentence-level speech perception across quiet and noise listening conditions.

4.1. Attention and differential effects of noise on neural coding of consonants and vowels

Our neurophysiological results are consistent with previous reports of noise-induced increases in P3 latency and decreases in P3 amplitude in response to speech (Bennett et al., 2012; Martin et al., 1997; Martin and Stapells, 2005; Whiting et al., 1998). Our data add to the literature by revealing a differential effect of background noise on the attentive cortical processing of consonant and vowel stimuli, such that background noise had a larger impact on P3 responses to the consonant change compared to the vowel change, which is consistent with results from our behavioral phoneme-change detection data. However, this pattern of noise-induced effects on the neural coding of consonant and vowel changes appears to be opposite to previous reports from the cortical, pre-attentive MMN response to consonants and vowels in noise (Koerner et al., 2016; Niemitalo-Haapola et al., 2015). Using the same consonant and vowel changes as in the current study, Koerner et al. (2016) showed that background noise actually had a larger impact on MMN amplitude and EEG theta power in response to the vowel

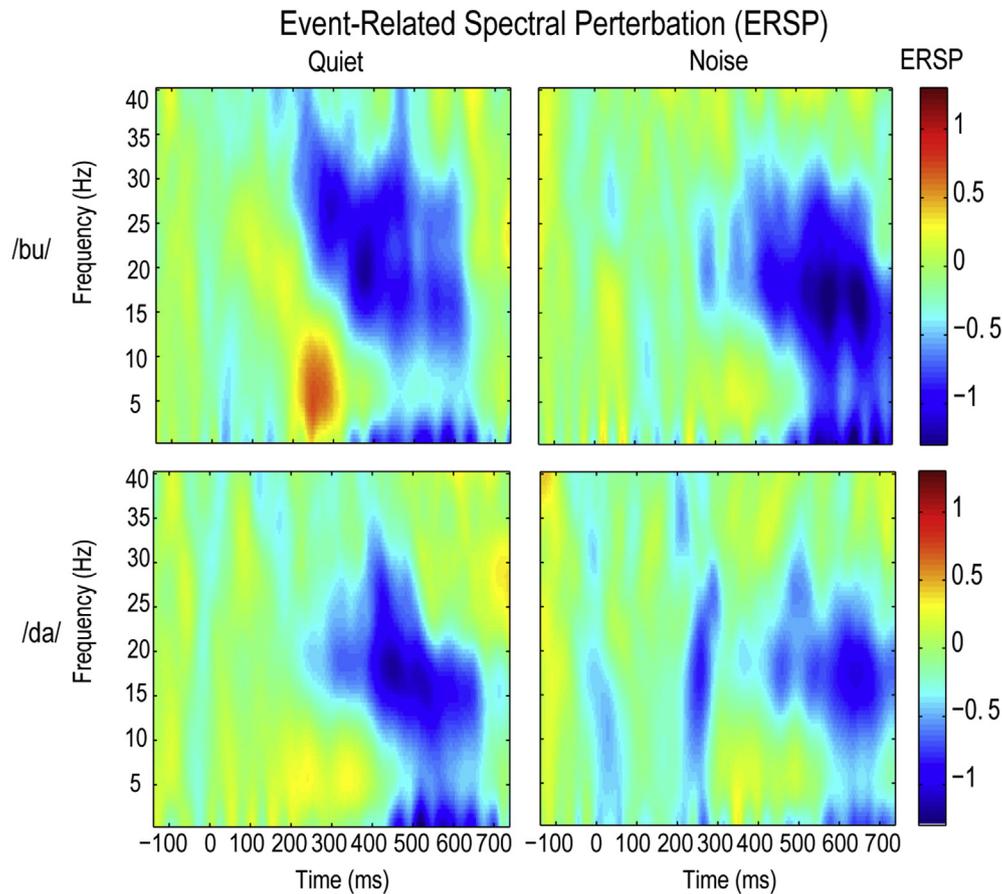


Fig. 3. Grand mean event-related spectral perturbation (ERSP) in response to the CV syllables /bu/ (top row) and /da/ (bottom row) in both quiet (left column) and babble noise (right column) listening conditions. ERSP was calculated using an analysis window of 300–400 ms for the quiet listening condition and a window of 400–500 ms for the listening condition with background noise.

Table 5

A comparison of mean (standard error) amplitudes from MMN (Koerner et al., 2016) and P3 responses to /bu/ and /da/ in quiet and noise listening conditions. The right column displays results from paired t-tests that examined differences between the quiet and noise listening conditions for each response.

	Quiet	Noise	Paired t-test (t-statistic)
MMN			
/bu/	−2.11 (0.3)	−0.40 (0.12)	−6.30***
/da/	−1.16 (0.20)	−1.02 (0.18)	−0.65
P3			
/bu/	7.64 (0.69)	6.29 (0.77)	1.29
/da/	5.97 (0.6)	3.31 (0.41)	3.69**

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

change compared to the consonant change (see direct comparison of MMN and P3 results in Table 5). Koerner et al. (2016) proposed that the larger noise induced effects on the neural processing of the vowel change may have occurred due to the need to internally resolve more transient, aperiodic consonant information in noise when deviant consonant and vowel changes were juxtaposed in a double-oddball paradigm at the pre-attentive processing level.

A main difference in EEG recording protocols between our previous MMN study (Koerner et al., 2016) and the current P3 study is that the P3 experiment required the participant to pay attention to all presented auditory stimuli and make overt behavioral responses to indicate the detection of the deviant sounds. The observed differences between the passive-MMN results and the active-P3 results may arise due to the fact that they are inherently

different components arising from different generators and underlying sensory/cognitive processes. In our pre-attentive MMN study (Koerner et al., 2016), participants were asked to ignore stimulus presentations and focus on a silent, subtitled movie. With no attention to the deviant stimuli in the double-oddball paradigm, the relative importance of the weaker, transient, and aperiodic consonant change may have increased, rendering the stronger vowel contrast to be more susceptible to noise as reflected in the MMN measure. Whereas with focused attention on the deviant stimuli, there could be more emphasis on the vowel contrast (or split emphasis to both contrasts), and the weaker consonant contrast may become more susceptible to noise as assessed by the P3 measure. Our interpretation is consistent with findings from a previous study (Gordon et al., 1993). In that study, the relative importance of voice onset time (VOT) and fundamental frequency (F0) cues during a /ba-/pa/ discrimination task in quiet was found to depend on the amount of attention paid to the task. VOT is considered the primary cue for discriminating phonemes that differ in voicing while the onset frequency of F0 is a weaker acoustic cue. When participants were asked to focus on the discrimination task, the stronger VOT cues dominated perception. However, when participants were asked to perform a distractor task during stimulus discrimination, results showed that the relative importance of the stronger VOT cues decreased while the impact of the weaker F0 cues to perception increased (Gordon et al., 1993). Similarly, when participants devoted full attention to the discrimination task in our active P3 paradigm, the weaker consonant contrast that was more difficult to behaviorally discriminate in noise showed significantly

weaker neural responses while the neural responses to the stronger vowel contrast were less impacted by the presence of background noise. An alternative explanation is that noise impacted the vowel-elicited MMN more than the consonant-elicited MMN in the double-oddball paradigm due to the autonomous preattentive processing of phonemic differences based on relative perceptual saliency of the phonemic contrasts. However, when the task required focused attention, the P3 responses for the weaker consonant contrast got more affected by noise due to the relative difficulty level of the contrasts during controlled effortful processing of the phonemic differences. Collectively, the data from our two studies suggest that the relative contributions of consonants and vowels to speech perception in noise can be modulated by attentional processing.

4.2. Effects of noise on event-related cortical oscillations

Our results revealed that ITPC in delta, theta, and alpha frequency bands in response to both deviant speech stimuli was significantly reduced by the presence of background noise. In contrast, results showed no significant noise-induced effects on ERSP across frequency bands in response to both deviant speech stimuli. On average, it appeared as if ERSP was reduced in noise compared to the quiet listening condition across stimuli; however, there was very large inter-subject variability in the data across listening conditions (see Table 3). These results are consistent with those from Koerner and Zhang (2015), which showed that background noise significantly disrupted averaged trial-by-trial neural synchrony, but did not significantly impact trial-by-trial spectral power (Koerner and Zhang, 2015). Fuentemilla et al. (2006) reported similar results while examining cortical N1 amplitude reduction in response to repeated pure tones; they found that stimulus-evoked phase synchronization occurred without any trial-by-trial spectral power modulation. These findings appear to support the “phase resetting” model, where stimulus-evoked phase synchronization of cortical oscillations has been shown to at least partially contribute to the neural generation of AERPs (Gruber et al., 2005; Klimesch et al., 2007; Makeig et al., 2004). In other words, reduced ITPC without noticeable concomitant changes in ERSP across quiet and noise listening conditions supports the possibility that partial stimulus-related phase synchronization of cortical oscillations drives the neural generation of the P3 response to speech stimuli.

4.3. The P3 as a neurophysiological marker of behavior

As described in previous work (Koerner et al., 2016), a -3 dB SNR was chosen because it was shown to result in performance on the speech perception measures used in this study without ceiling or floor effects. We assumed that the cortical P3 response would reflect this variability in behavioral performance based on findings from previous studies that revealed significant relationships between speech-evoked AERPs and behavioral speech perception in background noise (Anderson et al., 2011, 2010a, 2010b; Bennett et al., 2012; Billings et al., 2013, 2015; Cunningham et al., 2001; Koerner et al., 2016; Martin and Stapells, 2005; Martin et al., 1999; Muller-Gass et al., 2001; Song et al., 2011).

Results from our stepwise linear mixed-effects regression models are consistent with previous studies that have shown significant relationships between the cortical P3 amplitude and speech perception at the segmental-level (Kaplan-Neeman et al., 2006; Martin et al., 1997; Martin and Stapells, 2005; Whiting et al., 1998). During stepwise regression analyses, P3 latency was consistently removed from our final reduced regression models

(see Table 4). Although P3 latency is thought to represent the speed of stimulus classification, it has been reported to be unrelated to behavioral response accuracy or reaction time (McCarthy and Donchin, 1981; Verleger, 1997), which is consistent with our data showing that P3 latency was not able to significantly predict behavioral performance at the segmental- or sentence-levels across listening conditions and stimuli. Unlike results reported by Bennett et al. (2012), our data did not reveal that the averaged cortical P3 latency or amplitude was able to predict sentence-level recognition across listening conditions or deviant speech stimuli. This could be caused by differences in analysis methods, as Bennett et al. (2012) used Pearson correlations to examine brain-behavior relationships while the current work used stepwise regression models that contained multiple neural measures as fixed effects (Koerner and Zhang, 2017).

Our data also revealed that both ITPC and ERSP in various frequency bands at time points corresponding to the cortical P3 were predictive of segmental- and sentence-level behavioral performance. In other words, impaired speech perception may be indexed by noise-induced reductions in trial-by-trial neural synchrony or power recorded in response to consonant and vowel changes in an active change detection task. Although ITPC in the alpha band in response to the vowel change detection did not reach statistical significance as a predictor of sentence-level performance, ITPC in all three frequency bands in response to the vowel change were the most important for predicting sentence recognition. ERSP in the alpha frequency band for the consonant change detection was also a significant predictor of segmental-level performance across stimuli and listening conditions as well as sentence-level performance across conditions. These results support the functional associations between alpha oscillations and top-down inhibitory mechanisms (Klimesch, 2012; Straub et al., 2014; Wilsch et al., 2014). In the context of the present study, these selective inhibitory mechanisms may be related to the ability to accurately perceive speech in masking background noise.

4.4. Novelty and future directions

Unlike the previous studies using only one deviant speech stimulus (Kaplan-Neeman et al., 2006; Martin et al., 1997; Martin and Stapells, 2005; Whiting et al., 1998), the present study examined the role of attention in how background noise influenced the differential neural processing of consonant and vowel changes using a double-oddball paradigm in relation to behavioral perception. Although researchers have previously linked noise-induced variability in the P3 response with sentence-level perception in noise (Bennett et al., 2012), our results add to this body of knowledge by highlighting the different contributions from consonants and vowels to speech intelligibility in noise at both segmental and sentential levels. Furthermore, our results revealed that important complementary information can be gained about the effects of background noise on speech processing by breaking down the ERP response waveforms into individual oscillatory frequency bands of interest underlying ERP components. For instance, our results demonstrate that background noise impacted trial-by-trial neural synchrony during active speech discrimination tasks and that ITPC and ERSP measures were able to significantly predict speech perception. These results demonstrate that noise-induced disruption of event-related cortical oscillations may be functionally associated with variability in behavioral speech perception at both the segmental- and sentence-levels in background noise.

Due to time limitation in the current design, this work only used a speech-babble background noise at one SNR in a double-oddball paradigm. Future research is needed to examine how AERPs and

event-related cortical oscillations contribute to the neural generation of the cortical P3 response and how these measures are associated with behavioral performance in various listening conditions and populations with different degrees of language proficiency and degrees of hearing loss. As AERPs are non-invasive and used in difficult-to-test clinical populations for objective assessment of perceptual thresholds (Burkard et al., 2007), establishing correlations between AERP measures and cognitive measures with a range of speech-in-noise abilities would help elucidate the functional significance and clinical relevance of the current electrophysiological data. In addition, the MMN and P3 paradigms can include additional noise types, noise levels, as well as speech and non-speech stimuli for testing children, adults, and elderly to further study age-related developmental aspects of speech-in-noise processes.

5. Conclusions

This incremental work was designed to evaluate noise-induced effects on the speech-evoked cortical P3 response as well as event-related cortical oscillations in delta, theta, and alpha frequency bands in response to a consonant change and a vowel change in a double-oddball paradigm. Liner-mixed effects models were used to determine whether speech-evoked P3 latency, amplitude, ITPC, and ERSP were predictive of phoneme- and sentence-level speech recognition across listening conditions (quiet vs. noise), stimuli (/da/ vs. /bu/), and participants. As predicted, the presence of background noise significantly increased P3 latency, decreased P3 amplitude, and decreased ITPC in associated frequency bands without significant noise-induced effects on trial-by-trial power change across frequency bands. Consistent with behavioral performance at the segmental level, background noise had a larger impact on the neural processing of detecting the consonant change in comparison with detecting the vowel change. This pattern of differential noise-induced impacts on consonant vs. vowel discrimination in the P3 response was opposite to our previous MMN study (Koerner et al., 2016), indicating the important role of attention in modulating the ERP components of interest. Moreover, the cortical P3 response and its associated event-related cortical oscillations represent potential neural markers for speech perception at both segmental and sentence-levels. This work has important implications regarding the clinical utility of the P3 response that should be validated in future studies using different stimuli, listening conditions, and participant populations.

Acknowledgements

This research was supported in part by the University of Minnesota (Charles E. Speaks Graduate Fellowship to TKK, Brain Imaging Research Project Award to YZ, and Grand Challenges Exploratory Research Grant to YZ & HZ) and the Capita Foundation (YZ). Participant payment was partially supported by NIDCD R01-DC008306 (PBN). YZ conceived the study; YZ and TKK designed research; TKK and YZ performed research and analyzed data; PN, BW and HZ offered consultation; TKK and YZ wrote the article. The authors would like to thank Edward Carney, Sharon Miller, Adam Svec, Yingjiu Nie, and Luodi Yu for their assistance and two anonymous reviewers for their valuable comments and detailed suggestions to improve the manuscript. The authors have no conflicts of interest to disclose.

Appendix A. Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.heares.2017.04.009>.

References

- Anderson, S., Parbery-Clark, A., Yi, H.-G., Kraus, N., 2011. A neural basis of speech-in-noise perception in older adults. *Ear Hear* 32, 750–757. <http://dx.doi.org/10.1097/AUD.0b013e31822229d3.A>.
- Anderson, S., Skoe, E., Chandrasekaran, B., Kraus, N., 2010a. Neural timing is linked to speech perception in noise. *J. Neurosci.* 30, 4922–4926.
- Anderson, S., Skoe, E., Chandrasekaran, B., Zecker, S., Kraus, N., 2010b. Brainstem correlates of speech-in-noise perception in children. *Hear. Res.* 270, 151–157. <http://dx.doi.org/10.1016/j.heares.2010.08.001>.
- Baayen, R.H., Davidson, D.J., Bates, D.M., 2008. Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* 59, 390–412. <http://dx.doi.org/10.1016/j.jml.2007.12.005>.
- Bagiella, E., Sloan, R.P., Heitjan, D.F., 2000. Mixed-effects models in psychophysiology. *Psychophysiology* 37, 13–20.
- Başar, E., Demiralp, T., Schürmann, M., Başar-Eroglu, C., Ademoglu, A., 1999. Oscillatory brain dynamics, wavelet analysis, and cognition. *Brain Lang.* 66, 146–183. <http://dx.doi.org/10.1006/brln.1998.2029>.
- Basar-Eroglu, C., Başar, E., Demiralp, T., Schürmann, M., 1992. P300-response: possible psychophysiological correlates in delta and theta frequency channels. A review. *Int. J. Psychophysiol.* 13, 161–179.
- Bennett, K.O., Billings, C.J., Molis, M.R., Leek, M.R., 2012. Neural encoding and perception of speech signals in informational masking. *Ear Hear* 33, 231–238.
- Billings, C.J., Bennett, K.O., Molis, M.R., Leek, M.R., 2011. Cortical encoding of signals in noise: effects of stimulus type and recording paradigm. *Ear Hear* 32, 53–60.
- Billings, C.J., Mcmillan, G.P., Penman, T.M., Gille, S.M., 2013. Predicting perception in noise using cortical auditory evoked potentials. *J. Assoc. Res. Otolaryngol.* 14, 891–903. <http://dx.doi.org/10.1007/s10162-013-0415-y>.
- Billings, C.J., Penman, T.M., Mcmillan, G.P., Ellis, E.M., 2015. Electrophysiology and perception of speech in noise in older listeners: effects of hearing impairment and age. *Ear Hear* 36, 710–722.
- Burkard, R.F., Don, M., Eggermont, J.J. (Eds.), 2007. *Auditory Evoked Potentials: Basic Principles and Clinical Application*. Lippincott Williams & Wilkins, Philadelphia, PA.
- Caramazza, A., Papagno, C., Rumel, W., 2000. The selective impairment of phonological processing in speech production. *Brain Lang.* 75, 428–450. <http://dx.doi.org/10.1006/brln.2000.2379>.
- Carreiras, M., Price, C.J., 2008. Brain activation for consonants and vowels. *Cereb. Cortex* 18, 1727–1735. <http://dx.doi.org/10.1093/cercor/bhm202>.
- Carreiras, M., Vergara, M., Perea, M., 2009. ERP correlates of transposed-letter priming effects: the role of vowels versus consonants. *Psychophysiology* 46, 34–42. <http://dx.doi.org/10.1111/j.1469-8986.2008.00725.x>.
- Cole, R.A., Yan, Y., Mak, B., Fanty, M., Bailey, T., 1996. The contribution of consonants versus vowels to word recognition in fluent speech. *Proc. ICASSP 96*, 853–856.
- Cunningham, J., Nicol, T., Zecker, S.G., Bradlow, A., Kraus, N., 2001. Neurobiologic responses to speech in noise in children with learning problems: deficits and strategies for improvement. *Clin. Neurophysiol.* 112, 758–767. [http://dx.doi.org/10.1016/S1388-2457\(01\)00465-5](http://dx.doi.org/10.1016/S1388-2457(01)00465-5).
- David, O., Kilner, J.M., Friston, K.J., 2006. Mechanisms of evoked and induced responses in MEG/EEG. *NeuroImage* 31, 1580–1591. <http://dx.doi.org/10.1016/j.neuroimage.2006.02.034>.
- Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics (pdf, 0.7 MB). *J. Neurosci. Methods* 134, 9–21.
- Demiralp, T., Ademoglu, A., Istefanopoulos, Y., Başar-Eroglu, C., Başar, E., 2001. Wavelet analysis of oddball P300. *Int. J. Psychophysiol.* 39, 221–227.
- Fogerty, D., Humes, L.E., 2012. The role of vowel and consonant fundamental frequency, envelope, and temporal fine structure cues to the intelligibility of words and sentences. *J. Acoust. Soc. Am.* 131, 1490–1501. <http://dx.doi.org/10.1121/1.3676696>.
- Fogerty, D., Humes, L.E., 2010. Perceptual contributions to monosyllabic word intelligibility: segmental, lexical, and noise replacement factors. *J. Acoust. Soc. Am.* 128, 3114–3125. <http://dx.doi.org/10.1121/1.3493439>.
- Fogerty, D., Kewley-port, D., 2009. Perceptual contributions of the consonant-vowel boundary to sentence intelligibility. *J. Acoust. Soc. Am.* 126, 847–857. <http://dx.doi.org/10.1121/1.3159302>.
- Fuentemilla, L., Marco-Pallarés, J., Grau, C., 2006. Modulation of spectral power and of phase resetting of EEG contributes differentially to the generation of auditory event-related potentials. *NeuroImage* 30, 909–916. <http://dx.doi.org/10.1016/j.neuroimage.2005.10.036>.
- Gordon, P.C., Eberhardt, J.L., Rueckl, J.G., 1993. Attentional modulation of the phonetic significance of acoustic cues. *Cogn. Psychol.* 25, 1–42.
- Gruber, W.R., Klimesch, W., Sauseng, P., Doppelmayr, M., 2005. Alpha phase synchronization predicts P1 and N1 latency and amplitude size. *Cereb. Cortex* 15, 371–377.
- Intriligator, J., Polich, J., 1995. On the relationship between EEG and ERP variability. *Int. J. Psychophysiol.* 20, 59–74.
- Intriligator, J., Polich, J., 1994. On the relationship between background EEG and the P300 event-related potential. *Biol. Psychol.* 37, 207–218.
- Kaplan-Neeman, R., Kishon-Rabin, L., Henkin, Y., Muchnik, C., 2006. Identification of syllables in noise: electrophysiological and behavioral correlates. *J. Acoust. Soc. Am.* 120, 926–933. <http://dx.doi.org/10.1121/1.2217567>.
- Kewley-Port, D., Burkley, T.Z., Lee, J.H., 2007. Contribution of consonant versus vowel information to sentence intelligibility for young normal-hearing and elderly hearing-impaired listeners. *J. Acoust. Soc. Am.* 122, 2365–2375. <http://dx.doi.org/10.1121/1.2717567>.

- [dx.doi.org/10.1121/1.2773986](https://doi.org/10.1121/1.2773986).
- Klimesch, W., 2012. Alpha-band oscillations, attention, and controlled access to stored information. *Trends Cogn. Sci.* 16, 606–617.
- Klimesch, W., Sauseng, P., Hanslmayr, S., Gruber, W., Freunberger, R., 2007. Event-related phase reorganization may explain evoked neural dynamics. *Neurosci. Biobehav. Rev.* 31, 1003–1016. [http://dx.doi.org/10.1016/j.neubiorev.2007.03.005](https://dx.doi.org/10.1016/j.neubiorev.2007.03.005).
- Koerner, T., Zhang, Y., Nelson, P., 2013. Links between mismatch negativity responses and speech intelligibility in noise. *J. Acoust. Soc. Am.* 132, 2049.
- Koerner, T.K., Zhang, Y., 2017. Application of linear mixed-effects models in human neuroscience research: a comparison with Pearson correlation in two auditory electrophysiology studies. *Brain Sci.* 7, 26.
- Koerner, T.K., Zhang, Y., 2015. Effects of background noise on inter-trial phase coherence and auditory N1-P2 responses to speech. *Hear. Res.* 328, 113–119.
- Koerner, T.K., Zhang, Y., Nelson, P.B., Wang, B., Zou, H., 2016. Neural indices of phonemic discrimination and sentence-level speech intelligibility in quiet and noise: a mismatch negativity study. *Hear. Res.* 339, 40–49. [http://dx.doi.org/10.1016/j.heares.2016.06.001](https://dx.doi.org/10.1016/j.heares.2016.06.001).
- Kolev, V., Demiralp, T., Yordanova, J., Ademoglu, A., Isoglu-Alkac, U., 1997. Time-frequency analysis reveals multiple functional components during oddball P300. *NeuroReport* 8, 2061–2065.
- Korczak, P.A., Stapells, D.R., 2010. Effects of various articulatory features of speech on cortical event-related potentials and behavioral measures of speech-sound processing. *Ear Hear* 31, 491–504.
- Ladefoged, P., 2006. *A Course in Phonetics*, fifth ed. Thompson, Boston.
- Magazi, D.A., 2015. Linear mixed-effects models for within-participant psychology experiments: an introductory tutorial and free, graphical user interface (LMMgui). *Front. Psychol.* 6, 1–7. [http://dx.doi.org/10.3389/fpsyg.2015.00002](https://dx.doi.org/10.3389/fpsyg.2015.00002).
- Makeig, S., 1993. Auditory event-related dynamics of the EEG spectrum and effects of exposure to tones. *Electroencephalogr. Clin. Neurophysiol.* 86, 283–293. [http://dx.doi.org/10.1016/0013-4694\(93\)90110-H](https://dx.doi.org/10.1016/0013-4694(93)90110-H).
- Makeig, S., Debener, S., Onton, J., Delorme, A., 2004. Mining event-related brain dynamics. *Trends Cogn. Sci.* 8, 204–210. [http://dx.doi.org/10.1016/j.tics.2004.03.008](https://dx.doi.org/10.1016/j.tics.2004.03.008).
- Martin, B.A., Kurtzberg, D., Stapells, D.R., 1999. The effects of decreased audibility produced by high-pass noise masking on N1 and the mismatch negativity to speech sounds /ba/ and /da/. *J. Speech Lang. Hear. Res.* 42, 271–287.
- Martin, B.A., Sigal, A., Kurtzberg, D., Stapells, D.R., 1997. The effects of decreased audibility produced by high-pass noise masking on cortical event-related potentials to speech sounds/ba/and/da/. *J. Acoust. Soc. Am.* 101, 1585–1599.
- Martin, B.A., Stapells, D.R., 2005. Effects of low-pass noise masking on auditory event-related potentials to speech. *Ear Hear* 26, 195–213.
- McCarthy, G., Donchin, E., 1981. A metric for thought: a comparison of P300 latency and reaction time. *Science* 211, 77–80.
- Miller, G., Nicely, P., 1955. An analysis of perceptual confusions among some English consonant. *J. Acoust. Soc. Am.* 27, 338–352.
- Muller-Gass, A., Marcoux, A., Logan, J., Campbell, K.B., 2001. The intensity of masking noise affects the mismatch negativity to speech sounds in human subjects. *Neurosci. Lett.* 299, 197–200. [http://dx.doi.org/10.1016/S0304-3940\(01\)01508-7](https://dx.doi.org/10.1016/S0304-3940(01)01508-7).
- Näätänen, R., Paavilainen, P., Rinne, T., Alho, K., 2007. The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clin. Neurophysiol.* 118, 2544–2590. [http://dx.doi.org/10.1016/j.clinph.2007.04.026](https://dx.doi.org/10.1016/j.clinph.2007.04.026).
- Nash-Kille, A., Sharma, A., 2014. Inter-trial coherence as a marker of cortical phase synchrony in children with sensorineural hearing loss and auditory neuropathy spectrum disorder fitted with hearing aids and cochlear implants. *Clin. Neurophysiol.* 125, 1459–1470. [http://dx.doi.org/10.1016/j.clinph.2013.11.017](https://dx.doi.org/10.1016/j.clinph.2013.11.017).
- Nie, Y., Zhang, Y., Nelson, P.B., 2014. Auditory stream segregation using bandpass noises: evidence from event-related potentials. *Front. Neurosci.* 8, 1–12.
- Niemitalo-Haapola, E., Haapala, S., Jansson-Verkasalo, E., Kujala, T., 2015. Background noise degrades central auditory processing in toddlers. *Ear Hear* 36, 342–351.
- Niquette, P., Gundmundsen, G., Killion, M., 2001. *QuickSIN Speech-in-noise Test Version 1.3*.
- Owren, M.J., Cardillo, G.C., 2006. The relative roles of vowels and consonants in discriminating talker identity versus word meaning. *J. Acoust. Soc. Am.* 119, 1727–1739. [http://dx.doi.org/10.1121/1.2161431](https://dx.doi.org/10.1121/1.2161431).
- Parbery-Clark, A., Marmel, F., Bair, J., Kraus, N., 2011. What subcortical–cortical relationships tell us about processing speech in noise. *Cogn. Neurosci.* 33, 549–557. [http://dx.doi.org/10.1111/j.1460-9568.2010.07546.x](https://dx.doi.org/10.1111/j.1460-9568.2010.07546.x).
- Parikh, G., Loizou, P.C., 2005. The influence of noise on vowel and consonant cues. *J. Acoust. Soc. Am.* 118, 3874–3888. [http://dx.doi.org/10.1121/1.2118407](https://dx.doi.org/10.1121/1.2118407).
- Phatak, S.A., Allen, J.B., 2007. Consonant and vowel confusions in speech-weighted noise. *J. Acoust. Soc. Am.* 121, 2312–2326. [http://dx.doi.org/10.1121/1.2642397](https://dx.doi.org/10.1121/1.2642397).
- Pickett, J.M., 1957. Perception of vowels heard in noises of various spectra. *J. Acoust. Soc. Am.* 29, 613–620.
- Picton, T.W., 1992. The P300 wave of the human event-related potential. *J. Clin. Neurophysiol.* 9, 465–479.
- Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., R Core Team, 2016. *nlme: Linear and Nonlinear Mixed Effects Models*.
- Polich, J., 2004. Clinical application of the P300 event-related brain potential. *Phys. Med. Rehabil. Clin. N. Am.* 15, 133–161.
- Polich, J., 1997. EEG and ERP assessment of normal aging. *Electroencephalogr. Clin. Neurophysiol.* 104, 244–256.
- R Core Team, 2014. *R: a Language and Environment for Statistical Computing*.
- Rao, A., Rishiq, D., Yu, L., Zhang, Y., Abrams, H., 2017. Neural correlates of selective attention with hearing aid use followed by ReadMyQuips auditory training program. *Ear Hear* 38, 28–41.
- Russo, N., Nicol, T., Musacchia, G., Kraus, N., 2004. Brainstem responses to speech syllables. *Clin. Neurophysiol.* 115, 2021–2030. [http://dx.doi.org/10.1016/j.biotechadv.2011.08.021](https://dx.doi.org/10.1016/j.biotechadv.2011.08.021). *Secreted*.
- Russo, N., Zecker, S., Trommer, B., Chen, J., Kraus, N., 2009. Effects of background noise on cortical encoding of speech in autism spectrum disorders. *J. Autism Dev. Disord.* 39, 1185–1196. [http://dx.doi.org/10.1007/s10803-009-0737-0](https://dx.doi.org/10.1007/s10803-009-0737-0).
- Sauseng, P., Klimesch, W., Gruber, W.R., Hanslmayr, S., Freunberger, R., Doppelmayr, M., 2007. Are event-related potential components generated by phase resetting of brain oscillations? A critical discussion. *Neuroscience* 146, 1435–1444. [http://dx.doi.org/10.1016/j.neuroscience.2007.03.014](https://dx.doi.org/10.1016/j.neuroscience.2007.03.014).
- Song, J.H., Skoe, E., Banai, K., Kraus, N., 2011. Perception of speech in noise: neural correlates. *J. Cogn. Neurosci.* 23, 2268–2279. [http://dx.doi.org/10.1162/jocn.2010.21556](https://dx.doi.org/10.1162/jocn.2010.21556). *Perception*.
- Spencer, K.M., Polich, J., 1999. Poststimulus EEG spectral analysis and P300: attention, task, and probability. *Psychophysiology* 36, 220–232.
- Straub, A., Wostmann, M., Obleser, J., 2014. Cortical alpha oscillations as a tool for auditory selective inhibition. *Front. Neurosci.* 8, 107.
- Verleger, R., 1997. On the utility of P3 latency as an index of mental chronometry. *Psychophysiology* 34, 131–156.
- Whiting, K.A., Martin, B.A., Stapells, D.R., 1998. The Effects of Broadband Noise Masking on Cortical Event-Related Potentials to Speech Sounds /ba/ and /da/. *Ear Hear* 19, 218–231.
- Wilsch, A., Henry, M.J., Herrmann, B., Maess, B., Obleser, J., 2014. Alpha oscillatory dynamics index temporal expectation benefits in working memory. *Cereb. Cortex* 25, 1938–1946.
- Yordanova, J., Kolev, V., 1998. Event-related alpha oscillations are functionally associated with P300 during information processing. *NeuroReport* 3159–3164.
- Zhang, Y., Koerner, T., Miller, S., Grice-Patil, Z., Svec, A., Akbari, D., Tusler, L., Carney, E., 2011. Neural coding of formant-exaggerated speech in the infant brain. *Dev. Sci.* 14, 566–581.