Hearing Research 339 (2016) 40–49

Contents lists available at ScienceDirect

Hearing Research

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

Research paper

Neural indices of phonemic discrimination and sentence-level speech intelligibility in quiet and noise: A mismatch negativity 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 27 January 2016
Received in revised form
16 May 2016
Accepted 2 June 2016
Available online 4 June 2016

Keywords:
MMN
Time-frequency analysis
Theta band
Speech-in-noise perception
Linear mixed effects model

ABSTRACT

Successful speech communication requires the extraction of important acoustic cues from irrelevant background noise. In order to better understand this process, this study examined the effects of background noise on mismatch negativity (MMN) latency, amplitude, and spectral power measures as well as behavioral speech intelligibility tasks. Auditory event-related potentials (AERPs) were obtained from 15 normal-hearing participants to determine whether pre-attentive MMN measures recorded in response to a consonant (from /ba/ to /bu/) and vowel change (from /ba/ to /da/) in a double-oddball paradigm can predict sentence-level speech perception. The results showed that background noise increased MMN latencies and decreased MMN amplitudes with a reduction in the theta frequency band power. Differential noise-induced effects were observed for the pre-attentive processing of consonant and vowel changes due to different degrees of signal degradation by noise. Linear mixed-effects models further revealed significant correlations between the MMN measures and speech intelligibility scores across conditions and stimuli. These results confirm the utility of MMN as an objective neural marker for understanding noise-induced variations as well as individual differences in speech perception, which has important implications for potential clinical applications.

1. Introduction

Speech communication often takes place in the presence of background noise, which can be difficult for hard of hearing listeners as well as many listeners with normal hearing. In recent years, there has been a surge of interest investigating noise-induced modulatory effects on cortical/subcortical responses to examine the neural networks and brain mechanisms supporting higher-level cognitive and linguistic skills (Anderson et al., 2010a; Billings et al., 2013; Du et al., 2014; Koerner and Zhang, 2015; Mesgarani et al., 2014; Vaden et al., 2015; Wong et al., 2008). Cortical auditory event-related potentials (AERPs) are one representative method of measuring the neural coding of speech sounds in various listening conditions. In particular, the auditory mismatch negativity (MMN) response provides an objective electrophysiological measure of the neural timing and strength of pre-attentive auditory discrimination. It peaks at approximately 100–250 ms post-stimulus onset, which is typically generated when a participant's sensory memory trace of a “standard” stimulus detects a change by a less frequently occurring “deviant” stimulus in the absence of attention or any overt behavioral response (Näätänen et al., 2007). The present study attempts to address whether the MMN response is a good predictor of speech perception performance at both segmental and sentence levels in quiet and noise.

The pre-attentive cortical MMN response has been linked with behavioral speech perception in a number of studies. Representative topics include language learning and development in children (Kraus et al., 1996; Kraus and Cheour, 2000), native (Aaltonen et al., 1987; Christmann et al., 2014) and non-native speech perception in adults (Bidelman and Dexter, 2015; Brunellière et al., 2011; Näätänen et al., 1997; Winkler et al., 1999; Zhang et al., 2009), the

Abbreviations: EEG, electroencephalography; AERP, auditory event-related potential; MMN, mismatch negativity; SNR, signal-to-noise ratio; ANOVA, analysis of variance; SE, standard error; CV, consonant-vowel

* 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).

http://dx.doi.org/10.1016/j.heares.2016.06.001
0378-5955/

effects of hearing loss (Kraus et al., 1995b; Oates et al., 2002) and cochlear implantation (Groenen et al., 1996; Kraus et al., 1993, 1995b), and neural plasticity in auditory training (Kraus et al., 1995a; Tremblay et al., 1997; Tremblay et al., 1998; Zhang et al., 2009). Studies have shown that the MMN responses for phonetic discrimination in quiet can predict first- and second-language attainment (Garcia-Sierra et al., 2011; Jansson-Verkasalo et al., 2004; Jakoby et al., 2011; Kuhl et al., 2005; Molfese and Molfese, 1997). There is also evidence that pre-attentive speech perception in noise results in MMN amplitude decreases and latency increases when compared to quiet conditions (Kozou et al., 2005; Martin and Stapells, 2005; Martin et al., 1999; Muller-Gass et al., 2001). These noise-induced changes in the MMN response for detecting phonemic changes are associated with decrements in behavioral measures of discriminatory accuracy and increases in reaction time (Martin and Stapells, 2005; Martin et al., 1999; Muller-Gass et al., 2001). However, to our knowledge, no adult studies have examined brain-behavior relationships between changes in the pre-attentive MMN at the segmental level and performance on sentence-level word recognition performance across quiet and noise conditions.

In addition to the conventional ERP latency and amplitude measures, a recent trend in neurophysiological studies is the development of sophisticated time-frequency analyses to examine the role of various neural oscillation frequency bands of the EEG signal in the generation of AERP waveforms. These cortical oscillations are thought to modulate neural excitability and timing, which enables information exchange between cortical processes that are responsible for sensory and cognitive events (Başar et al., 1999; Klimesch et al., 2007; Koerner and Zhang, 2015; Makeig et al., 2004; Sauseng et al., 2007; Zhang et al., 2011). In particular, several studies have revealed the contribution of the theta frequency band (4–8 Hz) in driving the neuronal generation of the MMN in frontal and temporal areas (Bishop and Hardiman, 2010; Choi et al., 2013; Fuentemilla et al., 2008; Hsiao et al., 2009; Ko et al., 2012). Collectively, these studies show that neural generation of the MMN response are accompanied by phase alignment and power modulation of theta band activity. In the literature, the theta activity is proposed to be associated with several other cognitive functions including memory encoding, retrieval, and maintenance (Klimesch et al., 2008; Ward, 2003). Although previous studies have revealed prolonged latency and reduced amplitude in the MMN response due to the presence of background noise (Kozou et al., 2005; Martin and Stapells, 2005; Martin et al., 1999; Muller-Gass et al., 2001), it remains unknown how noise may modulate MMN spectral power in the theta band.

The present study aimed to investigate the effects of speech-babble background noise on the pre-attentive cortical processing of consonant and vowel changes by analyzing MMN latency, amplitude, and EEG spectral power measures. It is well established that the MMN responses can show high inter- and intra-subject variability in amplitude and latency (Kurtzberg et al., 1995; Lang et al., 1995; Martin et al., 1999, 2008; Näätänen et al., 2007; Näätänen, 1995; Stapells, 2002). This study was designed to take individual variability into account and investigate whether the objective neurophysiological measures in response to consonant and vowel phonemic contrasts in a double-oddball paradigm (Xi et al., 2010) can predict sentence-level speech intelligibility performance across quiet and noise listening conditions. The double oddball paradigm is a modified version of the conventional MMN protocol, in which the presentation trials for the single deviant stimulus are shared by two deviants (e.g., a consonant contrast and a vowel contrast) at equal probability of occurrence. Thus this paradigm allows the investigation of two MMN responses, one for each deviant stimulus, during the same recording session. Animal and human studies examining the neural processing of speech in noise have revealed differential effects of noise on consonant and vowel stimuli, such that the neural responses to steady-state vowel stimuli are more robust in noise than those to more transient, aperiodic consonant stimuli (Cunningham et al., 2002; Russo et al., 2004; Shetake et al., 2011; Song et al., 2011). Additionally, behavioral and neurophysiological research suggests that consonant and vowel stimuli may be processed by separate neural mechanisms in the auditory cortex (Caramazza et al., 2000; Carreiras et al., 2009; Fogerty and Humes, 2012; Fogerty et al., 2012; Kewley-Port et al., 2007; Liberman and Mattingly, 1985; Miceli et al., 2004). Thus the use of a double-oddball paradigm would allow us to test the differences in neural sensitivity to vowel and consonant changes across the quiet and noise conditions and their relative contributions to higher-level behavioral performance in sentence recognition.

We hypothesized that the introduction of background noise would result in increases in MMN latency and decreases in MMN amplitude, which would be accompanied by reduced spectral power in the theta band. In addition, noise would differentially affect the cortical processing of the consonant and vowel changes, such that the pre-attentive detection of the consonant change would be more vulnerable to disruption in noise than the vowel change. We further hypothesized that at least some of the MMN measures would be able to predict higher-level behavioral sentence recognition.

2. Methods

2.1. Subjects

The participants in the study were 15 individuals (mean age = 22.6 years, age range = 19–32 years, 5 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) and no history for speech, language, or cognitive difficulties. All participants were right handed and were native speakers of American English. 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

2.2.1. Stimuli for ERP measures

The consonant-vowel (CV) syllables, /ba/, /da/, and /bu/, were synthesized with the HLsyn software program (Sensimetrics Corporation, USA) using a 10 kHz sampling rate (Koerner and Zhang, 2015). All the syllables were 170 ms in duration with a steady fundamental frequency of 100 Hz and a steady F4 at 3300 Hz. 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 of the /ba/ sound. For /da/, the F1, F2, and F3 onset frequencies were 362 Hz, 1832 Hz, and 2540 Hz, and for /bu/, the formant onset frequencies were at 230 Hz, 900 Hz, and 2480 Hz. The steady center F1, F2, and F3 frequencies for the vowel portion (50–170 ms) of the /ba/ and /da/ syllables were 674 Hz, 1140 Hz, and 2350 Hz. The steady center F1, F2, and F3 frequencies for the vowel portion of /bu/ were 320 Hz, 860 Hz, and 2620 Hz, respectively. The background noise used in this study was a four-talker speech babble noise that was adopted from the Quick Speech In Noise Test (Quick-SIN) (Niquette et al., 2001). All of the CV syllables and the 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) (Koerner and Zhang, 2015).