NEUROSCIENCE
INTERNATIONAL BRAIN IBRO RESEARCH ORGANIZATION

RESEARCH ARTICLE

X. Zhang et al. / Neuroscience xxx (2018) xxx–xxx

Background Suppression and its Relation to Foreground Processing of Speech Versus Non-speech Streams

Xiaochen Zhang, [a†] **Xiaolin Li,** [a†] **Jingjing Chen** [a] **and Qin Gong** [a,b*]

[a] *Department of Biomedical Engineering, School of Medicine, Tsinghua University, Beijing, China*

[b] *Research Center of Biomedical Engineering, Graduate School at Shenzhen, Tsinghua University, Shenzhen, Guangdong Province, China*

Abstract—Since sound perception takes place against a background with a certain amount of noise, both speech and non-speech processing involve extraction of target signals and suppression of background noise. Previous works on early processing of speech phonemes largely neglected how background noise is encoded and suppressed. This study aimed to fill in this gap. We adopted an oddball paradigm where speech (vowels) or non-speech stimuli (complex tones) were presented with or without a background of amplitude-modulated noise and analyzed cortical responses related to foreground stimulus processing, including mismatch negativity (MMN), N2b, and P300, as well as neural representations of the background noise, that is, auditory steady-state response (ASSR). We found that speech deviants elicited later and weaker MMN, later N2b, and later P300 than non-speech ones, but N2b and P300 had similar strength, suggesting more complex processing of certain acoustic features in speech. Only for vowels, background noise enhanced N2b strength relative to silence, suggesting an attention-related speech-specific process to improve perception of foreground targets. In addition, noise suppression in speech contexts, quantified by ASSR amplitude reduction after stimulus onset, was lateralized towards the left hemisphere. The left-lateralized suppression following N2b was associated with the N2b enhancement in noise for speech, indicating that foreground processing may interact with background suppression, particularly during speech processing. Together, our findings indicate that the differences between perception of speech and non-speech sounds involve not only the processing of target information in the foreground but also the suppression of irrelevant aspects in the background.

Keywords: Speech-in-noise, oddball, mismatch negativity (MMN), N2b, P300, auditory steady-state response.

INTRODUCTION

Although humans communicate with each other mainly through the production and reception of speech, an understanding of speech perception remains elusive. It is still debated whether the perception of speech and non-speech signals involves different neural processes (Diehl et al., 2004; Samuel, 2011). Some claim that speech perception is a specialized mechanism (Liberman and Mattingly, 1985), others propose general mechanisms for both speech and non-speech sounds (Galantucci et al., 2006), and yet others suggest that both theories may account for speech processing at various stages (Zatorre and Gandour, 2008).

Comparisons between processing of speech phonemes and their non-speech counterparts have used recordings of event-related potentials (ERPs), including mismatch negativity (MMN), N2b, and P300 (e.g., Vihla et al., 2000; Sussman et al., 2004; Sorokin et al., 2010; Bennett et al., 2012). Sussman et al. (2004) found differences in later attention-related processing (reflected by N2b-P300) but not in early pre-attentive processing (reflected by MMN) of speech and non-speech streams. In contrast, some researchers found differences even in early pre-attentive processing, as suggested by stronger (Jaramillo et al., 2001; Sorokin et al., 2010) or weaker MMN (Vihla et al., 2000; Kozou et al., 2005). The aforementioned studies compared speech and non-speech processing mostly in noise-free backgrounds. However, since speech communication usually occurs in noisy environments, it seems appropriate to use experimental paradigms involving listening to speech accompanied by competing background noise. Noise deteriorates the neural representations of

---

*Correspondence to: Q. Gong, Room C212, School of Medicine, Tsinghua University, Beijing 100084, China.
E-mail address: gongqin@mail.tsinghua.edu.cn (Q. Gong).

[†] Contributed equally to this work.

Abbreviations: AM, amplitude-modulated; ASSR, auditory steady-state response; EEG, electroencephalography; ERP, event-related potential; FDR, false discovery rate; GFP, global field power; ITPL, inter-trial phase locking; MMN, mismatch negativity; RMANOVA, repeated measures analysis of variance; SNR, signal-to-noise ratio; SOA, stimulus onset asynchrony; SPL, sound pressure level; TANOVA, topographic analysis of variance; VEOG, vertical electro-oculograms.

https://doi.org/10.1016/j.neuroscience.2018.01.009
0306-4522/

Please cite this article in press as: Zhang X et al. Background Suppression and its Relation to Foreground Processing of Speech Versus Non-speech Streams. Neuroscience (2018), https://doi.org/10.1016/j.neuroscience.2018.01.009

speech (Parbery-Clark et al., 2009a, 2011), creating more challenging conditions than in a silent background. Also, people who perform better in speech-in-noise perception seem to be affected less by competing noise (Anderson et al., 2011), suggesting a connection between noise suppression and target signal processing. Furthermore, some measures of neural processing of speech vs. non-speech signals, such as MMNs and P300s, appear to be differentially affected by various types of noise (Kozou et al., 2005; Bennett et al., 2012). Since the human auditory system has to focus on relevant information while suppressing other signals (Hayrynen et al., 2016), the reported differences between the effect of noise on speech and non-speech perception may be attributed to (1) different processing of foreground stimuli, and/or (2) different suppression of background noise. Nevertheless, previous studies on speech perception in noise focused on processing of foreground stimuli (e.g., Cunningham et al., 2001; Kozou et al., 2005; Parbery-Clark et al., 2009a,b; Bennett et al., 2012; Jin and Liu, 2012), while neglecting mechanisms underlying the suppression of competing noise.

Noise suppression can be investigated using the auditory steady-state response (ASSR), an ERP following the periodic modulation of stimulus amplitude or frequency (Picton et al., 2003; Wang et al., 2012). Thus, if the noise competing with foreground stimuli is temporally modulated in amplitude, say 40 Hz (Pastor et al., 2002; Picton et al., 2003), the evoked 40-Hz ASSR can be regarded as reflecting the neural representation of background noise. Meanwhile, the neural processing of the foreground stimuli can be investigated via transient ERP components. Using an oddball paradigm with background amplitude-modulated (AM) noise, Rockstroh et al. (1996) found a reduction of ASSR power after the occurrence of foreground stimuli. A similar paradigm was also used to demonstrate the abnormal suppression of competing noise among individuals with schizophrenia (Hayrynen et al., 2016).

In the present study, we analyzed neural representations of both foreground stimuli (using transient ERPs) and background noise (using ASSR) to test the hypothesis that speech and non-speech processing in noisy environments differ not only in encoding of foreground target information but also in inhibition of irrelevant background signals.

EXPERIMENTAL PROCEDURES

Subjects

Fifteen subjects (6 females; mean ± SD age, 24 ± 2 years) participated in the experiment. All subjects were native Chinese speakers with normal hearing. Pure tone thresholds for both ears were all below 25 dB hearing level at octave frequencies from 125 to 8000 Hz. None of the subjects have history of neurological or psychiatric disorders. All subjects were paid for their time and gave informed consent to the experimental protocol approved by the Institutional Review Board at Tsinghua University (IRB00008273).

Stimuli

As shown in Fig. 1A, stimuli were presented in four separate blocks, as follows: (1) speech oddball stream against a noisy background (SN); (2) speech oddball stream against a quiet background (SQ); (3) non-speech oddball stream against a noisy background (nSN); (4) non-speech oddball stream against a quiet background (nSQ). Each block contained 700 standard stimuli and 100 deviant stimuli (with pitch contours different from standard ones). Standard stimuli had static pitch contours at 122 Hz, while deviant stimuli had rising pitch contours from 122 to 146 Hz. In speech conditions (SN and SQ), foreground stimuli were vowels /a/, generated in Praat (http://www.fon.hum.uva.nl/praat/) with a Klatt algorithm (Klatt and Klatt, 1990). The duration of foreground stimuli was 150 ms, including 10-ms rise/fall times. The stimulus onset asynchrony (SOA) was 1567 ms. In the SN block, broadband (0.5–4 kHz) AM noise with a 40-Hz envelope was the competing noise, which started 500 ms ahead of stimulus onset and had a duration of 1.5 s. In non-speech conditions (nSN and nSQ), the stimulus paradigm (including stimulus duration, rise/fall times, SOA, and AM noise) was the same as in the speech conditions, except that the foreground stimuli were complex tones which contained fundamental frequency components as well as five harmonic (3rd, 6th, 7th, 8th, and 12th) components. Similar to previous studies (Xi et al., 2010; Wang et al., 2017), the non-speech stimuli in our experiment were generated to match the speech stimuli in terms of fundamental frequency, amplitude, and duration parameters. Speech and non-speech stimuli differed only in spectral components; in non-speech complex tones, some harmonics were absent to create the non-speech percept. The complex tones were generated in Matlab (R2013b, MathWorks). The spectrograms of the vowels /a/ and the complex tones are shown in Fig. 1B. The foreground stimuli were presented at 75 dB sound pressure level (SPL), while the background AM noise was at 65 dB SPL. All sounds were binaurally presented to subjects via insert earphones (ER-3A, Etymotic Research) using STIM2 software (Compumedics NeuroScan). Q3

Electrophysiological recording

During data acquisition, participants were comfortably seated in an electro-acoustically shielded chamber. They were instructed to press a button in response to each deviant stimulus as quickly as possible. Continuous electroencephalography (EEG) data were recorded at a sampling rate of 10 kHz with SynAmps2 amplifier (Compumedics NeuroScan) and Curry software (7.0.9, Compumedics NeuroScan), using Ag/AgCl electrodes, simultaneously from 60 channels (FP1/2, AF3/4, F1/2/3/4/5/6/7/8, FT7/8, FC1/2/3/4/5/6, T7/8, C1/2/3/4/5/6, TP7/8, CP1/2/3/4/5/6, P1/2/3/4/5/6/7/8, PO3/4/5/6/7/8, O1/2, Fpz, Fz, FCz, Cz, CPz, Pz, POz, Oz) according to the international 10/10 system, referenced to nasal tip, with the ground electrode placed at AFz. Vertical electro-oculograms (VEOG) were also recorded for elimination of ocular artifacts. Electrode impedance was maintained below 10 kΩ.

Q1 Please cite this article in press as: Zhang X et al. Background Suppression and its Relation to Foreground Processing of Speech Versus Non-speech Streams. Neuroscience (2018), https://doi.org/10.1016/j.neuroscience.2018.01.009