



Effects of stimulus rate and noise on speech-evoked auditory brainstem responses



Les effets du rythme de présentation et du bruit sur les potentiels évoqués auditifs du tronc cérébral obtenus à l'aide de stimuli verbaux

KEY WORDS

SPEECH-EVOKED ABR

STIMULUS RATE

NOISE

AUDITORY STRESSORS

DISSOCIATION OF
NEURAL RESPONSES

Rida Al Osman
Christian Giguère
Hilmi R. Dajani

Abstract

Objective: To evaluate the effects of two auditory stressors, fast stimulus rate and/or addition of white noise, on auditory processing of speech in normal hearing subjects.

Design: Speech-evoked auditory brainstem responses (ABR) were collected using a 300 ms formant-synthesized /a/ vowel presented in four conditions: slow stimulus rate of 1.6/s in quiet, fast stimulus rate of 3.1/s in quiet, slow stimulus rate of 1.6/s with continuous white noise, fast stimulus rate of 3.1/s with continuous white noise.

Study sample: Twelve subjects (25 to 35 years old) with normal hearing thresholds.

Results: A fast stimulus rate in quiet reduced amplitude response corresponding to F1 but not to F0. The addition of white noise at the slow stimulus rate reduced amplitude of wave V and response amplitude at F1 but increased response amplitude at F0. Changing from quiet-slow to noise-fast increased response amplitude at F0 by 2.9 dB ($p < 0.05$), and decreased response amplitude at F1 and amplitude of wave V by 5.1 dB ($p < 0.01$) and 4.7 dB ($p < 0.05$) respectively. The interaction of fast stimulus rate with the addition of white noise demonstrated facilitative effects for response amplitude at F0, but occlusive or reversal effects for response amplitude at F1 and wave V amplitude.

Conclusions: This is the first study to investigate the interaction of fast stimulus rate and addition of white noise on speech-evoked ABR in human subjects. The results indicate a clear dissociation in the speech-evoked ABR between the steady-state response at F0, and that at F1 or the transient wave V. The findings advance our understanding of the complex interaction of perceptual stressors in young normal-hearing adults.

Rida Al Osman¹
Christian Giguère^{1,2}
Hilmi R. Dajani²

¹School of Rehabilitation Sciences,
Faculty of Health Sciences,
University of Ottawa,
Ottawa, ON
CANADA

²School of Electrical Engineering
and Computer Science,
Faculty of Engineering,
University of Ottawa,
Ottawa, ON
CANADA

Abrégé

Objectif : Évaluer les effets de deux facteurs de stress auditifs, soit un rythme rapide de présentation avec ou sans l'ajout de bruit blanc, sur le traitement auditif de la parole chez des sujets ayant une audition normale.

Devis : Des potentiels évoqués auditifs du tronc cérébral (PÉATC) obtenus à l'aide de stimuli verbaux furent recueillis en utilisant un stimulus d'une durée de 300 ms, correspondant à la voyelle /a/ synthétisée par formants, présenté dans quatre conditions : un rythme lent de présentation de 1,6/s dans le silence, un rythme rapide de présentation de 3,1/s dans le silence, un rythme lent de présentation de 1,6/s avec un bruit blanc continu et un rythme rapide de présentation de 3,1/s avec un bruit blanc continu.

Échantillon : Douze sujets (entre 25 et 35 ans) avec des seuils d'audition normaux.

Résultats : Un rythme rapide de présentation dans le silence a réduit l'amplitude de la réponse correspondant au F1, mais pas à la F0. L'ajout de bruit blanc sur le rythme lent de présentation a réduit l'amplitude de l'onde V et l'amplitude de la réponse au F1, mais a augmenté l'amplitude de la réponse de la F0. Le changement de la condition silence-lent à celle bruit-rapide a augmenté l'amplitude de la réponse de la F0 de 2,9 dB ($p < 0,05$), en plus de diminuer l'amplitude de la réponse du F1 et l'amplitude de l'onde V de 5,1 dB ($p < 0,01$) et 4,7 dB ($p < 0,05$), respectivement. L'interaction entre le rythme rapide de présentation et l'ajout de bruit blanc a démontré des effets facilitants pour l'amplitude de la réponse de la F0, mais des effets d'occlusion ou de renversement pour l'amplitude de la réponse du F1 et pour l'amplitude de l'onde V.

Conclusion : Il s'agit de la première étude qui explore l'interaction entre le rythme rapide de présentation et l'ajout d'un bruit blanc sur le PÉATC obtenus à l'aide de stimuli verbaux chez des sujets humains. Les résultats indiquent une dissociation évidente sur les PÉATC obtenus à l'aide de stimuli verbaux entre la réponse *auditory steady-state* de la F0 et la réponse du F1 ou l'onde transitoire V. Les découvertes font avancer notre compréhension de l'interaction complexe des facteurs de stress perceptuels chez les jeunes adultes ayant une audition normale.

Introduction

The human auditory system interprets the sounds of speech via sensory and cognitive processes, but understanding speech is not a simple task. Difficult listening conditions, such as the presence of noise, reverberation, or fast speech rate, may result in perceptual stressing of the auditory system, especially in individuals with hearing loss. The mechanisms under which the auditory system responds to these stressors and maintains robust speech representations are still not well understood (Tzounopoulos & Kraus, 2009), especially under conditions of combined or multiple auditory stressors. The present paper is concerned with the interaction of two types of auditory stressors: stimulus rate and noise.

Several authors have studied the interaction of stressors on speech perception (e.g. Adams, Gordon-Hickey, Morlas, & Moore, 2012; Adams & Moore, 2009; Raju & Alwan, 2013; Tun, 1998). Tun (1998) evaluated speech perception with sentences presented at various speech rates and different signal-to-noise ratios (SNRs) in young and older adults with normal hearing. The older group had greater difficulty than the young group in processing fast speech, especially at low SNRs. Adams et al (2012) investigated the effects of three speaking rates (slow, average, and fast) on speech perception in noise with older adults with normal hearing (ONH) and hearing impairment (OHI). The authors reported poorer speech perception for both groups of participants when the target speech was presented at a faster rate compared to the average and slower speaking rates. At all three speech rates, speech perception was significantly better for the ONH group than for the group with hearing impairment. Raju and Alwan (2013) evaluated the effects of presentation rate on the perception of consonant-vowel-consonants (CVC) in noise. The authors reported different results depending on the vowel identity. Speaking rate had the most pronounced effect on the /CuC/ stimuli, with fast speech rate being more intelligible than slow speech in noise. However, with the /CiC/ and /CaC/ stimuli, more intelligible speech in noise was found at the slow presentation rate.

Perceptual studies, such as described above, reflect the activity of the whole auditory system to speech stimuli, which involves the complex interaction of sensory, linguistic, cognitive, contextual cues, and other factors (Houtgast & Festen, 2008), in response to multiple stressors such as fast presentation rates and noise. However, they provide limited information about the internal representation of the different elements of speech, such as fundamental frequency and formants, within the auditory system, and on

their contribution to CVC discrimination and/or sentence perception. Electrophysiological studies, in contrast, provide an objective measure of the processing that occurs between the stimulus and a response from the generator site, making it possible to probe intermediate stages of the auditory pathway and the specific elements of speech being affected by a particular experimental manipulation.

Speech stimuli evoke both transient and steady-state electrophysiological responses at the level of the brainstem. In electrophysiological studies, the transient brainstem response is typically analyzed in the time domain, which involves measuring amplitudes and latencies of various peaks in the response. Apart from potentially being a neural marker of the start of a speech segment, the transient response has been shown to encode the acoustic structure at the start of the voiced consonant-vowel stop syllables /ga/, /da/, /ba/ (Johnson et al, 2008). The steady-state brainstem response, on the other hand, is typically analyzed in the frequency domain and can be categorized into the envelope-following response (EFR) and the fine structure frequency-following response (FFR). The EFR spectral components are generated primarily as a result of the non-linearities that are introduced by the rectification process of the speech envelope within the cochlea (Aiken & Picton, 2008; Cebulla, Stürzebecher, & Elberling, 2006). The EFR is commonly used to extract the evoked response that is phase-locked to the envelope of the speech stimulus which is modulated at the fundamental frequency F_0 . On the other hand, the FFR spectral content is generated as a result of auditory neural phase-locking that follows the fine structure of the speech stimulus. The FFR is used to extract the evoked response in the region of the first formant F_1 , and possibly the second formant F_2 if it is sufficiently low in frequency to allow neural phase-locking (Aiken & Picton, 2008; Prévost, Laroche, Marcoux, & Dajani, 2013).

Krizman, Skoe, and Kraus (2010) explored the effects of a single stressor, increasing stimulus rate, on auditory brainstem responses. They used both a click and a CV monosyllable (/da/) stimulus presented at three rates (6.9/s, 10.9/s, and 15.4/s). For the speech-evoked auditory brainstem response (ABR), they reported that the latency of peaks III, V, and A systematically increased in response to increasing stimulus presentation rates while latency of later peaks corresponding to the steady-state response were stable across the three presentation rates. Furthermore, they reported that the increased stimulus rate affected elements of the steady-state speech-evoked response differently; with higher frequencies (notably the F_1) being rate sensitive while lower frequencies (notably the F_0) remained rate resistant.

Some studies (e.g., Prévost et al, 2013; Russo, Nicol, Musacchia, & Kraus, 2004) explored the effects of another single stressor, background noise, using speech-evoked ABR. Prévost et al (2013) investigated the effects of different SNRs (+5, 0, -5, and -10 dB) on the transient and steady-state components of the ABR evoked by a vowel stimulus in white noise. They reported that in the presence of noise, transient response waves V and A were delayed when compared to those evoked in quiet, and the amplitude of waves V and A were strongly reduced. On the other hand, an increase in the speech-evoked ABR amplitude at F0 (but not F1) was found at all SNRs compared to the quiet condition. Similarly, Russo et al (2004) reported that the amplitude at F0 was more resistant to noise than the amplitude at F1. These results indicate that the different elements of speech may not be affected in the same direction under adverse conditions.

Several other studies (Johnson, Nicol, & Kraus, 2005; Wible, Nicol, & Kraus, 2004) have also reported a response dissociation at F0 and F1. These studies have reported that the response at F1 is diminished or delayed in children with language-based learning problems compared to normal children despite normal F0 encoding. This dissociation is also consistent with Bidelman and Krishnan (2010), who found that responses at F0 and F1 in the steady-state component of the speech-evoked ABR were affected differently under the influence of reverberation. They reported that reverberation had a minimal effect on the speech-evoked ABR at F0, but a large effect on the speech-evoked ABR at higher frequencies (notably the F1).

Using an animal model, the guinea pig, Cunningham, Nicol, King, Zecker, and Kraus (2002) evaluated the effects of multiple stressors at the level of the inferior colliculus (IC), medial geniculate body (MGB), and primary auditory cortex (AC) in response to a synthetic /ada/. Two speech rates, conversational and clear, were used in quiet and in noise. The authors reported that the onset and formant transition in the electrophysiological response were significantly degraded by the conversational rate compared to the clear rate, in noise at the IC, MGB, and AC levels. However, they found no difference across rates for the steady-state component of the response, corresponding to the vowel /a/ in the /da/ syllable, at the IC and MGB levels. They indicated that this response is conserved due to the phase-locking mechanism, which is pronounced for frequencies below 1 kHz (Assmann & Summerfield, 2004; Greenberg & Ainsworth, 2004).

To date, none of the speech-evoked ABR studies on humans have investigated the effects of multiple stressors.

Of interest, however, Burkard and Hecox (1983) investigated the effects of two stressors (noise and increased stimulus rate) using click-evoked ABR as a function of stimulus rate (15/s, 40/s, 65/s, and 90/s) and noise level (-40, 0, 10, 20, 30, and 40 dBEM), where dBEM (Effective Masking) was defined as the level of broadband noise which is just sufficient to perceptually mask a signal of the same nominal dBnHL. The authors studied the effects on wave V of the click-evoked ABR and reported that a higher noise level or a higher stimulus rate increases wave V latency. However, the combined effects were described as being occlusive, indicating that, at low noise levels, rate increases the latency of wave V but as noise level rises, the rate effect becomes less pronounced. These results raise the question as to whether the occlusive interaction found between presentation rate and noise level in the ABR response for click stimuli would also be present in the ABR transient and/or steady-state responses for speech stimuli.

The aim of this study is to investigate the interaction of two types of perceptual stressors (1) noise and (2) stimulus presentation rate on speech-evoked ABR. Responses to a synthetic vowel /a/ stimulus were recorded at two presentation rates, 3.1/s and 1.6/s, in quiet and continuous white noise conditions. For the transient response, we hypothesized that the amplitude of waves V and A would be reduced while latency of waves V and A would be increased under the effects of these two perceptual stressors. This hypothesis is consistent with single stressor studies (Prévost et al, 2013; Russo et al, 2004). For the steady-state response, we hypothesized that the combined effects of the two perceptual stressors would be different for the response amplitude at F0 and F1, as found for single stressor studies (Bidelman & Krishnan, 2010; Prévost et al, 2013; Russo et al, 2004), and that by extension, this dissociation would lead to differences in the interaction effect from the two stressors upon F0 and F1. Results from this study could therefore provide a window into subcortical processing of speech under the effects of noise and rate stressors.

SUBJECTS AND METHODS

Subjects

Twelve subjects (7 males and 5 females) participated in the study. The age of participants ranged from 25 to 35 years (mean = 29.4). All subjects had normal hearing thresholds, defined as ≤ 15 dB HL bilaterally at 0.25, 0.5, 0.75, 1, 2, and 4 kHz, and none of the subjects had a history of hearing difficulties. All subjects were compensated for their participation and provided their informed consent in compliance with a protocol approved by the University of Ottawa Research Ethics Board.

Stimuli

A synthetic vowel /a/ was generated using formant synthesis ($F_0 = 0.1$ kHz, $F_1 = 0.7$ kHz, $F_2 = 1.22$ kHz, $F_3 = 2.6$ kHz) based on a simplified version of the Klatt synthesizer (Klatt, 1980; Laroche, Dajani, Prévost, & Marcoux, 2013). This vowel was presented at 76.9 dB SPL, as measured in an ear simulator (IEC 60318-4, G.R.A.S. RA0045), in quiet or in noise, at a sampling frequency of 48 kHz and with 16-bit resolution. The vowel time-domain waveform, Hilbert envelope, and amplitude spectrum are presented in Figure 1.

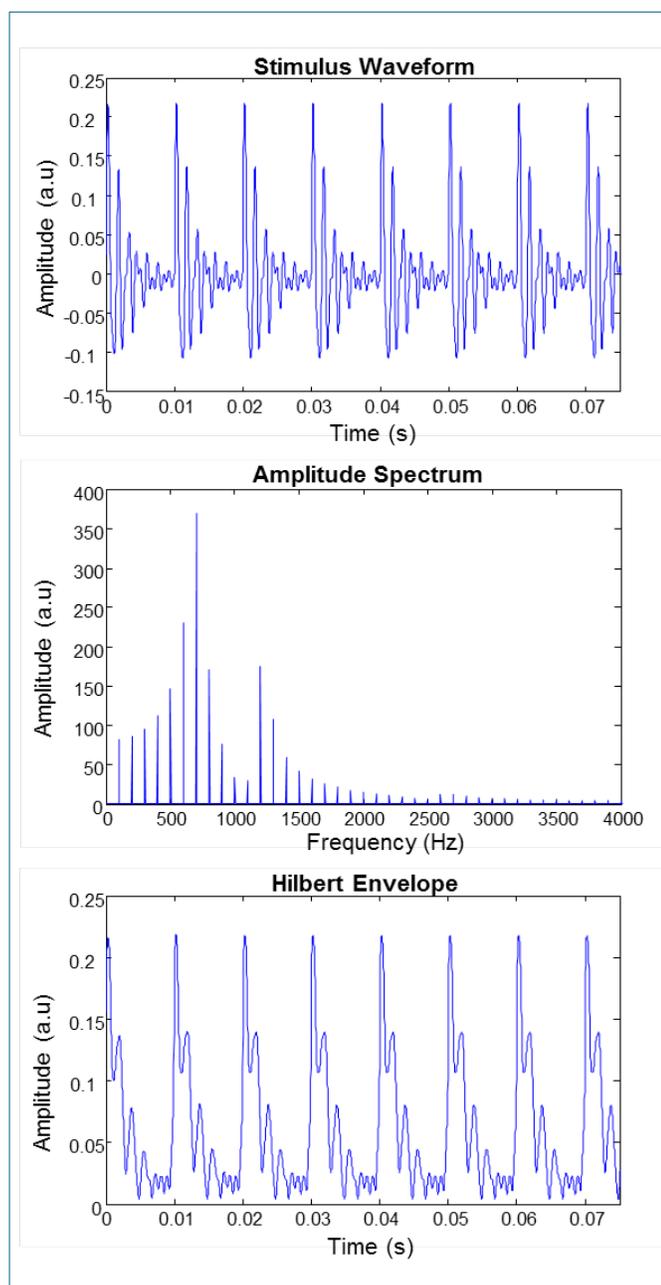


Figure 1. Time domain representation of the 300 ms /a/ vowel and its corresponding spectrum and Hilbert envelope.

The stimuli were delivered using the Bio-logic insert-earphone of the BioMARK v.7.0.2 system (Biological Marker of Auditory Processing, Biologic Systems Corp., Mundelein, IL). In the noisy condition, continuous white noise was added to the vowel and presented continuously at an SNR of 0 dB. The choice of the 0 dB SNR was based on Prévost et al (2013) who evaluated speech-evoked ABR in response to the same vowel /a/.

The study comprised four experimental conditions, in which the speech stimuli were delivered in the right ear of subjects at two different presentation rates (slow and fast) for each of two listening backgrounds (quiet and noise) (Table 1). The duration of the speech stimulus was 300 ms for both rates, but the inter-stimulus interval (ISI) was different. At the slow presentation rate (1.6/s), the ISI was 325 ms, while at the fast presentation rate (3.1/s), the ISI was 22.5 ms. Experimental conditions were presented in a pseudo-randomized order. The responses for the speech-evoked ABR were based on coherent averaging of responses to 1500 stimulus presentations in alternating polarities.

ABR recording

Measurements were performed in a shielded audiometric room with dim lighting. In order to minimize artifacts, the subjects were instructed to remain relaxed and to avoid abrupt movements while sitting in a comfortable reclining chair and watching a muted movie with subtitles. None of the subjects reported to have fallen asleep. Also, recorded sweeps in which the response exceeded $23.8 \mu\text{V}$ were discarded. In addition, we followed the recommendations of the manufacturer of BioMARK™ regarding the environmental electromagnetic noise reduction such as turning off fluorescent lights when operating the equipment, and making sure all the devices are connected with the iso-transformer provided with the system. The evoked potentials were recorded with a vertical one-channel electrode set-up. A recording electrode was placed at the vertex (Cz) and a reference electrode was placed on the right ear lobe. The ground electrode was placed on the left ear lobe. All electrode impedances were below 5 kOhm at 10 Hz. The response was amplified and filtered using an amplifier with a gain of 10,000 and a filter bandwidth extending from 30 to 1000 Hz. The evoked potentials were recorded using the BioMARK™ system over 319.8-ms epochs (1024 points/epoch corresponding to a sampling frequency of approximately 3202 Hz).

In order to ensure the absence of electromagnetic leakage contaminating the recorded responses, we replicated the ABR recording on one subject's scalp with

Table 1. The four listening conditions.

Conditions	Stimulus	Stressors
1	Quiet-slow	None
2	Quiet-fast	Rate
3	Noise-slow	Noise
4	Noise-fast	Combined

1. Slow stimulus rate of 1.6/s in quiet, 2. Fast stimulus rate of 3.1/s in quiet, 3. Slow stimulus rate of 1.6/s in continuous white noise at 0 dB SNR, 4. Fast stimulus rate of 3.1/s in continuous white noise at 0 dB SNR

the sole exception that instead of placing the foam insert earphone in the subject's right ear, it was inserted into an ear simulator (IEC 60318-4, G.R.A.S. RA0045), which presents approximately the same acoustic load to the electro-acoustic transducer as if the earphones were inserted in the ear. Spectral analysis of the recorded signal coherently averaged over 3000 trials showed that components at F0 in the EFR and at F1 in the FFR did not visually exceed the background noise, indicating no electromagnetic leakage from the sound-generating equipment to the electrodes.

Data Analysis

We analyzed both the transient and the steady-state responses to the speech stimulus. We use the terminology of Envelope Following Response or EFR (response at F0 and its early harmonics) and Frequency Following Response or FFR (response at F1) to distinguish between the responses that follow the envelope and those that follow the higher frequency fine structure. The time-domain EFR was computed by averaging the auditory brainstem responses to the original stimulus and to the inverted polarity stimulus, while the time-domain FFR was computed by averaging the responses to the original stimulus and the negative of the response to the inverted polarity stimulus (Aiken & Picton, 2008; Aiken & Purcell, 2013).

For the transient response, we followed the criteria from Skoe and Kraus (2010) for identifying peak latency and amplitude. The peak latency and amplitude were visually extracted from the time-domain EFR waveforms. The absolute peak amplitude was larger than the baseline activity recorded before the onset of the stimulus. The onset response "V" in quiet conditions started 6–10 ms following the stimulus, reflecting the time delay to the auditory brainstem, followed by negative wave "A" (Chandrasekaran & Kraus, 2010). An allowance for an increase in wave V latency was made when background

noise was present. After the peak latency and amplitude were identified, another rater who was blind to the experiment, visually extracted the peak latency and amplitude. No discrepancies were found between the peaks that were identified by us and ones that were identified by the blind rater.

For the steady-state response, the time-domain EFR and FFR waveforms were subjected to discrete Fourier transform with Matlab v.7.9 (MathWorks, Natick, MA) to extract the signal amplitude at F0 and F1 in order to evaluate the representation of speech under the effects of noise and/or stimulus rate. The signal amplitudes at F0 and F1 correspond to the height of the peak in the EFR and FFR root-mean-square (RMS) amplitude spectrum (in μV) at the frequencies of 100 Hz and 700 Hz, respectively. This was done in accordance with Aiken and Picton (2008). Examples of EFR and FFR from one subject in time and frequency domains in each condition are presented in Figures 2 and 3.

The differences in the amplitude of the EFR at F0 and FFR at F1, and in the amplitude and latency of waves V and A, across experimental conditions (Table 1), were evaluated separately with two-way repeated measures ANOVAs using SPSS, version 18.0 (SPSS Inc., Chicago, IL).

RESULTS

Results for the amplitudes and latencies corresponding to waves V and A are presented in Table 2 across all four experimental conditions. Results from the ANOVA are presented in Table 3. A background effect was observed for the amplitude of wave V ($p < .01$) and the latency of waves V ($P < .001$) and A ($p < .001$), but not for the amplitude of wave A. No rate effect was observed for the amplitude and latency of waves V and A. The interaction between background and stimulus rate showed a significant effect on the wave V amplitude only ($p < .05$).

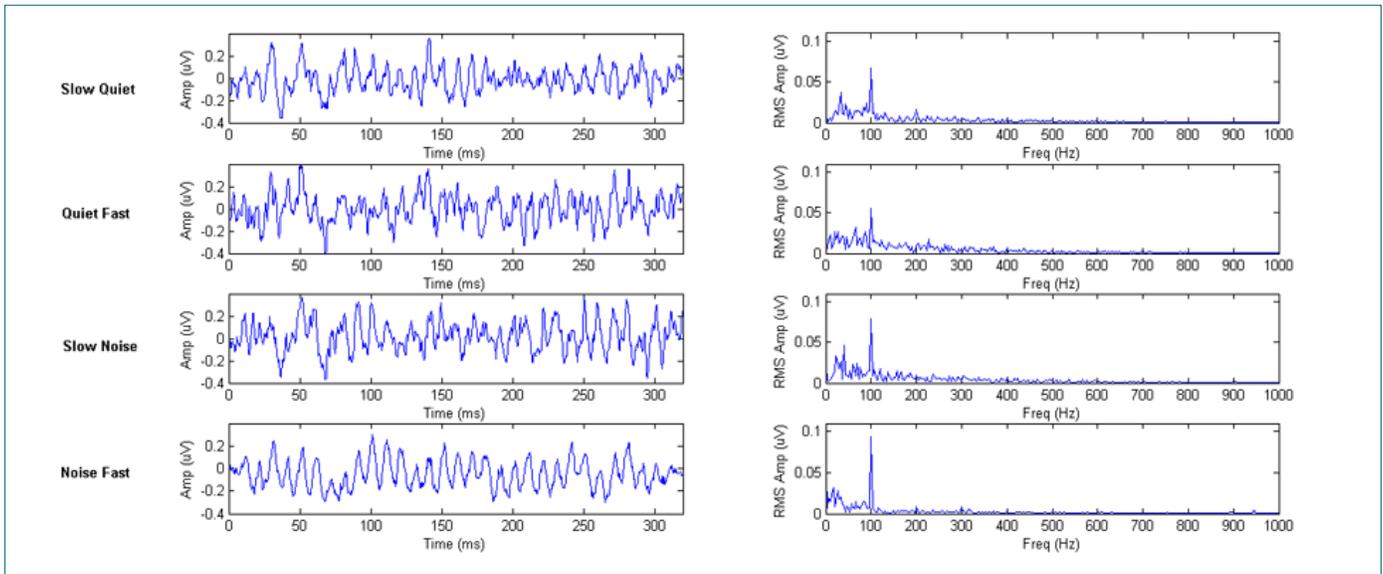


Figure 2. An example Envelope Following Response (EFR) from subject 4 in time and frequency domains in each condition. The signal amplitude at F0 corresponds to the height of the peak at 100 Hz.

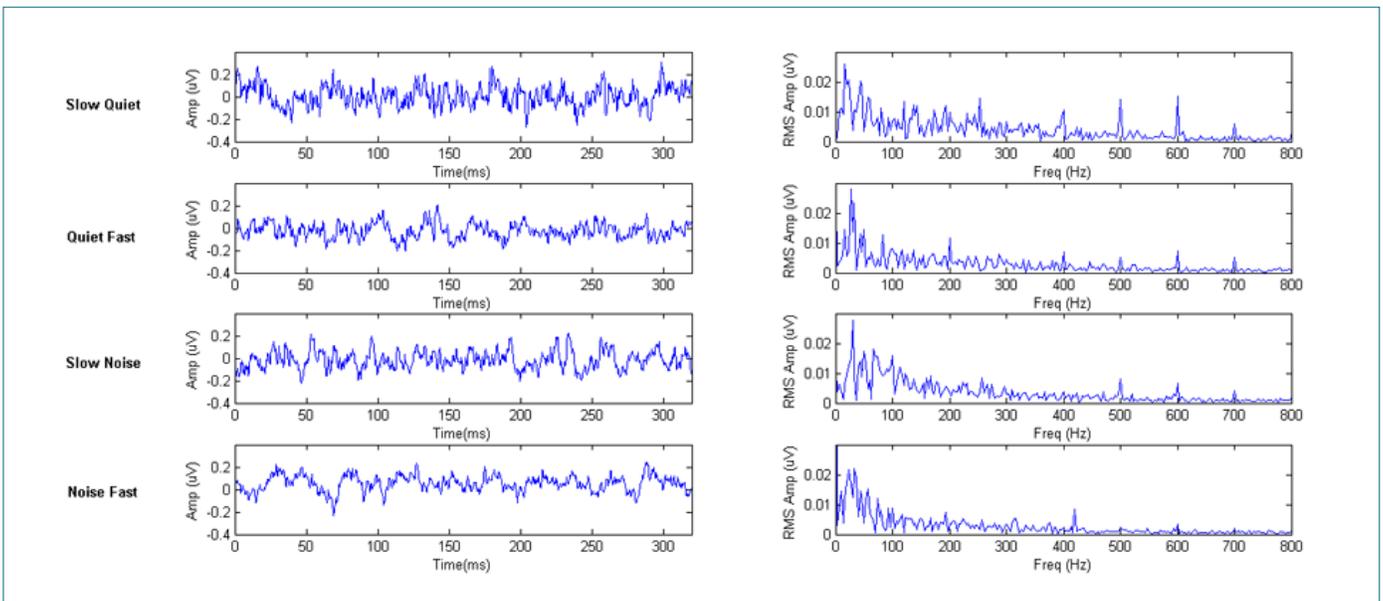


Figure 3. An example Frequency Following Response (FFR) from subject 4 in time and frequency domains in each condition. The signal amplitude at F1 corresponds to the height of the peak at 700 Hz.

Table 2. Mean (standard error) for the latency and amplitude of waves V and A over all subjects in each condition.

	Wave V latency [ms]	Wave A latency [ms]	Wave V amplitude [uV]	Wave A amplitude [uV]
Quiet-slow	7.42 (0.23)	9.16 (0.32)	0.31 (0.05)	-0.08 (0.02)
Quiet-fast	7.73 (0.82)	9.52 (0.84)	0.25 (0.03)	-0.1 (0.04)
Noise-slow	10.07 (0.42)	11.76 (0.42)	0.11 (0.02)	-0.08 (0.03)
Noise-fast	10.28 (0.49)	12.28 (0.59)	0.18 (0.03)	-0.04 (0.02)

Table 3. Statistical results from two-way repeated measures ANOVA for latency and amplitudes of waves V and A, and for amplitudes EFR and FFR.

	F	p
V (latency)		
Background (Quiet vs Noisy)	59.895	0.000 ***
Rate (Slow vs Fast)	0.559	0.470
Interaction (Background x Rate)	0.024	0.881
A (latency)		
Background (Quiet vs Noisy)	70.463	0.000 ***
Rate (Slow vs Fast)	0.94	0.353
Interaction (Background x Rate)	4.485	0.834
V (amplitude)		
Background (Quiet vs Noisy)	17.144	0.002 **
Rate (Slow vs Fast)	0.059	0.813
Interaction (Background x Rate)	8.311	0.015 *
A (amplitude)		
Background (Quiet vs Noisy)	2.222	0.164
Rate (Slow vs Fast)	0.039	0.847
Interaction (Background x Rate)	3.098	0.106
EFR (amplitude)		
Background (Quiet vs Noisy)	1.519	0.058 +
Rate (Slow vs Fast)	4.47	0.243
Interaction (Background x Rate)	5.173	0.044 *
FFR (amplitude)		
Background (Quiet vs Noisy)	22.535	0.001 **
Rate (Slow vs Fast)	0.777	0.397
Interaction (Background x Rate)	12.96	0.004 **

***p < 0.001; **p < 0.01; *p < 0.05; +p < 0.1. The degree of freedom is (3, 20).

The interaction of stressors found for wave V amplitude was further explored using post-hoc pairwise t-tests between experimental conditions. The results are found in Table 4, where the significance of changes in amplitude for single and multiple stressor conditions are reported. The effect sizes are expressed in dB as 20 times the logarithm of the amplitude ratio of the two conditions under test. This was calculated from the mean amplitude values listed in Table 2. For single stressors Quiet (slow to fast) and Slow (quiet to noise), the wave V amplitude trends towards a significant decrease by 1.9 dB and significantly decreases by 9.0 dB, respectively. For multiple stressors Fast (quiet to noise), Noise (slow to fast), and quiet-slow to noise-fast, the wave V amplitude changes correspond to a significant 2.9 dB decrease, a significant 4.3 dB increase, and a significant 4.7 dB decrease, respectively.

The amplitudes corresponding to F0 (labeled as EFR amplitude) and to F1 (labeled as FFR amplitude) across all experimental conditions are presented in Figure 4. Results from the ANOVA are presented in Table 3. A background effect or trend was observed for the EFR amplitude ($p = .058$) and the FFR amplitude ($p = .001$). On the other hand, no rate effect was observed for the EFR and FFR amplitudes. A significant interaction between background and stimulus rate was found for both the EFR amplitude ($P < .05$) and the FFR amplitude ($p < .01$).

The interaction of stressors found for EFR and FFR amplitudes was further explored using post-hoc pairwise t-tests between experimental conditions. The results are found in Table 4, where the significance of changes in amplitude for single and multiple stressors are reported. These changes were derived from Figure 4. For single stressor situations, the EFR amplitude shows a non-significant decrease by 0.3 dB for Quiet (slow to fast) and a non-significant increase by 1.3 dB for Slow (quiet to noise). On the other hand, the FFR amplitude significantly decreases by 2.3 dB for Quiet (slow to fast) and by 6.7 dB for Slow (quiet to noise). For multiple stressors, the EFR amplitude significantly increases by 3.3 dB for Fast (quiet to noise), significantly increases by 1.6 dB for Noise (slow to fast), and significantly increases by 2.9 dB for quiet-slow to noise-fast. In contrast, the FFR amplitude significantly decreases by 2.8 dB for Fast (quiet to noise) and by 5.1 dB for quiet-slow to noise-fast while it shows a non-significant increase by 1.6 dB for Noise (slow to fast) (Table 4).

DISCUSSION

The ability to communicate in the presence of auditory stressors is an important task for successful participation

in educational, social, and vocational environments. In this study, we explored the effects of two stressors, namely noise and fast stimulus rate, on the brainstem electrophysiological response to speech. These stressors were evaluated separately (effects of fast stimulus rate or noise) as well as jointly (combined effects of stimulus rate and noise). Although these stressors are fundamentally different, they represent realistic forms of signal degradation (Krizman et al, 2010; Prévost et al, 2013). Moreover, these stressors can differentially affect the processing of the different elements of speech such as fundamental frequency and formants. Results from this study could therefore provide a window into subcortical processing of speech under the effects of these stressors.

Single stressors

Quiet (slow-to-fast):

The EFR amplitude decreased by only 0.3 dB as a result of going from slow to fast stimulus rate conditions while the FFR amplitude decreased by 2.3 dB, as shown in Table 4. These results indicate a dissociation between responses at F0 and F1 with increasing stimulus rate in quiet. Such dissociation between responses at F0 and F1 was also reported by Krizman et al (2010). They indicated that the amplitude corresponding to F1 decreased systematically with the rate increase while the amplitude corresponding to F0 remained stable. They also found an increase in the wave V latency with increasing rate. In our study, the change in latency was very small, but the wave V amplitude decreased by 1.9 dB.

Slow (quiet-to-noise):

The EFR amplitude increased by 1.3 dB as a result of going from quiet to noise conditions while the FFR amplitude decreased by 6.7 dB, as shown in Table 4. These results again show a dissociation between responses at F0 and F1, this time with added noise. Such dissociation between responses corresponding to F0 and F1 is consistent with previous studies. Johnson et al (2005), Kraus and Nicol (2005), Prévost et al (2013), and Russo et al (2004) also reported a dissociation between F0 and F1 with the addition of noise, where the higher frequencies (including F1) were diminished despite normal - or in the case of Prévost et al (2013) enhanced - F0 encoding. Furthermore, in our study, the wave V amplitude decreased by 9 dB. This result is consistent with Russo et al (2004) and Prévost et al (2013) who reported a degradation of the transient response waves in the presence of noise.

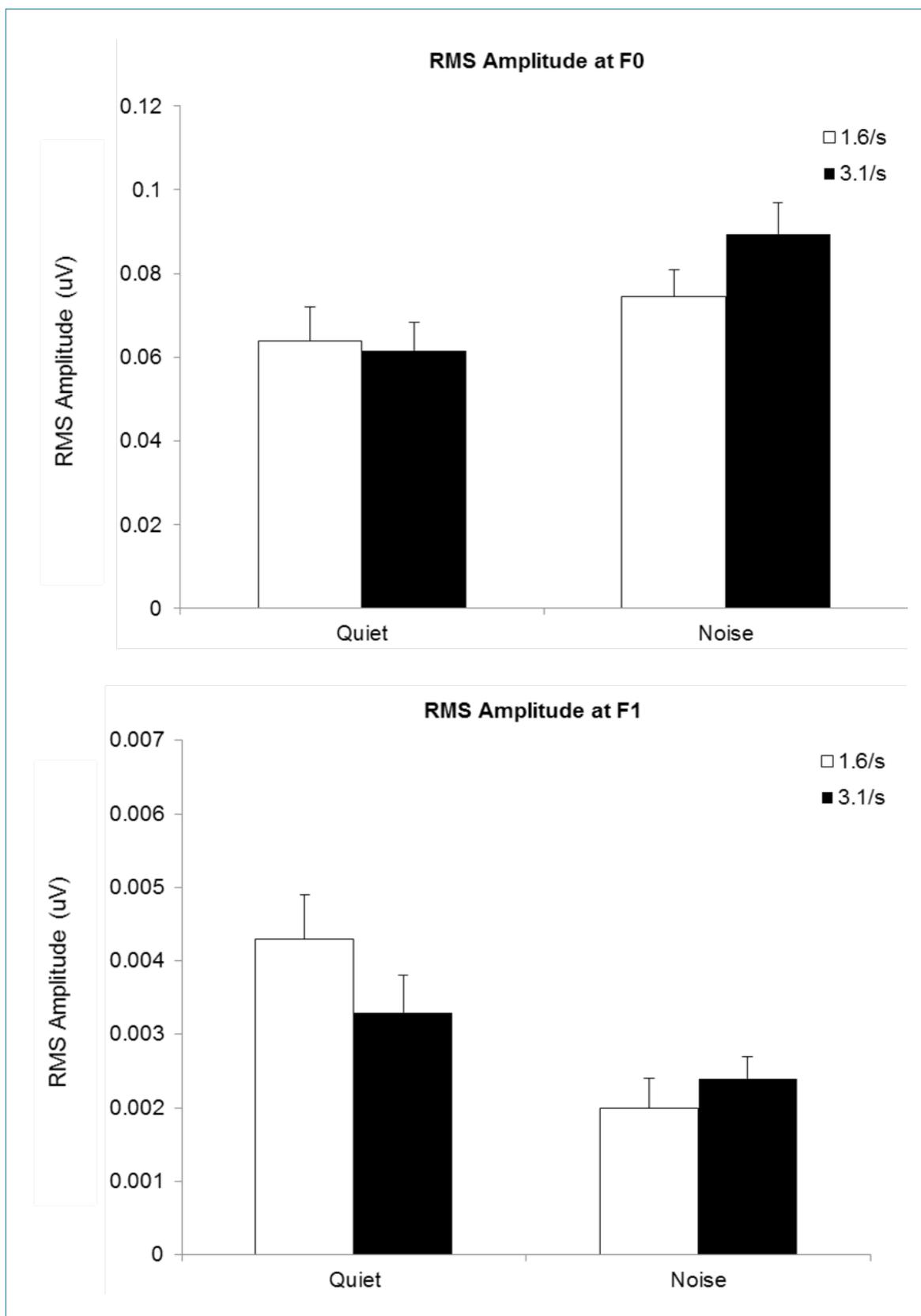


Figure 4. Comparison of response amplitudes in µV (mean and standard error) at F0 and at F1 for the two presentation rates (1.6/s and 3.1/s) in quiet and in noise.

Table 4. Mean change (standard error) of the transient response of wave V amplitude and EFR and FFR amplitudes over all subjects, and p-values obtained with post-hoc pairwise t-test comparisons between conditions. The type of effect (Facilitative, Reversal, Occlusive) is also indicated.

		V Amplitude change [dB]	F0 Amplitude change [dB]	F1 Amplitude change [dB]
Single Stressor	Quiet (slow to fast)	-1.9 (0.9) +	-0.3 (0.4)	-2.3 (0.6) *
	Slow (quiet to noise)	-9.0 (2.1) ***	1.3 (0.6)	-6.7 (1.3) ***
Multiple Stressors	Fast (quiet to noise)	-2.9 (0.26) * OE	3.3 (0.1) ** FE	-2.8 (1.3) * OE
	Noise (slow to fast)	4.3 (1) * RE	1.6 (0.3) ** FE	1.6 (0.6) RE
	Quiet-slow to Noise-fast	-4.7 (1.1) ** OE	2.9 (0.2) + FE	-5.1 (1.9) ** OE

FE denotes Facilitative Effect. RE denotes Reversal Effect. OE denotes Occlusive Effect. ***p < 0.001; **p < 0.01; *p < 0.05; +p < 0.1.

Multiple stressors

Fast (quiet to noise):

The wave V amplitude and the FFR amplitude showed a lesser decrease at the fast rate compared to the slow rate as a result of going from quiet to noise, (-2.9 dB vs. -9.0 dB) and (-2.8 dB vs. -6.7 dB), respectively. These results indicate that the effect of the baseline stressor (fast rate) provided an occlusive effect in the presence of the added stressor (noise). On the other hand, the EFR amplitude showed an increase by 3.3 dB. This increase is somewhat surprising in that the effect of going from quiet to noise is larger with fast rate than with slow rate for F0 amplitude (3.3 dB vs. 1.3 dB). This result indicates that the baseline stressor (fast rate) provided a facilitative effect in the presence of the added stressor (noise). The facilitative effect at F0 and occlusive effect at F1 and wave V of the baseline stressor (fast rate) are presented in Table 4.

Noise (slow to fast):

The wave V amplitude and the FFR amplitude showed a reversal in the direction of the effect in noise compared to quiet as a result of going from a slow to fast stimulus rate (4.3 dB vs. -1.9 dB) and (1.6 dB vs. -2.3 dB), respectively. On the other hand, the EFR amplitude showed an increase by 1.6 dB. This increase is again somewhat surprising in that the effect of increasing the stimulus rate is larger in noise than in quiet (1.6 dB vs. -0.3 dB). These results indicate that the baseline stressor (noise) produced a facilitative effect in the presence of the added stressor (fast rate). The facilitative effect at F0, reversal effect at F1 and wave V are presented in Table 4.

Quiet-slow to Noise-fast:

The wave V amplitude and the FFR amplitude showed a lesser decrease in comparison to the combined effects of the two single stressors (Quiet (slow to fast) and Slow (quiet to noise)), with changes of (-4.7 dB vs. -10.9 dB) and (-5.1 dB vs. -9 dB), respectively. These results indicate that the two stressors interacted in an occlusive fashion. On the other hand, the EFR amplitude showed a higher increase in comparison to the combined effects of the two single stressors (Quiet (slow to fast) and Slow (quiet to noise)), with changes of 2.9 dB vs. 1 dB. This result indicates that the two stressors (noise and fast rate) interacted in a facilitative fashion, resulting in an increase in the F0 response. The facilitative effect for F0 and occlusive effect for F1 and wave V amplitude are reported in Table 4.

Overall, the results for the multiple stressors show an essential dissociation between EFR and FFR amplitude changes (facilitative effect for EFR vs. occlusive or reversal effect for FFR). The results also show an association between FFR and wave V amplitudes changes (occlusive or reversal effect for both responses).

Relationship to other studies

The dissociation between responses corresponding to F0 on the one hand and responses corresponding to F1 and wave V of the transient response on the other hand might be explained in terms of a source and filter model of auditory processing of speech. Johnson et al (2005) suggested that specific components of the brainstem response reflect source (i.e., F0 and harmonics of the envelope) and filter (i.e., formants) stimulus

characteristics separately and along two separate neural streams. Results from Bidelman and Krishnan (2010) also supported this hypothesis. The authors reported that the neural representation of the filter related components are significantly degraded with reverberation while the neural representation of the source information remained relatively unchanged under the same conditions. However, this two-stream model has not been firmly established and the exact neural mechanisms are unclear.

Krizman et al (2010) reported a differential effect of increasing stimulus rate on the responses that correspond to F1 compared to those that correspond to F0, and concluded that these results support the involvement of different neural streams. The authors also proposed that this differential effect likely reflects an interaction of neural adaptation, neural fatigue, and refractory properties of individual nerve fibers, resulting in a desynchronization of the response at high stimulus rates (Hall, 1992).

Auditory brainstem responses to speech sounds are shaped by both the acoustic characteristics of the incoming speech signal and cognitive processes such as attention and memory (Galbraith, Bhuta, Choate, Kitahara, & Mullen, 1998). Specifically, auditory selective attention helps to extract relevant signal elements from competing background noise and stores them in working memory (Johnson & Zatorre, 2005). These steps enable top-down effects, thus enhancing the brainstem encoding of relevant and predictable features (pitch, timing, and harmonics) (Chandrasekaran & Kraus, 2010). Speech processing is related to a reliable transmission of speech in the brainstem (Parbery-Clark, Skoe, & Kraus, 2009) and to cognitive functions such as auditory attention that, although they are thought to take place in the cortex, may have top-down effects on brainstem processing (Anderson & Kraus, 2010).

Accordingly, the facilitative effect on the EFR amplitude when the two stressors are present could depend on active mechanisms that include top-down feedback, which in turn is "interrupted" with the longer inter-stimulus gaps found at the slower stimulus rate (i.e., 1.6/s). Hocherman and Gilat (1981) investigated the effects of inter-stimulus gaps using responses of single units in the primary auditory cortex. The durations of the stimulus and the ISI were varied (50-ms stimuli were presented at intervals of either 550 ms or 900 ms, while 100-ms stimuli were presented at intervals of either 900 ms or 1,600 ms). The authors reported that varying the ISI from 550 ms to 900 ms for the 50-ms stimuli or from 900 ms to 1,600 ms for the 100-ms stimuli resulted in a similar increase in evoked activity (67.5% and 67.1%, respectively). Furthermore, they reported that randomly

mixing two ISIs (550 and 900 ms) for the 50-ms stimuli or (900 ms and 1,600 ms) for the 100-ms stimuli caused a reduction in evoked activity in 29% of the units and an increase in 14%. The remaining 57% of the units did not follow a consistent reduction or increase in evoked activity. The authors speculated that the responses to varying the ISI could relate to cortical mechanisms that enable an evoked activity to such changes by some attentive mechanisms.

The neural basis of top-down attentional control of auditory processing at lower levels, such as the auditory brainstem and cochlea, was also investigated by Rinne et al (2008) using functional magnetic resonance imaging (fMRI). The authors suggested that auditory processing at the brainstem level is top-down modulated via selective attention. As reported in Du, Kong, Wang, Wu, and Li (2011), attentional top-down control of auditory processing is performed through enhancing synchronous phase-locked activities of brainstem neurons to behaviorally relevant stimulus. In our experiment we did not control for selective attention, but based on the reports in the literature (Anderson & Kraus, 2010; Bidelman & Krishnan, 2010), speech-evoked auditory brainstem responses may be useful measures for investigating how perceptual/cognitive cues can assist in selectively targeting a speech signal and improving recognition in the presence of multiple auditory stressors. To improve recognition, the processing at the level of the brainstem could involve internal noise suppression, signal enhancement, or a combination of both (Prévost et al, 2013).

Enhancement of the subcortical response that follows the envelope has been found in normal listeners who undergo auditory training for pitch discrimination (Carcagno & Plack, 2011) and for speech recognition in noise (Song, Skoe, Banai, & Kraus, 2012). Moreover, long-term experience with music or a tonal language has been found to correspond to a more robust subcortical representation of the pitch frequency (Bidelman, Gandour, & Krishnan, 2011; Wong, Skoe, Russo, Dees, & Kraus, 2007). The gain in the envelope-following evoked responses that we observed with multiple stressors could provide an electrophysiological substrate underlying noise robustness of normal hearing listeners with conversational speech. Furthermore, the strength of the subcortical representation of the envelope in normal listeners has been suggested to be a primary contributor to speech perception in noise (Song et al, 2012; Swaminathan & Heinz, 2012).

In hearing-impaired listeners, on the other hand, enhancement of the response at F0 appears to be associated with a degraded perceptual ability. Anderson,

Parbery-Clark, White-Schwoch, Dreihobl, and Kraus (2013) have evaluated the effects of sensorineural hearing loss (SNHL) on older adults using speech-evoked ABR and have reported a greater spectral representation for the response that follows the envelope in the SNHL group compared to an aged matched normal hearing group, and equivalent representation for the response corresponding to the temporal fine structure in the signal. The authors suggested that the perceptual deficit in the SNHL group might be related to an imbalanced subcortical representation of speech, with dominance of the neural activity that follows the signal envelope relative to the activity that follows the rapidly-varying temporal fine structure. Another plausible explanation may be related to hearing impaired listeners having widened auditory filters (Sharma & Chaudhari, 2013). As such, unlike in normal listeners, their auditory filters (particularly the lower frequency narrow filters) pass complex signals which are modulated at F0 because they combine multiple (instead of individual) harmonics of the fundamental, resulting in a stronger response at F0.

The effects of the multiple stressors on waves V amplitude and FFR amplitude showed an occlusive or reversal effect in the present study. For Fast (quiet to noise) and Quiet-slow to Noise-fast, both the wave V amplitude and the FFR amplitude showed an occlusive effect. This result is similar to what Burkard and Hecox (1983) found, where the effect of increased rate and addition of noise on wave V of the click-evoked response was purely occlusive. On the other hand, the Noise (slow to fast) effect on wave V amplitude and FFR amplitude showed a reversal effect. This result may be due to mechanisms similar to those proposed above for the facilitative effect with EFR.

The combined effects of fast rate and noise are complex, resulting in an interaction, rather than a summation. The neural mechanisms for such interaction are not fully defined yet. Further studies are required to reveal the neurophysiologic and cognitive processes involved in encoding speech stimuli under effects of perceptual stressors such as fast rate and noise.

CONCLUSION

This is the first study to investigate the interaction of auditory stressors such as fast stimulus rate and/or addition of white noise on the auditory system in humans using speech-evoked ABR. Perceptual studies have investigated the combined effects of these stressors. Such studies characterize the response of the whole auditory system, which involves the complex interaction of sensory, linguistic, cognitive, aging, contextual cues, and other factors. In our study, speech-evoked ABR provided an objective means to

assess processing at an intermediate level of the auditory system and probe into the neural representation of specific elements of speech.

For single stressors, our findings are consistent with previous studies that found a dissociation between responses corresponding to F0 and F1 under effects of either noise or increased stimulus rate. What differentiates this study from preceding work is the evaluation of how multiple stressors can interact together. This study has shown an essential dissociation between responses corresponding to F0 on the one hand and responses corresponding to F1 and wave V on the other hand. The two stressors have shown facilitative effects on the responses corresponding to F0, while they have shown occlusive or reversal effects on the responses corresponding to F1 and wave V. The facilitative effect found for F0 is particularly interesting and unexpected. However, the current study cannot determine whether this effect corresponds to a natural mechanism in normal listeners that underlies robustness in difficult environments, or whether it is an indicator of a degradation in the sub-cortical representation of a speech as found in studies with hearing impaired listeners in single stressor situations.

The encoding of the neural responses corresponding to F0 and F1 is important for both recognizing the speech content and identifying the speaker and voice emotion. Therefore, the recording of speech-evoked ABR with multiple stressors could be useful for investigating the neural mechanisms underlying how speech perception and recognition is achieved under difficult acoustic conditions. Further investigation into the effects of other auditory stressors will help reveal the physiological mechanisms that are selectively enhanced or diminished. Future work needs to evaluate the effects of age (younger vs. older adults) and hearing status (normal vs. impaired) on the interaction of rate and noise as well as the interaction between other auditory stressors.

References

- Adams, E. M., Gordon-Hickey, S., Morlas, H., & Moore, R. (2012). Effect of rate-alteration on speech perception in noise in older adults with normal hearing and hearing impairment. *American Journal of Audiology, 21*(1), 22-32.
- Adams, E. M., & Moore, R. E. (2009). Effects of speech rate, background noise, and simulated hearing loss on speech rate judgment and speech intelligibility in young listeners. *Journal of the American Academy of Audiology, 20*(1), 28-39.
- Aiken, S. J., & Picton, T. W. (2008). Envelope and spectral frequency-following responses to vowel sounds. *Hearing Research, 245*(1), 35-47.
- Aiken, S. J., & Purcell, D. (2013). Sensitivity to stimulus polarity in speech-evoked frequency-following responses. In *Proceedings of Meetings on Acoustics, 19*(1)
- Anderson, S., & Kraus, N. (2010). Sensory-cognitive interaction in the neural encoding of speech in noise: A review. *Journal of the American Academy of Audiology, 21*(9), 575-585.
- Anderson, S., Parbery-Clark, A., White-Schwoch, T., Dreihobl, S., & Kraus, N. (2013). Effects of hearing loss on the subcortical representation of speech cues. *The Journal of the Acoustical Society of America, 133*(5), 3030-3038.
- Assmann, P., & Summerfield, Q. (2004). The perception of speech under adverse conditions. In A. N. Popper & R. R. Fay (Eds.), *Speech processing in the auditory system* (pp. 231-308). New York: Springer.
- Bidelman, G. M., & Krishnan, A. (2010). Effects of reverberation on brainstem representation of speech in musicians and non-musicians. *Brain Research, 1355*, 112-125.
- Bidelman, G. M., Gandour, J. T., & Krishnan, A. (2011). Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. *Journal of Cognitive Neuroscience, 23*(2), 425-434.
- Burkard, R., & Hecox, K. (1983). The effect of broadband noise on the human brainstem auditory evoked response. I. Rate and intensity effects. *The Journal of the Acoustical Society of America, 74*(4), 1204-1213.
- Carcagno, S., & Plack, C. (2011). Subcortical plasticity following perceptual learning in a pitch discrimination task. *Journal of the Association for Research in Otolaryngology, 12*(1), 89-100.
- Cebulla, M., Stürzebecher, E., & Elberling, C. (2006). Objective detection of auditory steady-state responses: Comparison of one-sample and q-sample tests. *Journal of the American Academy of Audiology, 17*(2), 93-103.
- Chandrasekaran, B., & Kraus, N. (2010). The scalp-recorded brainstem response to speech: Neural origins. *Psychophysiology, 47*(2), 236-246.
- Cunningham, J., Nicol, T., King, C. D., Zecker, S. G., & Kraus, N. (2002). Effects of noise and cue enhancement on neural responses to speech in auditory midbrain, thalamus and cortex. *Hearing Research, 169*, 97-111.
- Du, Y., Kong, L., Wang, Q., Wu, X., & Li, L. (2011). Auditory frequency-following response: A neurophysiological measure for studying the "cocktail-party problem". *Neuroscience & Biobehavioral Reviews, 35*(10), 2046-2057.
- Galbraith, G. C., Bhuta, S. M., Choate, A. K., Kitahara, J. M., & Mullen, T. A. (1998). Brain stem frequency-following response to dichotic vowels during attention. *Neuroreport, 9*(8), 1889-1893.
- Greenberg, S., & Ainsworth, W. A. (2004). Chapter 1: Speech processing in the auditory system: An overview. In A. N. Popper & R. R. Fay (Eds.), *Speech processing in the auditory system* (pp. 1-63). New York, NY: Springer.
- Hall, J. W. (1992). *Handbook of auditory evoked responses*. Boston: Allyn & Bacon.
- Houtgast, T., & Festen, J. M. (2008). On the auditory and cognitive functions that may explain an individual's elevation of the speech reception threshold in noise. *International Journal of Audiology, 47*(6), 287-295.
- Hocherman, S., & Gilat, E. (1981). Dependence of auditory cortex evoked unit activity on interstimulus interval in the cat. *Journal of Neurophysiology, 45*(6), 987-997.
- Johnson, K. L., Nicol, T., & Kraus, N. (2005). The brainstem response to speech: A biological marker. *Ear and hearing, 26*(5), 424-434.
- Johnson, K. L., Nicol, T., Zecker, S. G., Bradlow, A. R., Skoe, E., & Kraus, N. (2008). Brainstem encoding of voiced consonant-vowel stop syllables. *Clinical Neurophysiology, 119*(11), 2623-2635.
- Johnson J. A., & Zatorre R. J. (2005). Attention to simultaneous unrelated auditory and visual events: Behavioral and neural correlates. *Cerebral Cortex, 15*(10), 1609-1620.
- Klatt, H. D. (1980). Software for a cascade/parallel formant synthesizer. *The Journal of the Acoustical Society of America, 67*(3), 971-995.
- Kraus, N., & Nicol, T. (2005). Brainstem origins for cortical "what" and "where" pathways in the auditory system. *Trends in Neurosciences, 28*(4), 176-181.
- Krizman, J., Skoe, E., & Kraus, N. (2010). Stimulus rate and subcortical auditory processing of speech. *Audiology & Neurotology, 15*(5), 332-342.
- Laroche, M., Dajani, H. R., Prévost, F., & Marcoux, A. M. (2013). Brainstem auditory responses to resolved and unresolved harmonics of a synthetic vowel in quiet and noise. *Ear Hearing, 34*(1), 63-74.
- Parbery-Clark A., Skoe E., & Kraus N. (2009). Musical experience limits the degradative effects of background noise on the neural processing of sound. *The Journal of Neuroscience, 29*(45), 14100-14107.
- Prévost, F., Laroche, M., Marcoux, A. M., & Dajani, H. R. (2013). Objective measurement of physiological signal-to-noise gain in the brainstem response to a synthetic vowel. *Clinical Neurophysiology, 124*(1), 52-60.
- Raju, A., & Alwan, A. (2013). The effect of speaking rate, vowel context, and speaker intelligibility on the perception of consonant vowel consonants in noise. *The Journal of the Acoustical Society of America, 134*(5), 4031-4031.
- Rinne, T., Balk, M. H., Koistinen, S., Autti, T., Alho, K., & Sams, M. (2008). Auditory selective attention modulates activation of human inferior colliculus. *Journal of Neurophysiology, 100*(6), 3323-3327.
- Russo, N., Nicol, T., Musacchia, G., & Kraus, N. (2004). Brainstem responses to speech syllables. *Clinical Neurophysiology, 115*(9), 2021-2030.
- Skoe, E., & Kraus, N. (2010). Auditory brainstem response to complex sounds: A tutorial. *Ear and Hearing, 31*(3), 302-324.
- Sharma S. D., & Chaudhari, D. S. (2013). Speech processing for sensorineural hearing impairment: A review. *International Journal of Advanced Research in Computer Science and Software Engineering, 3*(3), 710-712.
- Song, J. H., Skoe, E., Banai, K., & Kraus, N. (2012). Training to improve hearing speech in noise: Biological mechanisms. *Cerebral Cortex, 22*, 1180-1190.
- Swaminathan, J., & Heinz, M. G. (2012). Psychophysiological analyses demonstrate the importance of neural envelope coding for speech perception in noise. *The Journal of Neuroscience, 32*(5), 1747-1756.
- Tun, P. A. (1998). Fast noisy speech: Age differences in processing rapid speech with background noise. *Psychology and Aging, 13*(3), 424-434.
- Tzounopoulos, T., & Kraus, N. (2009). Learning to encode timing: Mechanisms of plasticity in the auditory brainstem. *Neuron, 62*(4), 463-469.
- Wible, B., Nicol, T., & Kraus, N. (2004). Atypical brainstem representation of onset and formant structure of speech sounds in children with language-based learning problems. *Biological Psychology, 67*(3), 299-317.
- Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience, 10*(4), 420-422.

Declaration of Interest

The authors report no conflict of interest. The authors alone are responsible for the content and writing of this paper.

Acknowledgements

This work was funded partially by the Natural Sciences and Engineering Research Council of Canada.

Authors' Note

Correspondence concerning this article should be addressed to Rida Al Osman at the School of Rehabilitation Sciences, Faculty of Health Sciences, University of Ottawa, 451 Smyth Road, Ottawa, ON K1H 8M5, CANADA Email: ralos010@uottawa.ca.