

A Study of Auditory Speech Processing Using Brainstem Evoked Responses Under the Effects of Stressors

Rida Al Osman

Thesis submitted to the Faculty of Graduate and Postdoctoral Studies in partial fulfillment of the requirements for the Doctorate in Philosophy degree in Rehabilitation Sciences

Faculty of Health Sciences

School of Rehabilitation Sciences

University of Ottawa

Abstract

This work investigated the effects of various stressors on auditory speech processing using speech-evoked auditory brainstem response (ABR). This work was carried out in three distinct studies: (1) the effects of stimulus rate and noise on the speech-evoked ABR, (2) the effects of direct, early, and later-arriving speech reflections on the speech-evoked ABR, and (3) the effects of self-masking and overlap masking in a reverberant environment using speech-evoked ABR.

Results from the first study indicate a complex interaction of fast stimulus rate and noise on the speech-evoked ABR. In particular, dissociation was found between electrophysiological responses to the envelope and the responses to the temporal fine structure or transient waves. The main finding from the second study is that direct sound and early reflections produce similar speech-evoked ABR, while the response to late reverberation shows different characteristics. Results from the third study indicate that severe reverberation significantly weakened the ABR when compared to anechoic or moderate reverberation, while moderate reverberation significantly enhanced the ABR when compared to anechoic speech.

Overall, the new findings from this research shed light on the effects of stressors such as noise, increased stimulus rate, and reverberation. Up until now, these effects have mostly been studied perceptually. Our research demonstrates that these effects can be studied electrophysiologically using speech-evoked brainstem responses. This method of investigation permits the probing of particular stages of the auditory pathway and evaluating how neural activity in response to specific elements of speech, such as the fundamental frequency and formants, is affected by different stressors.

Acknowledgments

I would like to thank my supervisors Drs. Christian Giguère and Hilmi Dajani for all their unwavering guidance, support, and prompt feedback. Their sustained encouragement has been invaluable in my endeavour to manage competing demands of school, employment, and family life.

I would like to thank all the participants who took part in my studies.

I would like to thank my brothers Hussein and Firas for their limitless support and encouragement.

Table of Contents

Chapter 1. Introduction.....	1
1.1. Background.....	1
1.2. Speech-evoked ABR.....	3
1.2.1. Speech-evoked ABR in quiet.....	4
1.2.2. Speech-evoked ABR to fast stimulus rate and noise.....	5
1.2.3. Speech-evoked ABR in reverberation	7
1.3. Objectives	9
1.4. Structure	10
1.5. Statement of Contribution to Jointly Authored Works Contained in the Thesis	11
Chapter 2. Effects of stimulus rate and noise on speech-evoked auditory brainstem responses (study #1).....	13
2.1. Introduction.....	14
2.2. Subject and Methods	19
2.2.1. Subjects	19
2.2.2. Stimuli.....	19
2.2.3. ABR recording.....	20
2.2.4. Data Analysis.....	22
2.3. Results	24
2.4. Discussion.....	29
2.4.1. Single stressors	30
2.4.2. Multiple stressors	31
2.4.3. Relationship to other studies	32
2.5. Conclusion	36
2.6. Declaration of Interest	38
Chapter 3. On the effects of direct, early, and late-arriving speech reflections using speech-evoked auditory brainstem response (study #2).....	39
3.1. Introduction.....	40
3.2. Methods	45
3.2.1. Subjects	45
3.2.2. Room Impulse Response	45

3.2.3. Stimuli.....	47
3.2.4. Recording and analysis of speech-evoked ABR.....	49
3.2.5. Behavioral Test.....	51
3.3. Results.....	52
3.3.1. Transient Response.....	52
3.3.2. Steady-state Response.....	52
3.3.3. Behavioral Response.....	52
3.4. Discussion.....	55
3.5 Conclusion.....	61
3.6 Declaration of Interest.....	62
Chapter 4. Effects of self-masking and overlap masking in a reverberant environment using speech-evoked auditory brainstem response (study #3).....	63
4.1. Introduction.....	64
4.2. Methods.....	67
4.2.1. Subjects.....	67
4.2.2. Room Impulse Response.....	68
4.2.3. Stimuli.....	69
4.2.4. Recording and analysis of speech-evoked ABR.....	70
4.3. Results.....	72
4.4. Discussion.....	74
4.6 Declaration of Interest.....	78
Chapter 5. General Discussion.....	79
5.1. Original findings.....	79
5.2. Top-down effects.....	84
5.3. Limitations.....	88
5.4 Future work.....	88
References.....	92
Appendix A: Informed Consent Form.....	109

List of Figures

Chapter 2

Figure 1. Time domain representation of the 300 msec /a/ vowel (top) and its corresponding spectrum and Hilbert envelope (bottom).22

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.26

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.....26

Figure 4. Comparison of response amplitudes in uV (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.27

Chapter 3

Figure 1. Representation of the full room impulse response (RIR) showing the direct sound, early reflections, and late reflections. The RIR was produced for a room of dimensions 3.5m×11m×14m and for source and listener coordinate positions of (2m, 2.5m, 1.5m) and (2m, 8m, 1.2m), respectively.46

Figure 2. Time-domain waveforms and amplitude spectra of the four stimuli (Direct, Early, Full, Late). Only the first 50 ms is shown.48

Figure 3. Average of EFR amplitudes (combination of the amplitudes at F0, 2F0, and 3F0 in the EFR spectrum) across subjects, and of FFR amplitudes (combination of the amplitudes at 5F0 to 9F0 in the FFR spectrum) across subjects, for each stimulus (Direct, Early, Full, Late). Error bars indicate standard errors of the mean.53

Figure 4. Average of individual latency (ms) and amplitudes (uV) of waves A and V for each stimulus (Direct, Early, Full, Late). Error bars indicate standard errors of the mean.53

Figure 5. Grand Average Speech-evoked ABR in time-domain and frequency domain of the four stimuli (Direct, Early, Late, and Full).....59

Chapter 4

Figure 1. The aggregate stimulus (/boy fell from the window / + 40 ms gap + vowel /a/) convolved with the RIR for all three reverberation conditions: anechoic, moderate reverberation, and severe reverberation are presented from top to bottom.69

Figure 2. The stimuli corresponding to the reverberation conditions: anechoic, moderate reverberation, and severe reverberation are presented from top to bottom. Only the first 50 ms is shown.70

Figure 3. Average of EFR amplitudes (combination of the amplitudes at F0, 2F0, and 3F0 in the EFR spectrum) across subjects, and of FFR amplitudes (combination of the amplitudes at 5F0 to 9F0 in the FFR spectrum) across subjects, for each stimulus (anechoic, moderate reverberation, and severe reverberation). Error bars indicate standard errors of the mean. ..73

Figure 4. Average of individual latency (ms) and amplitudes (uV) of waves A and V for each stimulus (anechoic, moderate reverberation, and severe reverberation). Error bars indicate standard errors of the mean.74

List of Tables

Chapter 2

Table 1. The four listening conditions. 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.....27

Table 2. Mean (standard error) for the latency and amplitude of waves V and A over all subjects in each condition.28

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.28

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 conditions29

Chapter 3

Table 1. Statistical results from one-way repeated measures ANOVA for latency and amplitudes of waves V and A, and for EFR and FFR amplitudes54

Table 2. Post-hoc HSD tests for latency and amplitudes of waves V and A, and for EFR and FFR amplitudes54

Chapter 4

Table 1. Statistical results from one-way repeated measures ANOVA for latency and amplitudes of waves V and A, and for EFR and FFR amplitudes74

Table 2. Post-hoc HSD tests for latency and amplitudes of waves V and A, and for EFR and FFR amplitudes74

List of Acronyms

ABR: Auditory brainstem response
AC: Auditory cortex
ANOVA: Analysis of Variance
dB SPL: Decibels (Sound Pressure Level)
dB HL: Decibels (Hearing Level)
CVC: Consonant-vowel-consonant
CV: Consonant-vowel
EEG: Electroencephalogram
EM: Effective masking
EFR: Envelope-following response
FFR: Fine structure frequency-following response
FFT: Fast Fourier transform
fMRI: Functional magnetic resonance imaging
F0: Fundamental (or Pitch) Frequency
F1: First formant
F2: Second formant
IC: Inferior colliculus
ISI: Inter-stimulus interval
MGB: Medial geniculate body
ONH: Older adults with normal hearing
OHI: Older adults with hearing impairment
RIR: Room impulse response
RMS: Root-mean-square
RT: Reverberation time
SNHL: Sensorineural hearing loss
SNR: Signal-to-noise ratio

Chapter 1. Introduction

Speech communication is essential for human interaction, but is often challenged by auditory stressors such as the presence of background noise (Pichora-Fuller, 2003; Wagener and Brand, 2005), fast speech rate (Tun, 1998; Adams et al., 2012), reverberation (Neuman et al., 2010; Hazrati and Loizou, 2012), and transmission channel distortion (Houtgast and Steeneken, 1985; Assman and Summerfield, 2004). In addition, the presence of hearing loss (Anderson et al. 2013), learning disorders (Ruggles et al., 2011), speech/language pathologies (Kraus, 1999) and/or non-native talkers (Dame et al. 2013) can also pose a challenge to speech communication.

1.1. Background

Several perceptual studies have investigated the effects of single stressors on speech perception. For instance, Festen and Plomp (1990) evaluated speech perception with normal-hearing listeners in various types of background noise presented at 80 dBA SPL using speech-reception threshold (SRT). The background noise included steady-state noise, modulated noise, and a single competing voice. The authors reported that for normal-hearing listeners the SRT for sentences increased by 4-8 dB with the different types of noise. Speech perception in noise is typically exacerbated in listeners with hearing impairment. For example, Kenyon et al (1998) evaluated speech perception using a speech discrimination test in noise and in quiet for normal hearing listeners and listeners with at least a 50-dB loss above 3000 Hz. The authors reported a 33% and 5% decrease in perception in noise (relative to quiet) for the listeners with and without hearing impairments, respectively. Similarly, several studies reported that non-native listeners show a sharper decline than native listeners in performance on speech perception tasks with increasing levels of signal distortion, either through the addition of background noise or

reverberation (e.g., Nábělek and Donahue, 1984; Mayo et al. 1997; Meador et al. 2000). Other perceptual studies investigated the effects of more than one auditory stressor on speech perception: fast presentation rates and noise (e.g. Adams and Moore, 2009; Adams et al. 2012; Raju and Alwan, 2013), and reverberation and noise (e.g. Hazrati and Loizou, 2012; Wróblewski et al. 2012). With exception to Raju and Alwan (2013), the authors reported poorer speech perception to fast speech rate in noise compared to slow speech rate in quiet (Adams and Moore, 2009; Adams et al. 2012). Similarly, the authors reported poorer speech perception under combined effects of noise and reverberation compared to noise or reverberation alone (Hazrati and Loizou, 2012; Wróblewski et al. 2012). Overall, these perceptual studies show that the combined effects of auditory stressors add up.

Perceptual studies reflect the activity of the whole auditory system and other brain systems to speech stimuli, and involve the complex interaction of sensory, linguistic, cognitive, contextual cues and other factors (Houtgast and Festen, 2008). However, these studies provide limited information about the internal representation of the different elements of speech, such as fundamental frequency and formants, within the auditory system.

Electrophysiological studies, in contrast, provide an objective measure of the processing that occurs between the stimulus and a response site, making it possible to probe intermediate stages of the auditory pathway and specific elements of the internal representation of speech being affected by a particular experimental manipulation.

The auditory brainstem response (ABR) provides a particularly useful objective measure of the auditory function. It has been employed in the evaluation of hearing thresholds in infants,

children, and adults individuals who are difficult to test, in assessment of auditory neuropathy, and in screening for retrocochlear dysfunction (Anderson and Kraus, 2013). In addition, the ABR is relatively easy to record and interpret compared to the cortical response.

1.2. Speech-evoked ABR

Early studies of the ABR employed simple stimuli such as clicks and sinusoidal tones. More recent studies have employed complex stimuli (e.g., speech, music, and environmental sounds). In particular, the use of speech stimuli can provide additional information about neural encoding compared to clicks (Song et al. 2006). The click stimulus is short in duration, while the speech stimulus is rich in both temporal and spectral content. Speech stimuli are also directly related to everyday communication tasks, thus highlighting the potential for the speech-evoked ABR to provide a very effective and relevant means of assessing auditory system function (Johnson et al. 2005; Anderson and Kraus, 2010).

Dajani et al. (2005) and Dajani and Giguère (2010) conducted two studies that further illustrate the potential of the speech-evoked response in investigating the auditory processing of speech. In the first study, the authors recorded evoked responses to a natural vowel and an altered natural vowel, in which the fundamental frequency was suppressed (Dajani et al., 2005). The evoked response to the altered vowel at the fundamental frequency F_0 was found to be as strong as it was for the unaltered vowel, and its contour also closely followed the acoustic pitch contour. These findings suggest that the measurement of pitch-evoked responses could be a marker for objective speech audiometry. In the second study, the authors compared the amplitude contours of the brainstem evoked response to a natural vowel with normal hearing subjects with the amplitude contours of the broadband vowel stimulus, and vowel filtered around F_1 , F_2 , F_3 , and

F4 (Dajani and Giguère, 2010). The study found a strong correlation of the amplitude contours of the brainstem evoked response and amplitude contours of vowel filtered around F3 and F4 compared to broadband signal and signal filtered around F1 and F2. The authors suggested that representations characterizing the amplitude perturbations in speech, or shimmer, are processed differently in the different formant regions.

Speech stimuli evoke both transient and steady-state responses at the level of the brainstem. The transient brainstem response is typically analyzed in the time domain, which involves measuring amplitude and latency of various peaks (V and A being most notable) in the response. The steady-state brainstem response 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 is obtained from the average between the response to the original stimulus and the response generated from the inverted polarity stimulus (Aiken and Picton, 2008), and follows the speech envelope. As a result, the EFR is primarily related to the responses at F0 and its lower harmonics. On the other hand, the FFR follows the actual harmonics in the stimulus, and not the envelope. The FFR is obtained by subtracting the average response generated using the inverted polarity stimulus from the average response obtained using the original polarity stimulus (Aiken and Picton, 2008). The FFR is primarily concerned with the responses around the formant frequencies, which in this study is restricted to the response in the region of F1.

1.2.1. Speech-evoked ABR in quiet

There have been several studies with speech-evoked ABR done in quiet conditions. Krishnan (2002) employed synthetic vowels /a/ and /u/ with normal hearing young adults. The duration of

each vowel was 60 ms. His results suggested that the human ABR shows clear and distinct peaks corresponding to formant frequencies of synthetic vowels. Subsequent studies by Krishnan and colleagues employed Chinese syllables to show that pitch representation in the ABR is based on temporal patterns of phase-locked neural activity of the fundamental frequency (Krishnan et al., 2005). These studies by Krishnan and others such as Aiken and Picton (2006) showed that speech-evoked ABR in quiet provides a way to access subcortical auditory processing and may be used as a biological marker of speech encoding in adverse listening conditions.

1.2.2. Speech-evoked ABR to fast stimulus rate and noise

Krivan et al. (2010) investigated effects of increasing stimulus rate on auditory processing using speech-evoked ABR. The authors employed a monosyllable (/da/) stimulus presented at three rates (6.9/s, 10.9/s, and 15.4/s) in quiet. They found that the latency of peaks III, V, and A of the transient response increased in response to increasing stimulus presentation rates while the latency of later peaks corresponding to the steady-state response were stable across the various 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 F1) being rate sensitive while lower frequencies (notably the F0) remained rate resistant.

Some studies (e.g. Russo et al. 2004; Prévost et al. 2013) also investigated effects of white noise on auditory processing using speech-evoked ABR. These studies found that the amplitude of peaks V and A, of the transient response, decreased and latencies of these peaks increased with noise. Moreover, they reported that the presence of background noise affected elements of the steady state speech-evoked ABR response differently; with higher frequencies (notably the F1)

being background sensitive while lower frequencies (notably the F0) remained robust (or increased in the case of Prévost et al. (2013).

There is a need to investigate the interaction of increased stimulus rate and noise, since such multiple auditory stressors may be present in real acoustic environments. Cunningham et al. (2002) evaluated this interaction at the level of the inferior colliculus (IC), medial geniculate body (MGB) and primary auditory cortex (AC) of anaesthetised guinea pigs in response to a synthetic /ada/. The authors reported a significant degradation for the onset and formant transition at the IC, MGB and AC levels corresponding to the vowel /a/ in the /da/, by the conversational rate compared to the clear rate, in noise. On the other hand, they reported no difference for the steady-state component of the response under the same conditions. Burkard and Hecox (1983) also investigated this interaction using click-evoked ABR with normal hearing subjects 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 reported that wave V latency of the click-evoked ABR increases with higher noise level or higher stimulus rate. However, the interaction of the increased stimulus rate and noise was described as being occlusive, indicating that, at low noise levels, rate increases wave V latency but as noise level rises, the rate effect becomes less pronounced. These results raise the question as to whether the occlusive interaction in the ABR response for click stimuli would also be present in the ABR transient and/or steady-state responses for speech stimuli. To date, none of the speech-evoked ABR studies have made this investigation yet.

1.2.3. Speech-evoked ABR in reverberation

Reverberation is defined as the sum of all sound reflections off the walls and objects in an enclosed setting (Blauert, 1997). The reflections arriving within the first 50 ms after the direct sound are often called early reflections and so are distinguished from late reflections (Bradley et al. 2003; Arweiler and Bucholz, 2011). The series of direct sound, early reflections, and late reflections are typically characterized by the room impulse response (RIR) (Blauert, 1997).

Reverberation alters the acoustical properties of the speech signals reaching the ear in face-to-face communication situations. It can cause degradation of speech signals in different ways, such as temporal smearing, filling dips and gaps in the temporal envelope, increasing the prominence of low-frequency energy, and flattening formant transitions (Assman and Summerfield, 2004; Sayles and Winter, 2008). Such degradation of speech signals results from masking of a speech segment by reflections of the segment itself, known as self-masking, and by reflections of previous segments, known as overlap-masking (Devore and Shinn-Cunningham, 2003; Kokkinakis and Loizou, 2011). Such deleterious effects of reverberation can detrimentally affect speech processing in real rooms. Therefore, it is important to better understand the consequences of reverberation at different stages of auditory processing.

On the other hand, the effect of reverberation on speech perception is not always detrimental as, in contrast to anechoic conditions, it also amplifies the speech signal level at a distance from the talker in real rooms. Perceptual studies have established the importance of such amplification using speech intelligibility test scores, in conditions similar to real rooms for both impaired and non-impaired subjects (e.g. Bradley et al. 2003; Yang and Bradley, 2009; Arweiler and

Buchholz, 2011). These studies showed that for situations where the direct sound is reduced, it is possible to understand the speech signal due to the presence of the early reflections.

Only a very limited number of studies have investigated the effects of reverberation on speech stimuli using electrophysiological methods. Bidelman and Krishnan (2010) investigated the effects of reverberation on speech-evoked ABR using a 250 ms synthetic vowel /i/ with normal hearing adults. Their results indicated that reverberation had little effect on the neural encoding of pitch, but significantly neural encoding of formant-related harmonics. Fujihira and Shiraishi (2015) investigated effects of reverberation using a 170 ms synthetic syllable /da/. Their results indicated that speech intelligibility under reverberation in elderly listeners is related to their ability to encode the temporal fine structure of speech, which is challenged under effects of reverberation.

It is important to note that in Bidelman and Krishnan (2010) and in Fujihira and Shiraishi (2015) the methodological choices and/or experimental design focused on the detrimental contribution of reverberation on speech communication. In Bidelman and Krishnan (2010), the amplitude of the reverberant speech stimulus was normalized to the same overall level as the non-reverberated condition, thus negating the natural amplification that typically occurs in reverberant conditions. In Fujihira and Shiraishi (2015), the early reflection components from the RIR were removed, thus eliminating the benefit they may have on speech perception. For a more comprehensive understanding of the effects of reverberation on speech in real rooms, all components of reverberation (direct sound vs. early reflections vs. late reflections vs. full reflections) need to be investigated. To our knowledge, no study has conducted this kind of investigation using the speech-evoked ABR.

Moreover, in Bidelman and Krishnan (2010) and in Fujihira and Shiraishi (2015), the stimuli included only the self-masking effect. In order to understand all the effects of reverberation on speech, both self-masking and overlap-masking need to be investigated. To our knowledge, no study has investigated the overlap masking effects using the speech-evoked ABR.

1.3. Objectives

In this thesis, we aim to evaluate the effects of various auditory stressors on auditory speech processing using speech-evoked ABR. The focus is on three gaps in knowledge (1) interaction of increased stimulus rate and noise, (2) effects of the different components (direct vs. early reflections vs. late reflections) of reverberation on speech, (3) effects of self-masking and overlap-masking of reverberation on speech. The objectives of the thesis can be summarized as follows:

- Study #1: We explore the effect of the interaction of increased stimulus rate and noise on auditory processing using speech-evoked ABR. Previous speech-evoked ABR studies explored effects of increased stimulus rate and noise on auditory processing, separately.
- Study #2: We explore effects of various components of reverberation (direct sound vs. early reflections vs. late reflections vs. full RIR reflections) on auditory processing using speech-evoked ABR. Previous speech-evoked ABR studies focused on the detrimental effect of reverberation on auditory processing.

- Study #3: We explore the self-masking and overlap-masking effects of reverberation on auditory processing using speech-evoked ABR. Previous speech-evoked ABR studies only explored the effects of self-masking on auditory processing.

The thesis is designed to add to the body of literature that describes how stressors such as noise and increased rate, interact in affecting auditory processing of speech. Furthermore, this research can lead to an improved understanding of the effects of the various components of reverberation (direct sound vs. early reflections vs. late reflections vs. full RIR reflections), and effects of effects of self-masking and overlap-masking of reverberation on auditory speech processing.

Moreover, this research may provide guidance on the importance of accounting for the perceptual stressors and the composition of reverberation (direct vs. early reflections vs. late reflections) and reverberation times into the development of clinical speech-evoked response testing to provide an accurate predictor of real-world performance in real rooms.

1.4. Structure

The thesis is article-based and is organized as follows. The next three chapters describe the three studies that were carried out in this thesis (1) Effects of stimulus rate and noise on speech-evoked auditory brainstem responses, (2) Effects of direct, early, and late-arriving speech reflections on the auditory brainstem response, and (3) Effects of self-masking and overlap-masking of reverberated speech on the auditory brainstem response. These chapters include a detailed literature review relevant to each study along with a description of the methodology and summary of the results. The final chapter includes a general discussion that draws together the findings from the three studies, limitations of this research, and suggestions for future work.

1.5. Statement of Contribution to Jointly Authored Works Contained in the Thesis

1. Al Osman, R., Dajani, H., Giguère, C. (2015). Encoding of reverberated speech using auditory brainstem response. 8th Annual CAA Conference and Exhibition, 2 (16).

Rida Al Osman was primarily responsible for concept development, data collection, analysis, reviewing drafts and editing. Christian Giguère contributed to concept development, analysis, reviewing drafts and editing. Hilmi Dajani contributed to reviewing drafts and editing.

2. Al Osman, R., Giguere, C., Dajani, H. (2016). Effects of stimulus rate and noise on speech-evoked auditory brainstem responses. Canadian Journal of Speech-Language Pathology and Audiology, 40(2), 104-119.

Rida Al Osman was primarily responsible for concept development, data collection, analysis, reviewing drafts and editing. Hilmi Dajani contributed to concept development, analysis, reviewing drafts and editing. Christian Giguère contributed to reviewing drafts and editing.

3. Al Osman, R., Dajani, H., Giguere, C. On the effects of direct, early, and late-arriving speech reflections using auditory brainstem response (**submitted**).

Rida Al Osman was primarily responsible for concept development, data collection, analysis, reviewing drafts and editing. Christian Giguère contributed to concept development, analysis, reviewing drafts and editing. Hilmi Dajani contributed to reviewing drafts and editing.

4. Al Osman, R., Dajani, H., Giguere, C. Effects of self-masking and overlap masking in a reverberant environment using speech-evoked auditory brainstem response (**to be submitted**).

Rida Al Osman was primarily responsible for concept development, data collection, analysis, reviewing drafts and editing. Christian Giguère and Hilmi Dajani contributed to concept development, analysis, reviewing drafts and editing.

Chapter 2. Effects of stimulus rate and noise on speech-evoked auditory brainstem responses (study #1)

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 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.

keywords: speech-evoked ABR; stimulus rate; noise; auditory stressors; dissociation of neural responses.

2.1. 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 and 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. Tun, 1998; Adams and Moore, 2009; Adams et al., 2012; Raju and Alwan, 2013). 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 and 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 (Cebulla et al., 2006; Aiken and Picton, 2008). 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 F0. 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 F1, and possibly the second formant F2 if it is sufficiently low in frequency to allow neural phase-locking (Aiken and Picton, 2008; Prévost et al., 2013).

Krizman et al (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 (speech-evoked 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 F1) being rate sensitive while lower frequencies (notably the F0) remained rate resistant.

Some studies (e.g. Russo et al., 2004; Prévost et al., 2013) 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 (Wible et al., 2004; Johnson et al., 2005) 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 et al (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 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 and Summerfield, 2004; Greenberg and 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 (Russo et al., 2004; Prévost et al., 2013). 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 (Russo et al., 2004; Bidelman and Krishnan, 2010; Prévost et al., 2013), 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.

2.2. Subject and Methods

2.2.1. 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.

2.2.2. Stimuli

A synthetic vowel /a/ was generated using formant synthesis (F0 = 0.1 kHz, F1 = 0.7 kHz, F2 = 1.22 kHz, F3 = 2.6 kHz) based on a simplified version of the Klatt synthesizer (Klatt, 1980; Laroche et al., 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.

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 ISI was different. At the slow presentation rate (1.6/s), the ISI was 325ms, 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.

2.2.3. 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 μ 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 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.

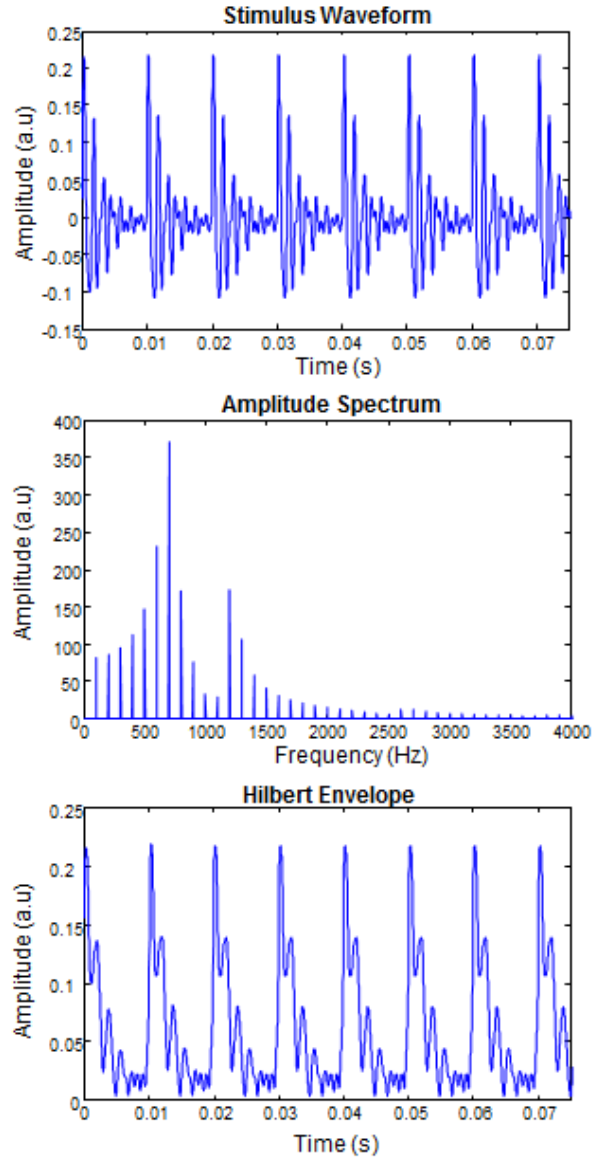


Fig1. Time domain representation of the 300 ms /a/ vowel (top) and its corresponding spectrum and Hilbert envelope (bottom).

2.2.4. 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 and Picton, 2008; Aiken and 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 and 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 study 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).

2.3. 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 ($p < 0.01$) and the latency of waves V ($p < 0.001$) and A ($p < 0.001$), 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 ($p < 0.05$).

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 ($p = 0.058$) and the FFR amplitude ($p = 0.001$). 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 ($p < 0.05$) and the FFR amplitude ($p < 0.01$).

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).

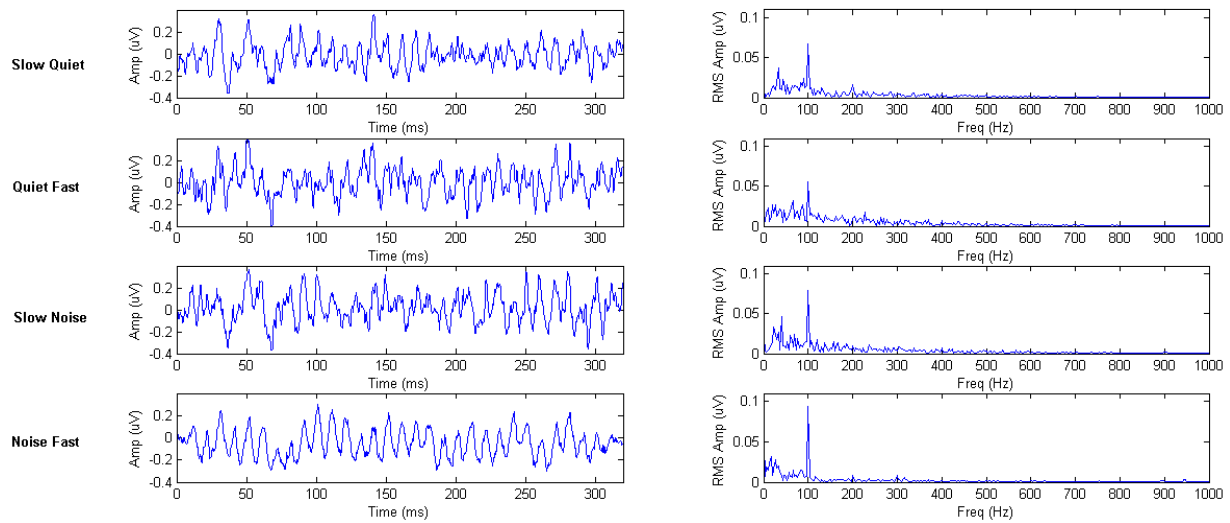


Fig. 2. An example Envelope Following Response (EFR) from subject 4 in time and frequency domains in each condition. The signal amplitude at F0 corresponds to the height of the peak at 100 Hz.

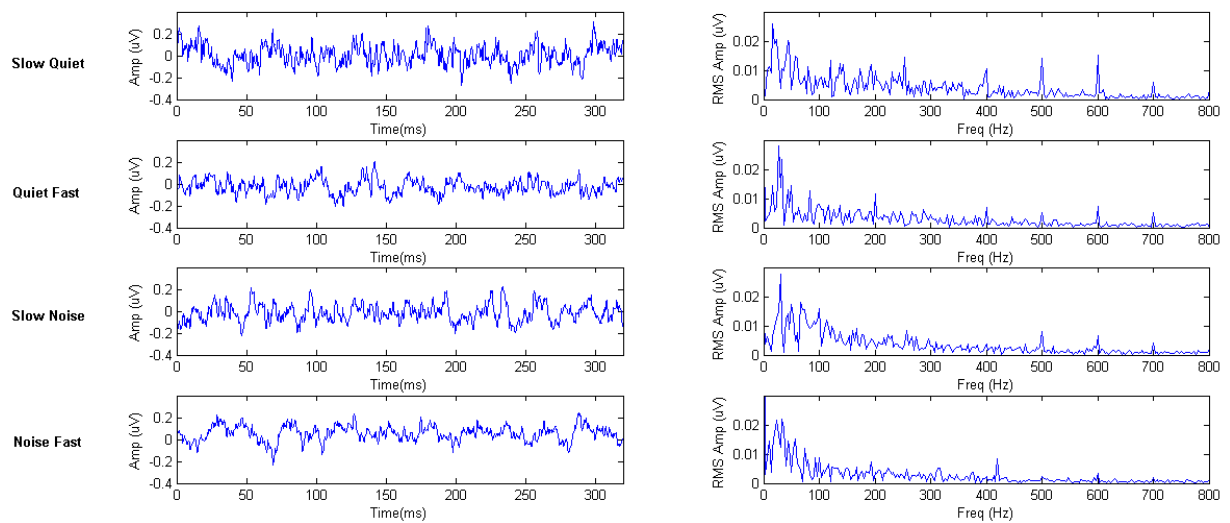


Fig. 3. An example Frequency Following Response (FFR) from subject 4 in time and frequency domains in each condition. The signal amplitude at F1 corresponds to the height of the peak at 700 Hz.

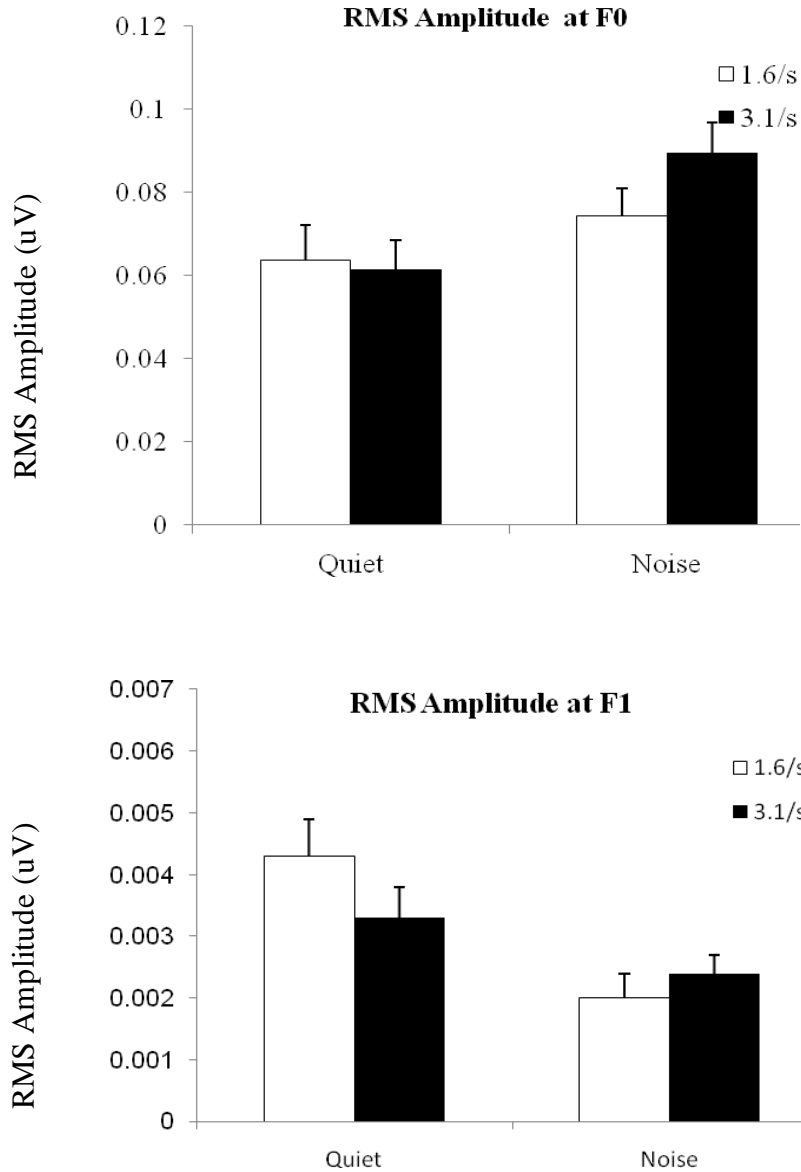


Fig. 4. Comparison of response amplitudes in uV (mean and standard error) at F0 and at F1 for the two presentation rates (1.6/s and 3.1/s) in quiet and in noise.

Table 1. The four listening conditions. 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

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

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

	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 **

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, and Occlusive) is also indicated. FE denotes Facilitative Effect. RE denotes Reversal Effect. OE denotes Occlusive Effect. ***p < 0.001; **p < 0.01; *p < 0.05; +p < 0.1.

		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

2.4. 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.

2.4.1. Single stressors

2.4.1.1. 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.

2.4.1.2. 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. Russo et al (2004), Johnson et al (2005), Kraus and Nicol (2005) and Prévost et al (2013) 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.

2.4.2. Multiple stressors

2.4.2.1. 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.

2.4.2.2. 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 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.

2.4.2.3. 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).

2.4.3. 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 et al., 1998). Specifically, auditory selective attention helps to extract relevant signal elements from competing background noise and stores them in working memory (Johnson and Zatorre, 2005). These steps enable top-down effects, thus enhancing the brainstem encoding of relevant and predictable features (pitch, timing, and harmonics) (Chandrasekaran and Kraus, 2010). Speech processing is related to a reliable transmission of speech in the brainstem (Parbery-Clark et al., 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 and 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 inter-stimulus interval (ISI) were varied (50 ms stimuli were presented at intervals of either 550 or 900 ms, while 100 ms stimuli were presented at intervals of either 900 or 1,600 ms). The authors reported that varying the ISI from 550 to 900 ms for the 50 ms stimuli or from 900 to 1,600 for the 100-ms stimuli resulted in 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 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 et al (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 study we did not control for selective attention, but based on the reports in the literature (Anderson and Kraus, 2010; Bidelman and 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 and Plack, 2011) and for speech recognition in noise (Song et al., 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 (Wong et al., 2007; Krishnan et al., 2009). 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 and 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 et al (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 SHNL 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

and 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.

2.5. 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, ageing, 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.

2.6. Declaration of Interest

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

Chapter 3. On the effects of direct, early, and late-arriving speech reflections using speech-evoked auditory brainstem response (study #2)

Objective: In perceptual studies, the direct wave and early reflections of reverberated speech have been found to constitute useful energy, while the late reflections constitute detrimental energy. This study investigated how various components of reverberated speech are encoded in the auditory system using the speech-evoked ABR.

Design: 12 subjects with normal hearing participated in this study. Speech-evoked ABRs were recorded using reverberant stimuli created as a result of the convolution between a 2-s synthetic vowel /a/ and each of the following room impulse response (RIR) components: direct sound, early reflections, late reflections, and full reverberation. Four stimuli were produced: direct component, early component, late component, and full component.

Results: For waves V and A amplitudes, a significant difference or a tendency towards significance was found between direct and late components, between direct and full components, and between early and late components. For waves V and A latencies, significant differences were found between direct and late components, between direct and full components, between early and late components, and between early and full components. For the Envelope Frequency Response and Fine Structure Frequency amplitudes, a significant difference or tendency towards significance was found between direct and late components, between direct and full components, and between early and late components. Moreover, in 67%, 27%, and 6% of the trials of a behavioural test, the early, full and late stimuli were, in decreasing order, the most perceptually similar to the direct stimulus.

Conclusion: Speech-evoked ABR to direct and early reflection components possess similar characteristics while responses to direct and early reflections possess different characteristics from responses to late reflections. Overall, these results show that effects of the different components of RIRs can be studied electrophysiologically, thereby providing another means than perceptual studies to investigate the effects of room acoustics on auditory processing.

3.1. Introduction

Speech communication is essential in human daily interaction, but it is often hindered by perceptual stressors such as background noise (Pichora-Fuller, 2003; Wagener and Brand, 2005), fast speech rate (Tun, 1998; Adams et al. 2012), reverberation (Neuman et al. 2010; Wróblewski et al. 2012), and transmission channel distortion (Houtgast and Steeneken, 1985; Assman and Summerfield, 2004). In particular, room reverberation alters the acoustical properties of the speech signals reaching our ears in situations of face-to-face communication, which can affect speech understanding (Crandell and Smaldino, 2000; Bradley et al., 2003). Therefore, it is important to understand the consequences of reverberation on auditory processing.

Reverberation is defined as the sum of all sound reflections off the walls and objects in an enclosed space (Blauert, 1997). The reflections arriving within the first 50 ms after the direct sound are often called early reflections and so are distinguished from late reflections (Bradley et al., 2003; Arweiler and Bucholz, 2011). The series of direct sound, early reflections, and late reflections are typically characterized by the room impulse response (RIR) (Blauert, 1997). In addition, the reverberation time (RT), defined as the time required for the sound level to decay by 60 dB after it has been turned off, is often used as a simple parameter to characterize room reverberation (Crandell and Smaldino, 2000). However, the overall effects of reverberation on

speech intelligibility depend on many other factors including the dimension of the room, the source-receiver distance, and the relative orientation of talkers and listeners (Blauert, 1997; Bradley et al., 2003). Boothroyd (2008) and Giguère (2013) discussed several of these factors in more detail in the context of speech understanding in children with hearing impairment.

From an acoustical standpoint, reverberation can cause degradation of speech signals, such as temporal smearing, filling dips and gaps in the temporal envelope, increasing the prominence of low-frequency energy, and flattening formant transitions (Assman and Summerfield, 2004; Sayles and Winter, 2008). Such degradation of speech signals mainly results from masking of a speech segment by reflections of the segment itself, known as self-masking, and by reflections of previous segments, known as overlap-masking (Devore and Shinn-Cunningham, 2003). On the other hand, reverberation amplifies the speech signal level. Such amplification can assist with speech perception particularly in instances where the direct sound is reduced (Bradley et al., 2003).

Speech stimuli evoke both transient and steady-state brainstem responses. The transient response refers to the initial part of the ABR (typically less than 20 ms), reflecting the transmission delay between the ear and the rostral brainstem structures, in response to any type of stimulus (speech or non-speech) (Skoe and Kraus, 2010). The transient response has been used as an objective tool by clinicians and researchers for assessing the neural activity of the ascending auditory pathway between the cochlear nerve and the inferior colliculus in midbrain. The transient response is typically analyzed in the time domain, which involves measuring amplitudes and latencies of various peaks (most notably V and A) in the waveform. Previous studies (e.g. Laroche et al., 2013; Prévost et al., 2013) using a synthesized speech vowel (/a/) have demonstrated that

background noise has an effect on waves V and A by increasing the latency and decreasing the amplitude. No study to our knowledge has investigated the effects of reverberation on waves V and A amplitudes and latencies. The steady-state response, on the other hand, is generated in response to a periodic speech stimulus, and provides information about the neural encoding of the periodic components of the stimulus. The steady-state response follows the speech stimulus acoustical features including fundamental frequency and low frequency formants (Skoe and Kraus, 2010). The steady-state response is typically analyzed in the frequency domain and can be categorized into the envelope-following response (EFR), representing the neural response that follows the speech envelope, and the fine structure frequency-following response (FFR), representing the neural response that follows the speech formants. The EFR characterizes the evoked response at F0 and its lower harmonics, while the FFR characterizes the evoked response at higher harmonics, typically in the region of the first formant F1 (Aiken and Picton, 2008; Laroche et al., 2013).

Timing measures of the speech-evoked ABR (e.g., peak latency) provide information about the accuracy with which the brainstem nuclei synchronously respond to acoustic stimulation. A peak latency difference on the order of a fraction of 1 ms in the speech-evoked ABR has been associated with difficulties in perceiving speech in noise (Wible et al., 2004; Banai et al., 2009; Hornickel et al., 2009), and can be of clinical significance (Wible et al., 2004; Chandrasekaran and Kraus, 2010). Magnitude measures of the steady-state speech-evoked ABR (e.g., root-mean-square amplitude) provide information about the robustness with which the brainstem nuclei respond to acoustic stimulation. A magnitude variability of less than 0.01 μv has been associated

with difficulties in perceiving speech in noise (Song et al., 2012; Anderson et al., 2013) and reverberation (Bidelman and Krishnan, 2010).

Given the importance of reverberation for real-world listening, and the increasing use of evoked responses to study auditory processing of more naturalistic signals, it is worthwhile to investigate the effects of reverberation on these responses. Only a few researchers have investigated the effects of reverberation on speech stimuli using electrophysiological methods. Unlike behavioral studies where overlap-masking and self-masking effects of reverberation can be easily explored using multisyllabic words or long-duration sentences, electrophysiological studies on reverberation pose challenges primarily due to the restriction of short stimulus duration. Bidelman and Krishnan (2010) produced reverberant stimuli by time domain convolution of a 250 ms synthetic vowel /i/ with RIRs recorded in a corridor at a distance of 0.63 m (RT = 0.7 s), 1.25 m (RT = 0.8 s), and 5 m (RT=0.9 s) from the sound source. Results indicated that reverberation had little effect on the neural encoding of pitch, characterized by the EFR response, but significantly decreased neural encoding of formant-related harmonics, characterized by the FFR response. Fujihira and Shiraishi (2015) produced reverberant stimuli by time convolution of a 170 ms synthetic /da/ with three RIRs (RT = 0.5, 1.0 and 1.5 s) chosen from the environment/architectural acoustics “SMILE 2004” database (Architectural Institute of Japan, 2004). Results from this study indicated that speech intelligibility under reverberation in elderly listeners is related to their ability to encode the temporal fine structure of speech, characterized by the FFR response - which is decreased under effects of reverberation.

It is important to note that in Bidelman and Krishnan (2010) and Fujihira and Shiraishi (2015) the methodological choices and/or study design promoted the detrimental contribution of reverberation on speech processing. In Bidelman and Krishnan (2010), the amplitude of the reverberant speech stimulus was normalized to the same overall level as the non-reverberated condition, thus negating the natural amplification that typically occurs in reverberant conditions. In Fujihira and Shiraishi (2015), the early reflection components from the RIR were removed, thus eliminating the benefit they may have on speech perception. These studies were not designed to reveal the separate effects of early and late reflections on the speech-evoked EFR and FFR.

The purpose of the current study is to shed light on how various components (direct sound vs. early reflections vs. late reflections vs. full RIR) of the reverberated speech are encoded in the auditory system using an electrophysiological test. No other study to our knowledge has made this comparison using the speech-evoked ABR.

Given the beneficial effects of early reflections and the detrimental effects of late reflections, we hypothesized that the speech-evoked ABR to the direct and early stimuli would be similar, while the responses to direct or early stimuli would be different from responses to late stimuli. The similarity and difference are established with respect to measures of latency of waves V and A, amplitude of waves V and A, and amplitude of EFR and FFR. Furthermore, given the effects of reverberation on speech signals, such as temporal smearing of the speech envelope, and flattening formant transitions, we hypothesized that the speech-evoked ABR to the late and full reverberant stimuli would be degraded (i.e., longer latencies of waves V and A, and reduced

amplitudes of waves V and A, and EFR and FFR) compared to the response to the direct stimulus. Results from this study show that effects of the different components of RIRs can be studied electrophysiologically, thereby providing another means than perceptual studies to investigate the effects of room acoustics on auditory processing of speech.

3.2. Methods

3.2.1. Subjects

Twelve subjects (2 males and 10 females) participated in the study. The age of participants ranged from 21 to 31 years (mean = 24). 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 reported a history of hearing or ear-related issues. All subjects provided their informed consent in compliance with a protocol approved by the University of Ottawa Research Ethics Board.

3.2.2. Room Impulse Response

The Room Impulse Response (RIR) was computed following the algorithm from McGovern (2003), which is based on the image method from Allen and Berkley (1979). This method consists of modeling reflections as if they originated from virtual locations behind room boundaries, and the contributions from all the virtual sources are summed. Each virtual source contributes a delayed impulse (echo), whose time delay is equal to the distance between the source and the listener divided by the speed of sound. The echo amplitude is inversely proportional to the distance traveled and proportional to the room reflection coefficient to the power of the number of reflections (Khars and Bradenburg, 1998).

The algorithm from McGovern (2003) assumes that the room model is rectangular, making it easy to simulate using a computer program. Furthermore, it assumes that the average room reflection coefficient is uniform in the room. The algorithm takes into account the position of the n virtual sources, position of the listener, delay and magnitude of each reflection, room dimensions ($H \times L \times W$), number of virtual sources, and average room reflection.

The RIR in this study was generated for a room of dimensions $3.5\text{m} \times 11\text{m} \times 14\text{m}$ and for source and listener coordinate positions of $(2\text{m}, 2.5\text{m}, 1.5\text{m})$ and $(2\text{m}, 8\text{m}, 1.2\text{m})$, respectively. The source-listener distance (5.508 m) was selected to reflect an equal ratio of early and late energy. The early energy value was calculated as the sum of the amplitude components in RIR up to 50 ms following (but not including) the direct sound. Similarly, the late energy was calculated as the sum of the amplitude components in RIR following the early reflections (Figure 1).

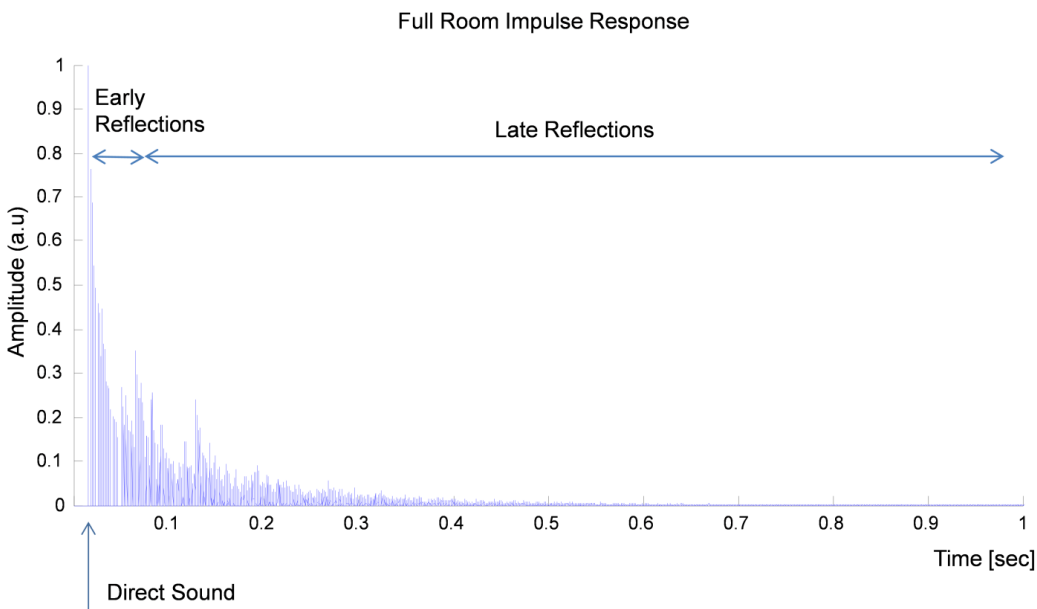


Fig. 1. Representation of the full room impulse response (RIR) showing the direct sound, early reflections, and late reflections. The RIR was produced for a room of dimensions $3.5\text{m} \times 11\text{m} \times 14\text{m}$ and for source and listener coordinate positions of $(2\text{m}, 2.5\text{m}, 1.5\text{m})$ and $(2\text{m}, 8\text{m}, 1.2\text{m})$, respectively.

Bradley (1986) indicated that the level of reverberation could vary in typical classrooms from 0.4 to 1.2 s. The RT in this study was chosen as 1 s, simulating a classroom near the upper end of this range. The average room reflection was calculated from the desired RT based on the Eyring formula (Fischetti, 2003).

As noted earlier, the overall effects of reverberation on speech intelligibility depend on many factors including the dimension of the room, the source-receiver distance, and the relative orientation of talkers and listeners (Blauert, 1997; Bradley et al., 2003). In our study, we modeled a situation where the source and listener positions reflected an equal ratio of early and late energy.

3.2.3. Stimuli

A 2-s 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 et al. 2013). The reverberant stimuli were created as a result of the convolution between the 2-s synthetic vowel /a/ and each of the RIR components: direct sound, early reflections and late reflections, as well as the full RIR (Figure 1) as follows:

1. Direct component = vowel /a/ * RIR direct sound
2. Early reflections = vowel /a/ * RIR early reflections only (i.e. reflections within 50 ms following the direct sound)
3. Late reflections = vowel /a/ * RIR late reflections only (i.e. reflections following the early reflections)

4. Full RIR= vowel /a/ * full RIR (i.e. direct sound + early reflections + late reflections), where the symbol * denotes convolution.

The reverberant stimuli were chopped between 1200 ms and 1500 ms in order to include self-masking from the initial portion of the vowel. We selected 1200 ms to ensure the room response had stabilized, given the 1s RT. The end result was four stimuli (direct, early, full, and late), each with a 300-ms duration generated at a sampling frequency of 48 kHz and with 16-bit resolution. The direct stimulus was presented at 76.9 dB SPL, as measured in an ear simulator (IEC 60318-4, G.R.A.S. RA0045). The levels of the reverberant stimuli (early at 79.4 dB SPL, full at 79.2 dB SPL, and late at 79.7 dB SPL) were not normalized in full order to be consistent with real room environments. The different time-domain stimulus waveforms and amplitude spectra are shown in Figure 2. The spectral components of full are not always higher than those of direct, early, and late. We suspect that this is due to phase effects, also affecting the time-domain waveforms.

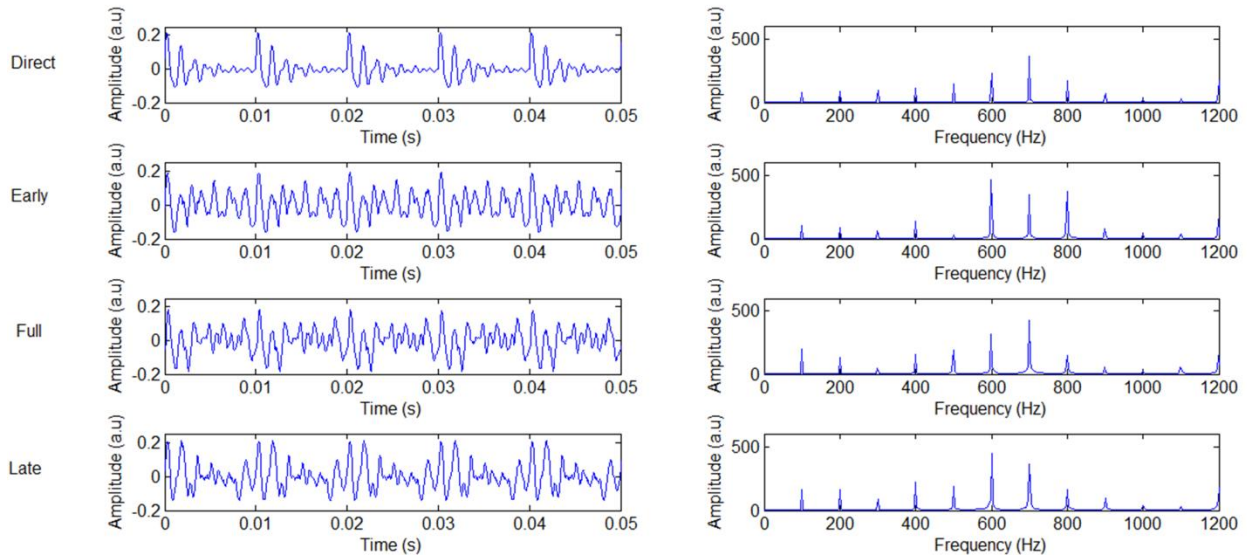


Fig. 2. Time-domain waveforms and amplitude spectra of the four stimuli (Direct, Early, Full, Late). Only the first 50 ms is shown.

3.2.4. Recording and analysis of speech-evoked ABR

Recordings of the speech-evoked ABR were performed in a shielded audiometric room at the University of Ottawa, in dim lighting. Responses were recorded with a vertical montage using BioMARK™ v.7.0.2 system (Biological Marker of Auditory Processing, Biologic Systems Corp., Mundelein, IL) from the vertex (Cz), with right ear lobe as reference, and left ear lobe as ground. Responses were recorded over 319.8-ms epochs (1024 points/epoch corresponding to a sampling frequency of approximately 3202 Hz) and were amplified and filtered using an amplifier with a gain of 10,000 and a filter bandwidth extending from 30 to 1000 Hz. All electrode impedances were below 5 kOhm at 10 Hz.

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 μ V were discarded. In addition, we followed the recommendations of the manufacturer of BioMARK™ regarding environmental electromagnetic noise reduction such as turning off fluorescent lights when operating the equipment, and making sure all the devices are connected to the iso-transformer provided with the system.

The stimuli (direct, early, full, and late) were delivered using the Bio-logic insert-earphone of the BioMARK system, monaurally in the right ear and in alternating polarity, and were presented in a pseudo-randomized order. For each of the stimuli (direct, early, late, and full), two sets of 1500 ABR responses were recorded consecutively. The 1500 artifact-free responses were added and divided by the total number of responses to minimize the influence of the cochlear microphonic and stimulus artifacts on the response (Aiken and Picton, 2008; Campbell et al., 2012), and to

obtain response components related to the stimulus envelope. We term the responses calculated with the addition method as the envelope following responses (EFR). Similarly, the 1500 artifact-free responses to the two polarities were subtracted from each other and divided by the total number of responses to obtain response components that follow the stimulus temporal fine structure (Aiken and Picton, 2008; Anderson et al. 2010). We term the responses calculated with the subtraction method as the frequency following responses (FFR). In order to test the absence of electromagnetic leakage, the electrodes were placed on one subject's scalp, and in all other aspects the experiments proceeded as normal, with the sole exception that instead of placing the 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 transducer as if the earphones were inserted in the ear. Spectral analysis of the recorded signal coherently averaged over 3000 trials showed no significant components in the steady-state response, indicating no contamination of the recorded responses from the sound-generating equipment to the electrodes.

We analyzed both the transient and the steady-state components of the responses to the speech stimuli. The steady-state response followed the transient response and was taken to start at 20 ms and for a duration of 300 ms.

For the transient response, the amplitude and latency of waves V and A were visually extracted from the time-domain EFR waveforms. The prominent onset wave "V" in the response starts 6–10 ms following the stimulus, reflecting the time delay to the upper auditory brainstem, and is followed by the negative wave "A" (Chandrasekaran and Kraus, 2010).

For the steady-state response, spectral analysis was performed using the Fourier transform in Matlab v.7.9 (MathWorks, Natick, Massachusetts, USA), in order to extract the amplitudes of different harmonics from the EFR and FFR. The EFR amplitude was computed from the fundamental and harmonics below the first formant in the EFR spectrum, as the square-root of the combined power of F0, 2F0, and 3F0 (Figure 3). The FFR amplitude was computed from the harmonics around the first formant in the FFR spectrum, as the square-root of the combined power of 5F0 to 9F0 (Figure 3). The harmonic amplitudes correspond to the height of the peak in the EFR and FFR root-mean-square (RMS) amplitude spectrum (in μV).

Waves V and A amplitudes and latencies as well as EFR and FFR amplitudes were evaluated separately with one-way repeated measures ANOVAs, using SPSS version 18.0 (SPSS Inc., Chicago, IL) to determine the effect of stimulus. Post-hoc comparisons using Tukey's honestly significant difference test were performed to assess significant differences between pairs of stimulus conditions.

3.2.5. Behavioral Test

The subjects were instructed to listen, monaurally in the right ear via a shielded earphone and at the same levels as in the electrophysiological experiment, to the direct sound and subsequently listen to the early, late, and full stimuli. They were then asked to select, on a computer screen, their choice of sound (early, late, or full) that most resembles the direct sound. The direct sound was presented first, and then the other three sounds were presented in pseudo-randomized order. This test was repeated three times.

3.3. Results

3.3.1. Transient Response

The waves V and A amplitudes and latencies across all stimuli (direct, early, full, and late) are shown in Figure 4. The ANOVA results show that there is a main effect of stimulus components with all of these measures (Table 1). Post-hoc tests were used to determine the pairwise differences among the different stimulus conditions (Table 2). For waves V and A amplitudes, a significant difference or a tendency towards significance was found between direct and late, between direct and full, and between early and late, among the possible six pairwise comparisons. For waves V and A latencies, significant differences were found between direct and late, between direct and full, between early and late, and between early and full.

3.3.2. Steady-state Response

The EFR and FFR amplitudes across all stimuli (direct, early, full, and late components) are shown in Figure 3. The ANOVA results show that there is a main effect of stimulus components with these two measures (Table 1). Post-hoc tests for the EFR and FFR amplitudes showed that there was a significant difference or tendency towards significance between direct and late, and between early and late among the six possible pairwise comparisons (Table 2).

3.3.3. Behavioral Response

In a total of 36 trials conducted by the 12 subjects, the early, full, and late stimulus reflection components were the most perceptually similar to direct sound in 67%, 27%, and 6% of the trials, respectively.

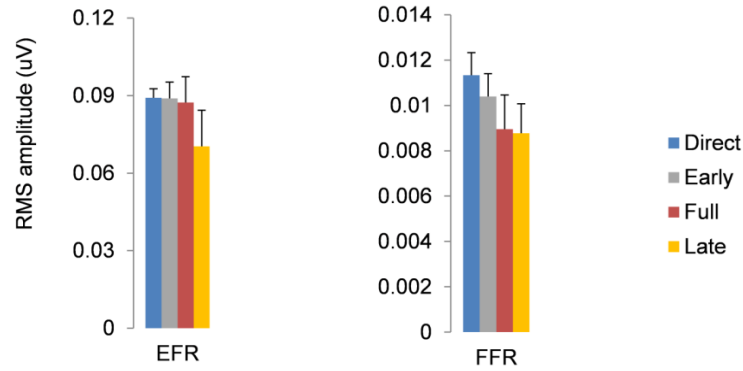


Fig.3. Average of EFR amplitudes (combination of the amplitudes at F0, 2F0, and 3F0 in the EFR spectrum) across subjects, and of FFR amplitudes (combination of the amplitudes at 5F0 to 9F0 in the FFR spectrum) across subjects, for each stimulus (Direct, Early, Full, Late). Error bars indicate standard errors of the mean.

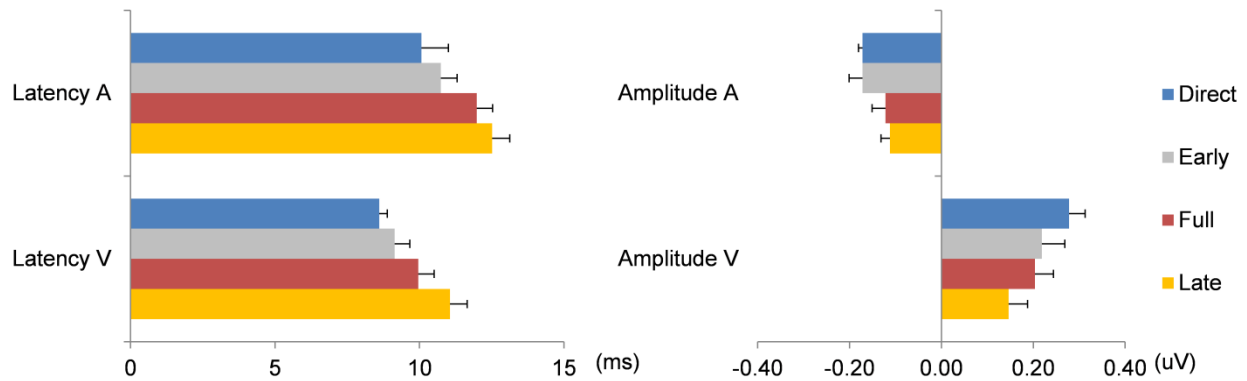


Fig.4. Average of individual latency (ms) and amplitudes (uV) of waves A and V for each stimulus (Direct, Early, Full, Late). Error bars indicate standard errors of the mean.

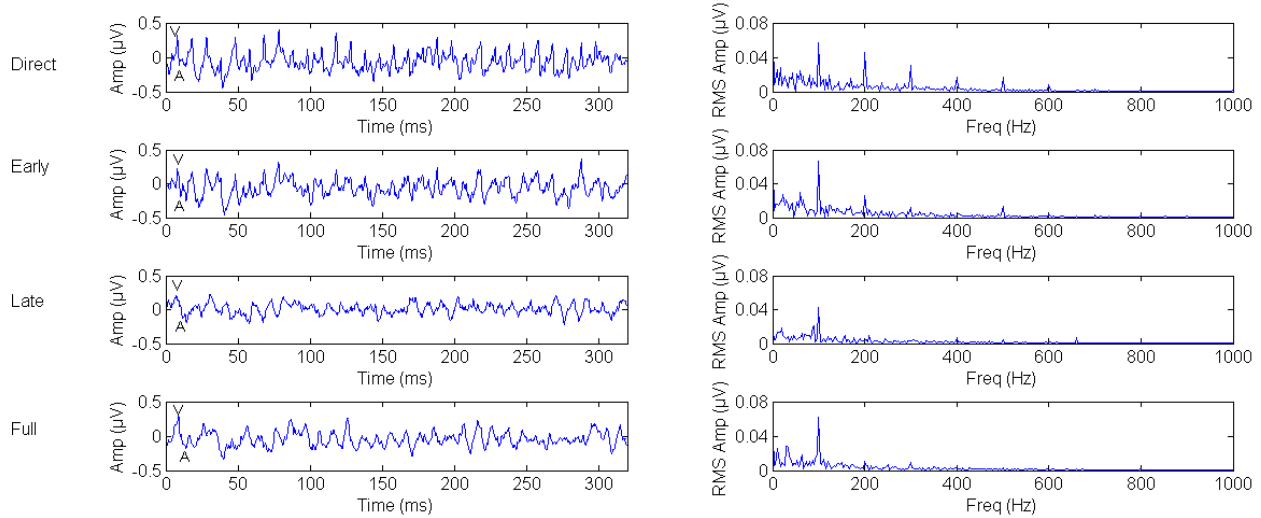


Fig.5. Grand Average Speech-evoked ABR in time-domain and frequency domain of the four stimuli (Direct, Early, Late, and Full).

Table 1. Statistical results from one-way repeated measures ANOVA for latency and amplitudes of waves V and A, and for EFR and FFR amplitudes. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; + $p < 0.1$.

	F	P
Latency V	20.062	0.0001 ***
Amplitude V	3.177	0.021 *
Latency A	16.363	0.0001 ***
Amplitude A	2.291	0.05 +
EFR amplitude	3.327	0.031 *
FFR amplitude	3.009	0.041 *

Table 2. Post-hoc HSD tests for latency and amplitudes of waves V and A, and for EFR and FFR amplitudes. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; + $p < 0.1$. D, E, F, and L denote the four stimuli: direct sound, early reflections, full reflections, and late reflections, respectively.

Post-Hoc	Latency V	Amplitude V	Latency A	Amplitude A	EFR	FFR
D-E	0.204	0.137	0.138	0.917	0.993	0.147
D-L	0.007 **	0.014 *	0.0001 ***	0.048 *	0.050 +	0.019 *
D-F	0.001 **	0.036 *	0.001 **	0.085 +	0.1	0.127
E-L	0.002 **	0.050 +	0.004 **	0.049 *	0.035 *	0.050 +
E-F	0.023 *	0.680	0.008 **	0.194	0.187	0.554
L-F	0.462	0.860	0.174	0.841	0.268	0.413

3.4. Discussion

In this study, we explored the effects of various components of the Room Impulse Response (RIR), namely direct sound, early, full, and late reflections, on auditory processing of speech using speech-evoked ABR and behavioral testing. Our study modeled a situation where the source and listener positions reflected an equal ratio of early and late energy. The full components consisted of the sum of direct, early, and late components.

Our results revealed four key findings (Table 2):

- (1) Responses to direct and early components are not significantly different;
- (2) Responses to full and late components are not significantly different;
- (3) Responses to direct/early and late components are all significantly different;
- (4) Responses to direct/early and full components present a mixed picture; some response measures are significantly different or tend towards significance, while others are not.

Regarding finding (1), waves V and A amplitudes and latencies of the transient response, and EFR and FFR amplitudes of the steady-state response, all showed no statistically significant difference between direct and early reflection components. Furthermore, the results from the behavioral test corroborate these findings. In 67 % of the trials, our subjects reported that the early stimulus is the most perceptually similar to the direct stimulus. These electrophysiological and perceptual results support findings from Bradley et al. (2003) and Roman and Woodruff (2013) who investigated the effects of early and late reflections on the intelligibility of reverberated speech with normal hearing adults. The authors reported that early reflections occurring within the first 50 ms are beneficial to speech perception and can be integrated by the auditory system with the direct sound.

Regarding finding (2), all transient and steady-state response measures showed no statistically significant difference between full and late components. This result is not in line with Hu and Kokkinakis (2014) who reported that late reflections are different from full RIR.

Regarding finding (3), all transient and steady-state response measures showed statistically significant differences between direct and late reflection components. This is consistent with the results from the behavioral test where in only 6 % of the trials, our subjects reported that the late stimulus was the most perceptually similar to the direct stimulus. These electrophysiological and perceptual results support the findings of Bradley et al. (2003) and Roman and Woodruff (2013) who indicated that late reflections have a detrimental effect on speech perception and cannot be integrated with the direct sound.

All transient and steady-state response measures also showed a difference between early and late reflection components, despite their equal energy. These results support findings from Bradley et al. (2003), Boothroyd (2008), Yang and Bradley, (2009), and Roman and Woodruff (2013) which indicated that early reflections contain useful energy while late reflections possess detrimental energy. This makes it likely that electrophysiological responses to these two types of reflections would also differ.

Regarding finding (4), waves V and A amplitudes and latencies of the transient response showed a statistically significant difference between direct and full components or a trend towards significance, but EFR and FFR amplitudes of the steady-state response showed no difference. Likewise, waves V and A latencies of the transient response showed a difference between early

and full components, while waves V and A amplitudes of the transient response, and EFR and FFR amplitudes of the steady-state response, showed no significant difference.

Our results corroborate some of the findings from Hu and Kokkinakis (2014). For instance, the authors reported that the early and late stimuli produced the lowest and highest decrease in speech intelligibility scores compared to the direct stimulus, respectively, while the full stimulus produced a higher score than early but lower than late. Furthermore, the authors reported that the differences in speech intelligibility scores to early vs. late, and late vs. full were statistically significant, while the scores to direct vs. early were marginally statistically non-significant. These results follow the same pattern as our findings with the exception of full vs. late stimuli being statistically significant (our results showed that they are not). In our study, the energy of late reflections and early reflections were equal, and the listeners had normal hearing, while in Hu and Kokkinakis (2014), the listeners wore cochlear implants. It is understood that listeners with cochlear implants are generally more susceptible to detrimental effects of reverberation than normal hearing listeners (Hazrati and Loizou, 2012) and dominance of late reflection components is typically emphasized with high reverberation time (RT = 1.0 s in Hu and Kokkinakis (2014)). As such, this could explain this inconsistent result with Hu and Kokkinakis (2014).

The direct stimulus was presented at a lower level (76.9 dB SPL) than the reverberant stimuli, early (79.4 dB SPL), full (79.2 dB SPL), and late (79.7 dB SPL); however, it resulted in responses that were higher in amplitude and shorter in latency than responses to the full and late reflection stimuli and not significantly different from responses to the early reflection stimuli.

This result can be attributed to the smearing effects of reverberation that blur the speech signal, resulting in a degraded ABR. The smearing effects in the reverberant stimuli (early, full, and late) are shown in Figure 2. In Figure 2, the direct stimulus shows the strongest amplitude modulation compared to the reverberant stimuli (early, full, and late). In Figure 5, the response to the direct stimulus shows a more clearly-defined envelope than responses to the reverberant stimuli. Also, the response to direct stimulus produces richer harmonics compared to the response to the reverberant stimuli, particularly to the late reflections. This finding can be attributed to the detrimental effects of self-masking from the previous vowel /a/, which is typically associated with the late reflections rather than the early reflections. This finding corroborates results from previous perceptual studies which showed that speech intelligibility scores decreased the least and most with early and late reflections stimuli, respectively. This finding therefore provides a window into auditory speech processing and suggests that reverberation should be optimized (i.e. increase early reflections and decrease late reflections) rather than simply minimized.

In addition, our results could be compared to some of the findings of Bidelman and Krishnan (2010). The authors did not decompose the RIR into direct, early, full, and late but rather compared direct (or anechoic) with full RIR at various RTs (0.6, 0.7, and 0.8 s). Also, the authors did not measure the waves V and A amplitudes and latencies. However, they indicated that with increasing reverberation, spectral components of the speech-evoked ABR (EFR and FFR amplitudes) are degraded. This is consistent with our results that showed that the speech-evoked ABR in terms of EFR and FFR amplitudes, and waves V and A latencies and amplitudes are degraded with full RIR compared to direct. Nonetheless, our post-hoc tests showed that there

was not a significant difference between full RIR and direct for the EFR and FFR amplitudes (Table 2). In our study, we modeled a situation where the source and listener positions reflected an equal ratio of early and late energy whereas Bidelman and Krishnan (2010) modeled situations with increasing distances (up to 5 m for RT = 0.8 s). It is understood that at larger distances, the energy of late reflections is higher than the energy of the early reflections. This could explain our inconsistent results with Bidelman and Krishnan (2010).

Furthermore, different interference patterns of the various components of the Full RIR could have boosted or reduced certain frequency regions of the speech signals, which in turn could have amplified or attenuated the neural responses in the auditory system. In order to capture the essence of full vs. direct/early/late comparisons, various source-receiver configurations need to be established, simulating short and long distances, and employing various reverberation times.. The method and tools used in this study could be very helpful when designing future experiments on electrophysiological responses under different reverberant conditions.

A correspondence between speech-evoked ABR and speech-in-noise perception has been reported in several studies. For instance, Kraus and her colleagues recorded speech-evoked ABR in six-talker babble noise using the syllable /da/, and measured speech-in-noise perception using the Hearing-in-Noise Test (HINT), with various age groups: children (Anderson et al., 2010), young adults (Anderson and Kraus, 2010; Song et al., 2011), and older adults (Anderson and Kraus, 2012). The authors reported that all subjects with poor SIN perception had deficits in the subcortical spectral and temporal representation, including low-frequency spectral magnitudes and the timing of transient response peaks, compared to subjects with good SIN perception. Furthermore, Hornickel et al (2009) evaluated the relationship between the subcortical

representation of stop-consonant timing and SIN perception. They recorded speech-evoked ABR to the syllables /ba/, /da/, and /ga/, in children with a wide range of reading abilities. They reported that subcortical differentiation of 3 speech stimuli ([ba], [da], [ga]) was found to be correlated with reading, and speech-in-noise perception, with better performers showing greater differences among brainstem responses to the 3 syllables. On the other hand, Anderson et al. (2013) reported that the steady-state speech-evoked ABR amplitude at F0 increased with older adults who had mild to moderate sensorineural hearing loss compared to older adults with normal hearing, in noise (but not in quiet). Their findings do not correspond to perceptual studies that reported that older adults with hearing loss have poorer speech perception compared to older adults with normal hearing, in the presence of background noise (e.g. Glyde, 2011; Mukari et al, 2014).

In our study, the results show that stimuli that are acoustically most similar (direct and early) result in electrophysiological responses that are not significantly different. On the other hand, stimuli that are acoustically most different (direct and late, early and late) result in responses that are significantly different across all response measures. Meanwhile, the full stimulus is intermediate acoustically in that it contains direct, early, and late energy. None of the response measures with the full stimulus were significantly different from those with the late stimulus, while some of the response measures with this stimulus were different from those obtained with direct and early stimuli.

3.5 Conclusion

This is the first study to investigate how various components (direct sound vs. early reflections vs. late reflections vs. full RIR) of reverberated speech are encoded in the auditory system using the speech-evoked ABR. The amplitude of the reverberant stimuli in our study was not normalized, as in previous studies on reverberation with speech-evoked ABR, thus allowing for the natural amplification that typically occurs in reverberant conditions.

Our results showed that responses to direct and early reflections component are fairly similar with respect to measures of waves V and A amplitudes and latencies as well as EFR and FFR amplitudes. Also, our results showed that responses to late reflections component are different from responses to direct or early reflections component. These findings provide insights towards the understanding of the effects of the different components of the RIRs on auditory processing of speech. Overall, our study shows that effects of the different components of the RIRs can be studied electrophysiologically, thereby providing another means than perceptual studies to investigate the effect of room acoustics on auditory processing. Future work should explore the correlation between reduced/increased speech-evoked ABR measures (EFR and FFR amplitudes, and waves V and A amplitude and latencies), and perceptual studies of room acoustics.

This study was conducted for one specific room and one specific source–receiver set-up. The RIR was created with the image source method for a rectangular room with a reverberation time of 1.0 s. A vowel /a/ was convolved with different components of RIR (direct, early, late, and full). The convolved speech stimuli were delivered monaurally to the subjects' right ear. In order to generalize the effects of reverberation on auditory processing of speech, additional acoustic scenarios, with different source-receiver set-ups, different room shapes, and different

reverberation times, need to be investigated. Also, additional sets of speech stimuli that are reverberated in a real acoustic environment that includes background noise need to be investigated. Furthermore, different populations should be investigated – especially with children with normal hearing, with hearing impairment, and with learning difficulties, because of the important effect of reverberation on classroom learning.

3.6 Declaration of Interest

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

Chapter 4. Effects of self-masking and overlap masking in a reverberant environment using speech-evoked auditory brainstem response (study #3).

Objective: This study investigated the effects of self-masking and overlap masking in a reverberant environment using speech-evoked ABR.

Design: 12 subjects with normal hearing participated in this study. Speech-evoked ABRs were recorded in response to reverberant speech. The speech sample (/A boy fell from the window / + 40 ms gap + vowel /a/) was convolved with the room impulse response (RIR). Three acoustic scenarios were created: anechoic (reverberation time (RT) = 0 s), moderate reverberation (RT = 0.6 s) with dominance of early reflections, and severe reverberation scenario (RT = 1.5 s) with dominance of late reflections.

Results: The wave V and A amplitudes and latencies, and Envelope Frequency Response (EFR) and Frequency Following Response (FFR) amplitudes, decreased significantly with severe reverberation relative to anechoic and moderate reverberation conditions. The wave V and A amplitudes and latencies, and FFR amplitudes, were not significantly different between the anechoic and moderate reverberation. On the other hand, a strengthening of the EFR amplitude with moderate reverberation compared to the anechoic condition tended towards significance.

Conclusion: Our results showed a strengthening in the speech-evoked ABR to moderate reverberation compared to severe reverberation. Overall, our results showed that effects of self-masking and overlap-masking of reverberation can be studied electrophysiologically, thereby

providing objective means to investigate the effect of room acoustics on auditory speech processing.

4.1. Introduction

In everyday life, human listeners are challenged by various auditory stressors that degrade the processing of speech. Indoors, the speech sound reaching the ears is a combination of energy received directly from the sound source and indirectly from a series of reflections from the surrounding surfaces. Such reflections are collectively known as reverberation (Sadjadi and Hansen, 2014).

Reverberation introduced by an enclosed space such as a classroom is characterized by its room impulse response (RIR). The RIR contains three main components: direct sound wave, the early reflections and the late reflections. Reverberation is associated with two types of masking: overlap masking and self-masking. Overlap masking occurs when reflections due to a preceding speech segment overlaps with a subsequent speech segment and is typically associated with late reflections (Sadjadi and Hansen, 2014). Self-masking refers to masking of a speech segment by reflections of the segment itself (Devore and Shinn-Cunningham, 2003; Sadjadi and Hansen, 2014), and is typically associated with early reflections (Sadjadi and Hansen, 2014).

On the other hand, one potentially positive effect of reverberation is that it amplifies the speech signal level. In particular, when the direct sound is blocked, the early reflections may assist with speech perception (Bradley et al. 2003). There are several situations in an enclosed room where the presence of early reflections could be beneficial, such as when (1) the listener is sitting far from the speaker, (2) the speaker directs his/her attention to one part of the room, and (3) the

speaker is speaking with his/her back directed to the listener (Bradley et al. 2003). Results from Bradley et al. (2003) showed that for young adult listeners, the added early reflections arriving within 50 ms following the direct sound have the same effect as increasing the level of the direct sound equivalent to an increase in the SNR up to 9 dB.

Some behavioural studies investigating the effects of reverberation on speech perception reported that direct sound and early reflections component can be integrated together, thus forming useful energy for speech perception, while late reflections are an unwanted signal distortion which degrades speech intelligibility (e.g. Bradley et al. 2003; Boothroyd, 2008; Yang and Bradley, 2009). Roman and Woodruff (2011) investigated the relative contribution of early and late reflections to speech intelligibility with twelve young normal hearing listeners by processing speech to produce three stimuli: direct sound only, direct sound and early reflections, and reverberant only. The direct component was obtained by filtering the target signal through the anechoic impulse response. The direct sound and early reflections component was obtained by convolving the target signal with the first 50 ms of the room impulse response. Speech reception thresholds were calculated for the aforementioned speech stimuli with reverberation equal to 0.4 s, and 0.8s. The authors reported a significant improvement of 8.3 and 11.8 dB at RT = 0.4 and 0.8 s, respectively with the direct and early reflections component, compared to the late reflections component. The authors therefore concluded that in order to achieve significant intelligibility improvements, either the direct sound or the early reflections should be preserved in the speech signal. These results lent further support to the suggestion that early reflections are beneficial to speech intelligibility in normal hearing listeners.

Some electrophysiological studies investigating the effects of reverberation reported that reverberation degrades auditory processing of speech. For example, Bidelman and Krishnan (2010) investigated the effects of reverberation by convolving a 250 ms synthetic vowel /i/ with RIRs recorded in a corridor at a distance of 0.63 m (RT = 0.7 s), 1.25 m (RT = 0.8 s), and 5 m (RT=0.9 s) from the sound source. Their results indicated that reverberation had little effect on the neural encoding of pitch, characterized by what can be referred to as the Envelope Following Response (EFR), but significantly degraded neural encoding of formant-related harmonics, characterized by the Frequency Following Response (FFR). In addition, Fujihira and Shiraishi (2015) investigated effects of reverberation by convolving a 170 ms synthetic /da/ with three RIRs without early reflections (RT = 0.5, 1.0 and 1.5 s) chosen from the environment/architectural acoustics “SMILE 2004” database (Architectural institute of Japan, 2004). Results from Fujihira and Shiraishi (2015) indicated that intelligibility of reverberant speech in elderly listeners is related to their ability to encode the temporal fine structure of speech, characterized by the FFR - which is decreased under effects of late reverberation.

It is important to note that in Bidelman and Krishnan (2010) and Fujihira and Shiraishi (2015) the methodological choices and/or experimental design promoted the adverse contribution of reverberation on speech processing. In Bidelman and Krishnan (2010), the amplitude of the reverberant speech stimulus was normalized to the same overall level as the non-reverberated condition, thus negating the natural amplification that typically occurs in reverberant conditions. In Fujihira and Shiraishi (2015), the early reflection components were removed from the RIR, thus eliminating the benefit they may have on speech perception. These studies were not designed to assess the importance of early reflections on EFR and FFR responses.

In a previous study, we focused on the relative contributions from direct, early, and late reflections in a situation where the effects of speech self-masking were investigated (Al Osman et al., 2015). Our results showed that Speech-evoked ABR to direct and early reflection components possess similar characteristics while responses to direct and early reflections possess different characteristics from responses to late reflections. Overall, our results showed that the effects of reverberation can be studied electrophysiologically, thereby providing another means than perceptual studies to investigate the effects of room acoustics on auditory processing of reverberant speech.

The current study builds upon our previous work in which we investigated effects of self-masking on speech-evoked ABR. Here, we investigated the effects of self-masking and overlap masking in young normal hearing subjects using speech evoked ABR. Three acoustic scenarios were used: An anechoic scenario (RT = 0 s), a moderate reverberation scenario (RT = 0.6 s) with dominance of early reflections, and a severe reverberation scenario (RT = 1.5 s) with dominance of late reflections. We hypothesized that speech-evoked ABR (transient and steady-state responses) would be enhanced with moderate reverberation due to the dominance of the early reflections.

4.2. Methods

4.2.1. Subjects

Twelve participants took part in the study (3 males and 9 females). Their age ranged from 20 to 29 years (mean = 25). All participants had normal hearing thresholds, defined as ≤ 15 dB HL bilaterally at 0.25, 0.5, 0.75, 1, 2, and 4 kHz. None of the participants reported a history of

hearing or ear-related issues. Informed consent was obtained for all testing procedures, which were approved by the University of Ottawa Research Ethics Board.

4.2.2. Room Impulse Response

The Room Impulse Response (RIR) was computed following the algorithm from McGovern (2003), which is based on the image method from Allen and Berkley (1979). This method considers a rectangular room and assumes that the average reflection off the room surfaces is uniform. The algorithm takes into account various parameters such as: the position of the n virtual sources, position of the listener, delay and magnitude of each reflection, room dimensions ($H \times L \times W$), number of virtual sources, and average room reflection.

A hypothetical scenario for a classroom of dimensions $3.5\text{m} \times 11\text{m} \times 14\text{m}$ and for source and listener coordinate positions of (width=2m, length=2.5m, height=1.5m) and ($x=2\text{m}$, $y=8\text{m}$, $z=1.2\text{m}$), respectively, were selected to generate the RIR in this study. Three reverberation times ($RT = 0, 0.6, 1.5$ s) were selected simulating the following scenarios: (1) anechoic, (2) moderate reverberation with dominance of early reflections energy, and (3) severe reverberation with dominance of late reflections energy. The early reflections energy was calculated as the sum of energy in RIR up to 50 ms following (but not including) the direct sound. Similarly, the late reflections energy value was calculated as the sum of energy in RIR following the early reflections. The early reflections to late reflections ratios for moderate and severe reverberation scenarios were 8 and 0.64, respectively. This confirms that the early reflections energy were dominant with the moderate reverberation scenario while the late reflections energy were dominant with the severe reverberation scenario.

4.2.3. Stimuli

A 300 ms synthetic version of the vowel /a/ was generated using a simplified version of the Klatt cascade formant synthesizer (Klatt, 1980; Laroche et al., 2013). The fundamental frequency and upper three formants were static throughout the stimulus ($F_0 = 0.1$ kHz, $F_1 = 0.7$ kHz, $F_2 = 1.22$ kHz, $F_3 = 2.6$ kHz). In order to include effects of overlap masking, a 2.0 s natural sentence /A boy fell from the window/ spoken by a male (Vermiglio, 2008), preceded the 300 ms synthetic vowel /a/. A silent gap of 40 ms was placed between the sentence and the synthetic vowel /a/ to create a natural pause. The reverberant materials (/A boy fell from the window / + 40 ms gap + vowel /a/) was convolved with the RIR for all three reverberation times (RT = 0, 0.6, and 1.5 s). This is shown in Figure 1.

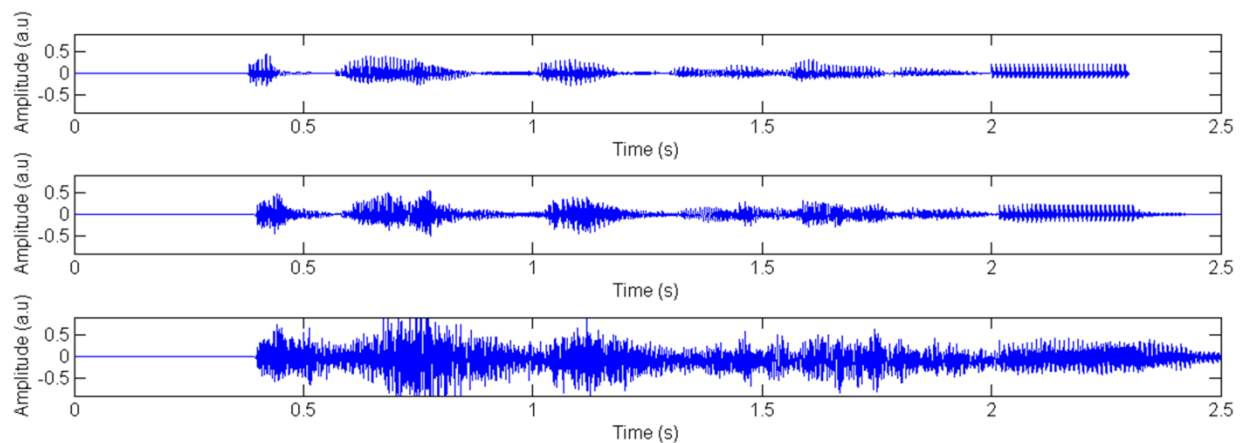


Fig. 1. The aggregate stimulus (/boy fell from the window / + 40 ms gap + vowel /a/) convolved with the RIR for all three reverberation conditions: anechoic, moderate reverberation, and severe reverberation are presented from top to bottom.

The reverberant materials were chopped between 2.04 s and 2.34 s thus including effects of self-masking from the initial portion of the 300 ms synthetic vowel. The end result is three stimuli (anechoic, moderate reverberation, and severe reverberation), each with a 300-ms duration

generated at a sampling frequency of 48 kHz and with 16-bit resolution. These stimuli were presented at 75.3, 78.5, and 83.8 dB SPL, respectively, as measured in an ear simulator (IEC 60318-4, G.R.A.S. RA0045). The amplitudes of the reverberant stimuli were not normalized in order to remain consistent with real room environments. The stimuli time domain and spectra are shown in Figure 2.

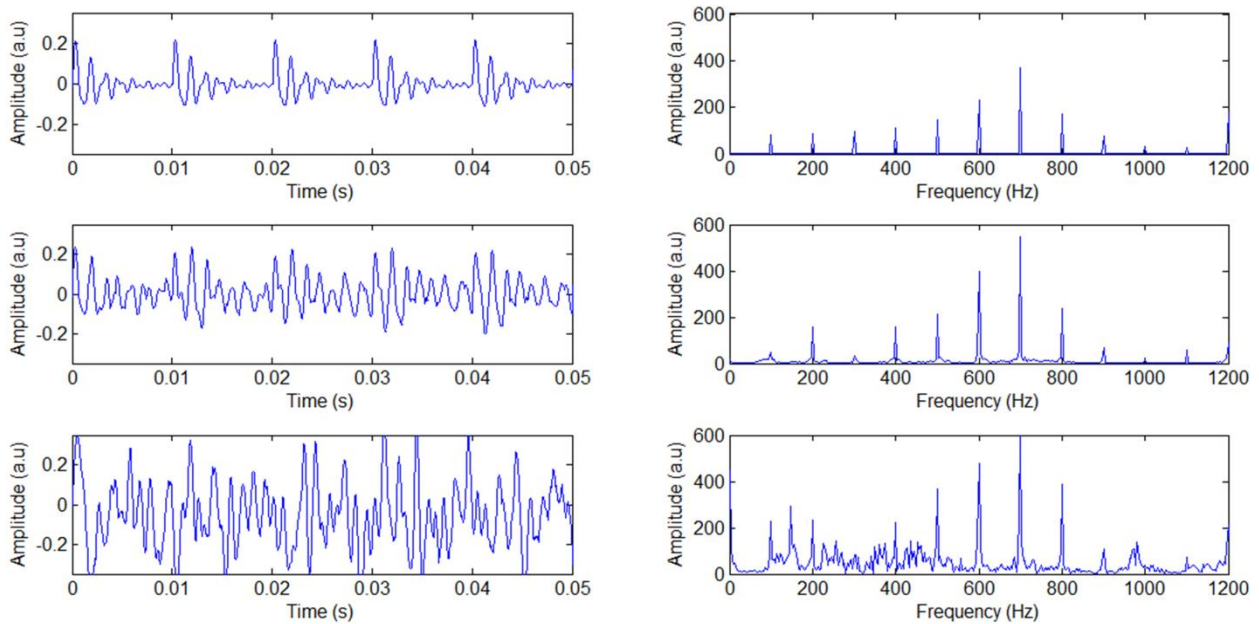


Fig. 2. The stimuli corresponding to the reverberation conditions: anechoic, moderate reverberation, and severe reverberation are presented from top to bottom. Only the first 50 ms is shown.

4.2.4. Recording and analysis of speech-evoked ABR

The speech-evoked ABRs were recorded while the participants sat on a reclining chair in a shielded audiometric room at the University of Ottawa. The stimuli (anechoic, moderate reverberation, and severe reverberation) were presented in a pseudo-randomized order, using the Bio-logic insert-earphone of the BioMARK system, monaurally in the right ear. The responses were recorded, in dim lighting, with a vertical montage using BioMARK™ v.7.0.2 system (Biological Marker of Auditory Processing, Biologic Systems Corp., Mundelein, IL) from the

vertex (Cz), with right ear lobe as reference, and left ear lobe as ground. The contact impedance for all electrodes was kept below 5 kOhm at 10 Hz throughout the experiment. The responses were recorded over 319.8-ms epochs (1024 points/epoch corresponding to a sampling frequency of approximately 3202 Hz) and were amplified and filtered using an amplifier with a gain of 10,000 and a filter bandwidth extending from 30 to 1000 Hz.

We minimized the artifacts throughout the ABR recordings. This was done by instructing the participants to remain relaxed, minimize muscle movement, and stay awake. Also, recorded sweeps in which the response exceeded 23.8 μ V were discarded. Furthermore, we followed the recommendations of the manufacturer of BioMARK™ regarding environmental electromagnetic noise reduction such as turning off fluorescent lights when operating the equipment, and ensuring that all the devices are connected with the iso-transformer provided with the system.

We ensured there was no electromagnetic leakage throughout the ABR recordings. This was done by recording the ABR with one subject as per the procedure noted above. But, instead of placing the insert earphone in the subject's right ear, we inserted it into an ear simulator (IEC 60318-4, G.R.A.S. RA0045), which presents approximately the same acoustic load to the transducer as if the earphones were inserted in the ear. We performed spectral analysis of the recorded signal coherently averaged over 3000 trials, which showed no significant components in the steady-state response, indicating no contamination of the recorded responses from the sound-generating equipment to the electrodes.

The responses were based on coherent averaging of responses to 2000 artifact-free trials in alternating polarities. This method allows the evaluation of the neural coding of two separate

temporal components in speech: the temporal envelope and the temporal fine structure (Anderson et al., 2013). The responses to the alternating polarities were added to emphasize the envelope-following response (EFR) and subtracted to emphasize the temporal fine structure-frequency following response (FFR) (Aiken and Picton, 2008).

Two different types of analysis were conducted: temporal analysis of the transient response and spectral analysis of the steady-state response. In the temporal analysis, the amplitude and latency of waves V and A were visually extracted from the time-domain EFR waveforms. The prominent onset wave “V” in the response starts 6–10 ms following the stimulus, reflecting the time delay to the upper auditory brainstem, and is followed by negative wave “A” (Chandrasekaran and Kraus, 2010). In the spectral analysis, we estimated the root-mean-square (RMS) amplitude spectra (in μV) in Matlab v.7.9 (MathWorks, Natick, Massachusetts, USA) of the steady-state response (20 – 300 ms). We then extracted the amplitudes of different harmonics from the EFR and FFR spectra. The EFR amplitude was computed from the fundamental and two harmonics, as the square root of the combined amplitude power of F_0 , $2F_0$, and $3F_0$ (Figure 3). The FFR amplitude was computed from the harmonics around the first formant, as the square root of the combined amplitude power at $5F_0$ to $9F_0$ (Figure 3).

4.3. Results

The waves V and A amplitudes and latencies, and EFR and FFR amplitudes across all conditions (anechoic, moderate reverberation, and severe reverberation) are presented in Figures 4 and 3. One-way ANOVA with repeated measures indicated a significant effect of the reverberation condition with wave V amplitude ($p = 0.0001$), wave V latency ($p = 0.024$), wave A amplitude ($p = 0.04$), wave A latency ($p = 0.0001$), and EFR amplitude ($p = 0.0001$) and FFR amplitude ($p =$

0.001). Post-hoc comparisons using Tukey's honestly significant difference tests were performed to assess significant differences between pairs of reverberant conditions. As shown in Table 2, waves V and A amplitude and latency, and EFR and FFR amplitudes, deteriorated significantly with severe reverberation relative to anechoic and moderate reverberation conditions. The waves V and A amplitudes and latencies, and FFR amplitudes, were not significantly different between the anechoic and moderate reverberation. On the other hand, a strengthening of the EFR amplitude with moderate reverberation compared to the anechoic condition tended towards significance.

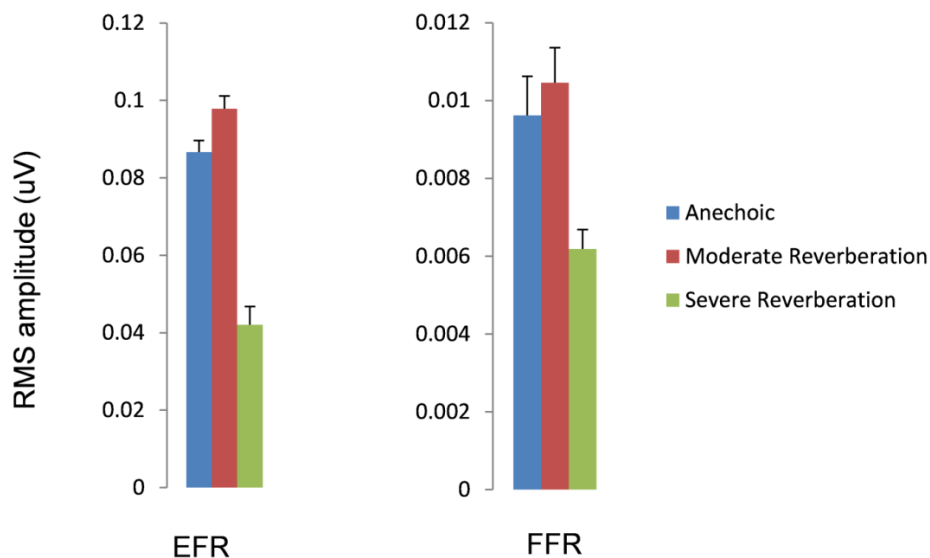


Fig. 3. Average of EFR amplitudes (combination of the amplitudes at F0, 2F0, and 3F0 in the EFR spectrum) across subjects, and of FFR amplitudes (combination of the amplitudes at 5F0 to 9F0 in the FFR spectrum) across subjects, for each stimulus (anechoic, moderate reverberation, and severe reverberation). Error bars indicate standard errors of the mean.

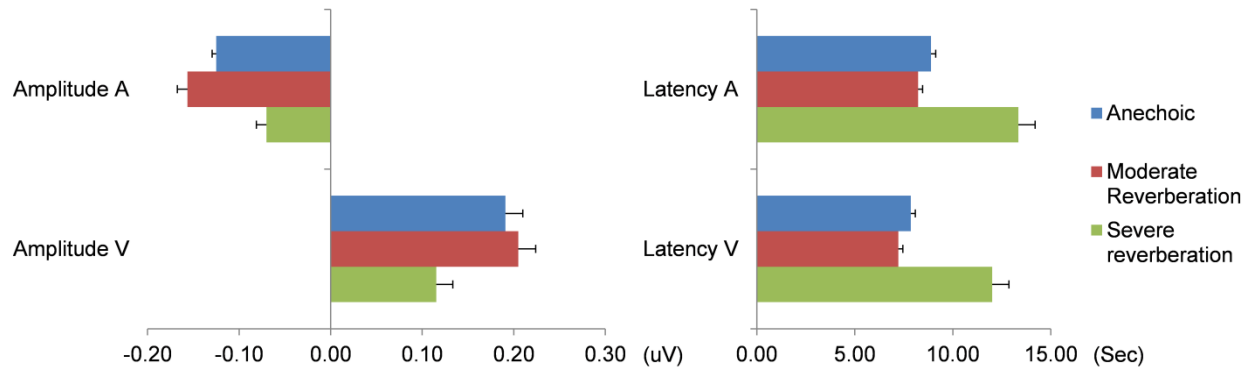


Fig. 4. Average of individual latency (ms) and amplitudes (uV) of waves A and V for each stimulus (anechoic, moderate reverberation, and severe reverberation). Error bars indicate standard errors of the mean.

Table 1. Statistical results from one-way repeated measures ANOVA for latency and amplitudes of waves V and A, and for EFR and FFR amplitudes. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; + $p < 0.1$.

	F	P
Latency V	40.046	0.0001 ***
Amplitude V	4.532	0.024 *
Latency A	50.231	0.0001 ***
Amplitude A	5.920	0.04 *
EFR amplitude	68.803	0.0001 ***
FFR amplitude	18.322	0.0001 ***

Table 2. Post-hoc HSD tests for latency and amplitudes of waves V and A, and for EFR and FFR amplitudes. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; + $p < 0.1$.

Post-Hoc	Latency V	Amplitude V	Latency A	Amplitude A	EFR	FFR
Anechoic-Moderate	0.655	0.625	0.384	0.127	0.051 +	0.157
Anechoic-Severe	0.0001 ***	0.0001 ***	0.046 *	0.059 +	0.0001 ***	0.0001 ***
Moderate-Severe	0.0001 ***	0.0001 ***	0.006 **	0.019 *	0.0001 ***	0.0001 ***

4.4. Discussion

In an earlier study, we examined the relative contributions from direct, early, and late reflections in a situation where the effects of speech self-masking were investigated (Al Osman et al., 2015). The effects of overlap-masking from other speech segments were excluded. Results from (Al Osman et al., 2015) showed that responses to direct and early reflection components were

fairly similar with respect to measures of waves V and A amplitudes and latencies as well as EFR and FFR amplitudes. In contrast, responses to direct/early and late were different for the same measures. Furthermore, the results supported the notion that the combination of direct wave and early reflections constitutes useful energy, while the late reverberation constitutes detrimental energy.

The current study included effects of self-masking from the vowel /a/ itself as well as overlap-masking from a sentence that preceded the vowel, thereby extending our previous study and other works on reverberation with speech-evoked ABR which included self-masking only. Three conditions of reverberation were investigated: (1) anechoic, (2) moderate reverberation with dominance of early reflections, and (3) severe reverberation with dominance of late reflections.

Our results revealed three key findings:

- (A) Responses were significantly stronger with moderate reverberation compared to severe reverberation.
- (B) Responses were significantly degraded with severe reverberation compared to anechoic or moderate reverberation.
- (C) Responses to anechoic and moderate reverberation were not significantly different (except for the EFR amplitude which was stronger with moderate reverberation).

Regarding finding (A), the strength of speech-evoked ABR with moderate reverberation compared to severe reverberation can be attributed to the relative contribution of early reflections or the detrimental effects of the late reflections. The early reflections to late reflections ratios for moderate and severe reverberation scenarios were 8 and 0.64, respectively. This shows that the early reflections' energy was dominant with the moderate reverberation scenario while the late

reflections' energy was dominant with the severe reverberation scenario. As outlined above, early reflections are reported to be beneficial for the perception of speech in a reverberant environment. Roman and Woodruff (2011) and Roman and Woodruff (2013) suggested that the early reflections can be integrated by the auditory system with the direct sound and thus can be advantageous in auditory processing. For example, early reflections in classrooms can promote the students' ability to hear their teacher (Bradley et al., 2003).

The strength of speech-evoked ABR with moderate reverberation is consistent with Bidelman and Krishnan (2010). The authors reported that the EFR and FFR amplitudes of the speech-evoked ABR remained robust with mild reverberation (RT = 0.7 s).

Regarding finding (B), the degradation of speech-evoked ABR with severe reverberation can be attributed to the relative contribution of late reflections. The early reflections to late reflections ratios for moderate and severe reverberation scenarios were 8 and 0.64, respectively. This shows that the early reflections' energy were dominant with the moderate reverberation scenario while the late reflections' energy were dominant with the severe reverberation scenario. The late reflections are understood to degrade listener performance in reverberant speech-based tasks. Furthermore, Roman and Woodruff (2011) and Roman and Woodruff (2013) indicated that late reflections have a detrimental effect on speech perception and cannot be integrated with the direct sound. The deterioration of speech-evoked ABR with severe reverberation is also fairly consistent with Bidelman and Krishnan (2010). The authors reported that the EFR and FFR amplitudes of the speech-evoked ABR highly deteriorated with severe reverberation (RT = 0.9 s).

Regarding finding (C), the results are consistent with our previous work that showed that responses to direct and early reflections components are fairly similar with respect to measures of waves V and A amplitudes and latencies as well as EFR and FFR amplitudes. The early reflections component was dominant in the moderate reverberation scenario. This explains why our results showed that speech-evoked ABR of anechoic and moderate reverberation were not significantly different. These results corroborate findings from Bradley (2003), Yang and Bradley (2009), and Roman and Woodruff (2011) who showed that early reflections occurring within the first 50 ms may be beneficial to speech perception by normal hearing listeners, and can be integrated by the auditory system with the direct sound.

On the other hand, a strengthening of the EFR amplitude with moderate reverberation compared to the anechoic condition tended towards significance. The strengthening of the EFR amplitude has generally been correlated with better perceptual outcomes in normal hearing listeners (Krishnan, et al., 2005; Wong et al., 2007; Bidelman et al., 2009; Song et al., 2012).

4.5 Conclusion

Our results showed a strengthening in the electrophysiological responses to moderate reverberation compared to severe reverberation. This strengthening was attributed to the dominance of early reflections and late reflections components with moderate and severe reverberation scenarios, respectively.

Overall, our results showed that reverberated speech can be studied electrophysiologically, thereby providing objective means to investigate the effect of room acoustics on auditory

processing. Our study was conducted with young adults that have normal hearing. Future work should include various populations - especially with children with normal hearing, with hearing impairment, and with learning difficulties, because of the important effect of reverberation on classroom learning. Also, future work should investigate the combined effects of noise and reverberation.

4.6 Declaration of Interest

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

Chapter 5. General Discussion

The effects of various perceptual stressors on auditory processing were explored in this thesis using speech-evoked auditory brainstem responses. Speech-evoked ABR is an objective measure of auditory processing which allows probing of particular stages of the auditory pathway and evaluating how neural activity in response to specific elements of speech, such as the fundamental frequency and formants, is affected by a certain experimental manipulation. This is in contrast to perceptual studies that reflect the activity of many auditory and other brain centres, and which involve a complex interaction of sensory, linguistic, cognitive, contextual cues and other factors (Boothroyd, 1988; Houtgast and Festen, 2008; Strait and Kraus, 2011).

This chapter summarises the main results of the three studies carried out in this thesis, study #1: effects of stimulus rate and noise on the speech-evoked ABR, study #2: effects of direct, early, and later-arriving speech reflections on the auditory brainstem response, and study #3: effects of self-masking and overlap masking in a reverberant environment using speech-evoked ABR. This chapter also provides a general discussion of the findings from these studies, including the potential limitations of the work and suggestions for future work.

5.1. Original findings

Perceptual studies (Tun, 1998; Adams and Moore, 2009; Adams et al., 2012; Raju and Alwan, 2013) have investigated the interaction between two stressors such as stimulus presentation rate and noise, while previous speech-evoked ABR studies have only investigated the effects of a single stressor such as stimulus presentation rate (Krizman et al., 2010) or noise (e.g. Prévost et al., 2013; Russo et al., 2004) on auditory processing. Study #1 was the first study to investigate

the interaction of two types of perceptual stressors (1) noise and (2) stimulus presentation rate on speech-evoked ABR with normal hearing young adults.

Some behavioural studies investigating the effects of reverberation on speech perception have reported that direct sound and early reflections component can be integrated together, thus forming useful energy for speech perception, while late reflections are an unwanted energy which degrades speech intelligibility (e.g. Bradley et al. 2003; Boothroyd, 2008; Yang and Bradley, 2009). While a few studies (Bidelman and Krishnan, 2010; Fujihira and Shiraishi, 2015) have investigated the effect of reverberated stimuli using speech-evoked ABR, study #2 was the first one to focus on the individual components of room response (direct, early, and late reflections) on auditory processing. Also, study #3 was the first one to show that reverberation can enhance or suppress the ABR response when compared to anechoic conditions, depending on the amount of reverberation.

Overall, our research demonstrated that the effects of stressors such as noise, increased stimulus rate, and reverberation can be studied electrophysiologically using speech-evoked brainstem responses, thereby providing another means than perceptual studies to investigate the effects of these stressors on auditory processing. The major findings arising from studies 1, 2, and 3 are as follows:

- (1) The combined effects of fast stimulus rate and noise are complex, resulting in an interaction, rather than an addition. This was shown in study #1.
- (2) There is a dissociation between the responses corresponding to F0, and the responses corresponding to F1 or transient waves. This was shown in study #1.

- (3) Direct and early reflections resulted in ABR with similar characteristics, while direct or early reflections vs. late reflections resulted in ABR with different characteristics. This was shown in study #2.
- (4) Severe reverberation significantly degraded ABR when compared to anechoic or moderate reverberation, while moderate reverberation significantly enhanced ABR when compared to anechoic speech. This was shown in study #3.
- (5) Human listeners with normal hearing demonstrated neural robustness especially in responses to the envelope under the effects of auditory perceptual stressors such as combined noise and increased stimulus rate in study 1, and reverberation in Studies #2 and #3.

Regarding finding (1), the interaction resulting from combining fast stimulus rate and noise demonstrated facilitative effects for response amplitude at F0, but occlusive or reversal effects for response amplitude at F1 and wave V amplitude. The neural mechanisms for such interaction are not fully defined yet. Further studies are required to reveal the neuro-physiologic and cognitive processes involved in encoding speech stimuli under effects of perceptual stressors such as fast rate and noise.

Regarding finding (2), 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. Indeed, Kraus and co-workers in a series of studies made parallels between the acoustic properties of the speech signal as described in the source-filter model of speech production and the speech-evoked

ABR. The source-filter production model describes the acoustics of speech as a two stages process. The source-filter production model describes the acoustics of speech as a two stages process. This process involves the generation of a sound source, as a result of the vibration of the vocal folds reacting to airflow from the lungs, which is then shaped or filtered by the resonant properties of the oral cavity, tongue, lips and jaw – comprising the filter (Fant, 1960). Kraus and her colleagues suggested that specific components of the speech-evoked ABR reflect the source (i.e. F0 and harmonics of the envelope) and filter (i.e. formants) characteristics separately and using two separate neural streams (e.g., Johnson et al., 2005; Kraus and Nicol, 2005). 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, the source-filter model has not been firmly established and the physiological origins of the two auditory mechanisms are unclear.

Regarding finding (3), we showed in study #2 that transient (wave V and A amplitude and latencies) and steady-state (EFR and FFR amplitudes) components of the speech-evoked ABR were comparable in response to direct and early reflections, while they were different when responses to direct and early reflections were compared to responses to late reflections. Furthermore, the results from the behavioral test corroborate with these findings. In 67 % of the trials, our subjects reported that the early stimulus is the most perceptually similar to the direct stimulus, while only in 6 % of the trials, our subjects reported that the late stimulus was the most perceptually similar to the direct stimulus.

Regarding finding (4), we showed in study #3 that transient (wave V and A amplitude and latencies) and steady-state (EFR and FFR amplitudes) components of the speech-evoked ABR were significantly degraded with severe reverberation compared to anechoic or moderate reverberation. We reported that the significant deterioration in the ABR is attributed to the dominance of the late reflections. Also, we showed that these responses were significantly enhanced with moderate reverberation compared to anechoic speech. We reported that the significant strengthening in the ABR with moderate reverberation is attributed to the dominance of the early reflections.

The results from Studies #2 and #3 corroborate findings from perceptual studies by Arweiler and Buchholz (2011) who reported that early reflections occur sufficiently close in time to the direct sound, thereby they may be perceptually fused with it. Also, the results corroborate findings from Roman and Woodruff (2011) and Roman and Woodruff (2013) who suggested that the early reflections can be integrated by the auditory system with the direct sound and thus can be advantageous in auditory processing, while late reflections have a detrimental effect on speech perception and cannot be integrated with the direct sound. Overall, the results support the notion that direct and early reflections possess useful energy while late reflections possess detrimental energy (Bradley et al. 2003; Boothroyd 2008; Yang and Bradley, 2009; Roman and Woodruff, 2011).

Regarding finding (5), several psychoacoustic studies have investigated how the human auditory system processes speech under the effects of auditory perceptual stressors such as fast presentation rates and noise (e.g. Tun, 1998; Adams and Moore, 2009; Adams et al. 2012; Raju and Alwan, 2013), and reverberation (e.g. Hazrati and Loizou, 2012; Wróblewski et al. 2012).

However, little is yet known about the physiological processes that yield neural robustness under effects of auditory stressors. One study has attributed this neural robustness to the efferent feedback in auditory processing via selective attention, which in turn reinforces our ability to understand speech in noisy and reverberant environments (Song et al., 2008).

5.2. Top-down effects

Speech processing under effects of auditory stressors cannot be fully accounted for via bottom-up encoding. Top-down encoding modulated by selective attention allows us to focus our attention on a particular speech signal while ignoring background noise or another speech signal (Fritz et al., 2007). This type of attention has been shown to evoke changes in activity at the cortical level (Ding and Simon, 2013). However, there is no consensus yet that this type of attention affects the ABR.

Early studies using brief acoustic clicks (i.e., Picton and Hillyard, 1974; Picton et al., 1981; Woods and Hillyard, 1978) showed that attention does not affect the ABR. However, recent studies using pure tones and speech have shown an attentional effect on the ABR (Galbraith et al., 1998, 2003; Sorqvist et al., 2010; Du et al., 2011; Lehmann and Schonwiesner, 2014). For example, Galbraith et al. (2003) have shown that steady-state speech-evoked ABR amplitudes are larger when participants direct their attention towards the evoking stimuli. Also, Galbraith et al. (1998) have shown that steady-state speech-evoked ABR amplitudes to the F0 of each vowel are significantly larger when that vowel is attended to than when ignored. Lehmann and Schonwiesner (2014) have reported an increase in spectral power to the attended vowel's F0 and a decrease in spectral power to the unattended vowel's F0, although the size of the effect was

small. The authors suggested that selective attention modulates the human brainstem frequency-following response. Their results, therefore, indicate that aspects of auditory processing of speech are the result of a dynamic interplay between top-down and bottom-up processing.

We did not control for selective attention, but based on the aforementioned studies, speech-evoked ABR may be useful measures for investigating how top-down effects can assist in selectively targeting a speech signal and improving processing of speech in the presence of auditory stressors.

Beyond the effects of selective attention, Skoe et al. (2013) reported that short-term training of the order of minutes could result in neuroplastic changes that occur in higher-order centers of the brain which in turn alters the functional state of the speech-evoked ABR. Similarly, long-term musical training and auditory training have also been shown to alter the speech-evoked ABR especially under challenging listening conditions (Anderson and Kraus, 2010). A growing body of research has demonstrated that musical training shapes functional brain networks, enhances selective auditory attention, and improves hearing of speech in noise. Kraus and Chandrasekaran (2010) demonstrated that the brains of musicians could provide insights into the neural attentional mechanisms and their potential for experience-dependent plasticity. Neural enhancements in musicians are particularly evident in the context of challenging listening environments, such as in the presence of background noise (Strait and Kraus, 2011) and reverberation (Bidelman and Krishnan, 2010). These studies demonstrated that musically trained adults show less noise or reverberation induced degradation in the subcortical encoding of speech than non-musicians. Strait and Kraus (2011) reported that, in contrast to non-musicians, musicians recruit extra-sensory cortical areas (such as the prefrontal, superior parietal, and

inferior frontal cortices) associated with attention and working memory during challenging auditory tasks that require discriminatory attentiveness (e.g., when subjects are instructed to listen for a certain speech signal target while their auditory system is being stressed or overloaded). The recruitment by musicians of extra-sensory networks involved in attention and working memory may account for their enhanced performance in auditory tasks such as pitch discrimination, sound recall, the hearing of speech in noise (Strait et al., 2010). It may also account for enhancement of the speech-evoked ABR in noise particularly with the response at F0 (Strait and Kraus, 2011), and the speech-evoked ABR in reverberation (Bidelman and Krishnan, 2010), possibly through top-down effects.

Likewise, some studies have demonstrated that auditory training improves processing of speech at the brainstem level. Carcagno and Plack (2011) reported that an enhancement of the subcortical response that follows the envelope with normal hearing listeners who pursued auditory training for pitch discrimination and for speech recognition in noise (Song et al., 2012). Moreover, long-term experience with a tonal language has been found to correspond to a more robust subcortical representation of the pitch frequency (Wong et al., 2007; Krishnan et al., 2009).

Our work adds to the body of ABR literature that describes how auditory stressors such as increased speech rate and background noise interact together in affecting auditory processing of speech. Also, our work improves our understanding of the effects of the various components of reverberation (direct sound vs. early reflections vs. late reflections vs. full RIR reflections), and effects of self-masking and overlap-masking of reverberation on auditory processing of speech.

Overall, our work demonstrated that different aspects of the speech-evoked ABR are affected differently and in complex ways by the aforementioned auditory stressors, and in the case of reverberation, by different components of the reverberant speech. By developing a better understanding of different aspects of the speech-evoked ABR and how they are affected by the different auditory stressors, we may be able to identify particular types of intervention to improve auditory processing of speech in challenging acoustic environments. Such intervention could strengthen top-down mechanisms, which in turn may tune lower-level sensory mechanisms to increase the neural signal-to-noise ratio. For example, Russo et al (2005) investigated the effects of auditory training on speech-evoked ABR with children with a language-based learning problem (e.g., dyslexia). Speech-evoked ABR in response to the syllable /da/ were recorded in quiet and noise with two groups (1) experimental group who pursued the auditory training program over a six months period and (2) control group who did not participate in the auditory training program. The authors reported that the steady-state speech-evoked ABR latencies decreased for the experimental group compared to the control group, in noise (but not in quiet). The authors concluded that auditory training alters the speech-evoked ABR with children with learning problems particularly in challenging listening conditions. Furthermore, Song et al (2012) investigated the effects of auditory training on speech-evoked ABR with young adults with normal hearing. Two groups participated in the study (1) trained subjects who pursued the auditory training over a six months period and (2) untrained subjects who did not undergo the training. The authors reported that the trained subjects showed larger steady-state speech-evoked ABR amplitudes at F0 and H2 compared to the non-trained subjects in noise (but not in quiet). The authors concluded that the young adult nervous system's ability to process speech in adverse listening conditions such as background noise can be improved via an intervention program.

5.3. Limitations

There are several limitations in our studies. One limitation is related to the frequency components that can be captured via the Biologic measurement device. This device has a cut-off frequency at 1000 Hz which prevented us from investigating ABR corresponding to higher stimulus formants such as F2 which may still be at least partially encoded through neural phase-locking. Another limitation is related to the use of a synthetic speech stimulus rather than a natural speech stimulus, which is likely a better stimulus for studying the effects of perceptual stressors especially with reverberation. On the other hand, natural speech typically requires considerably longer recording times as stimulus energy is not concentrated at harmonic frequency components. Even with the use of a synthetic stimulus, another limitation is that a recording time of approximately 15 several minutes per condition was often required to obtain an acceptable ABR and thus we could not include more than 4 conditions in each of our studies.

It should be noted that our results only apply to young adults with normal hearing, and populations with hearing impairments may have different patterns of responses (Anderson et al. 2013). In addition, we did not distinguish between individuals with and without musical training, whereas musicians have been reported to have an advantage in speech-evoked ABR under effects of perceptual stressors (Bidelman and Krishnan, 2010).

5.4 Future work

The results of this research show that auditory stressors such as increased rate, noise, and reverberation, with normal hearing young adults, affect waves V and A amplitudes and latencies, and EFR and FFR amplitudes differently. It would be worthwhile to investigate effects of these

stressors with different populations such as (younger adults vs. older adults), different hearing status (normal vs. impaired), and different listening condition (monaural vs binaural).

Anderson et al. (2012) evaluated the effects of aging on speech-evoked ABR in quiet conditions. The authors reported that older adults had delayed ABRs compared to young adults. Also, they found that older adults had smaller EFR and FFR amplitudes compared to younger adults. The authors concluded that older adults have a loss of temporal precision in the ABR encoding of sound, which may account, at least in part, for their difficulties with speech perception. We speculate that speech-evoked ABR under effects of auditory stressors would be degraded with older adults compared to young adults. However, just as we found complex interactions between the responses to more than one stressor in young adults, it is possible that complex and unexpected interactions will be found with the older group.

The effects of sensorineural hearing loss (SNHL) on speech-evoked ABR were investigated by Anderson et al (2013) with older adults. The authors reported an enhancement to the response that follows the envelope in the SNHL group compared to an age matched normal hearing group, and equivalent representation for the response corresponding to the temporal fine structure in the signal. Similarly, we expect that wave V and A amplitudes and latencies, and FFR amplitude would be degraded under effects of auditory stressors with subjects with hearing loss compared to subjects with normal hearing. Also, we expect that EFR amplitude would increase under these conditions.

Stimuli were presented monaurally in the studies described in this thesis; however, it has been shown that speech perception is facilitated in noise by binaural listening if there is a spatial

separate of the signal and noise (Dubno et al., 2002). Furthermore, Shinn-Cunningham (2000) reported that normal human listeners' robustness to reverberation might arise from their ability to internally model rooms with both ears. The listeners might build up internal geometrical representations of the various reflective surfaces in their environment, and mentally deconvolve the associated signal and room effects. Therefore, it would be worthwhile to examine the possible effect of binaural listening on the ABR in response to reverberant speech.

Also, future work needs to compare the performance of subjects in perceptual tests with speech-evoked ABR when multiple stressors, including reverberation, are present. This type of comparison was used previously by Anderson et al. (2010) in the context of a single noise stressor, where they recorded the speech-evoked ABR in response to the /da/ stimulus in quiet and with +10 dB SNR. The ABRs were then compared to the subjects' performance on the speech in noise tests. An increase in the steady-state speech-evoked ABR latencies was related to poor performance on the Hearing in Noise Test (HINT) (Anderson et al., 2010). Future work should explore the correlation between reduced/increased speech-evoked ABR measures (EFR and FFR amplitudes, and waves V and A amplitude and latencies), and perceptual studies of room acoustics.

Furthermore, future work needs to evaluate different types of speech stimuli such as words or sentences, different room dimensions and different source-receiver configurations. The use of longer stimuli in multiple acoustic conditions could require longer recording sessions but the investigation of these parameters is important because they could have an effect on how the auditory system encodes speech with various auditory stressors.

Finally, future work needs to separate the top-down from the bottom-up effects on the speech-evoked ABR. One possible experiment that could provide insights on the top-down and bottom-up separation would be to collect speech-evoked ABR in response to a speech stimulus by native English speakers in the presence of two types of masking - one that is intelligible to the listener (e.g. story narrated in English) and another type of masking that is unintelligible to the listener (story narrated in a foreign language such as Korean). If there were no top-down effects, then the masker type would not have an effect on the speech-evoked ABR, thus we can conclude that only speech-evoked ABR are affected only by bottom-up encoding. Otherwise, we can conclude that both bottom-up and top-down encoding are at interplay.

References

1. 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.
2. 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.
3. Aiken, S. J., & Picton, T. W. (2006). Envelope following responses to natural vowels. *Audiology and Neurotology*, *11*(4), 213-232.
4. Aiken, S. J., & Picton, T. W. (2008). Envelope and spectral frequency-following responses to vowel sounds. *Hearing Research*, *245*(1), 35-47.
5. Aiken, S. J., & Purcell, D. (2013). Sensitivity to stimulus polarity in speech-evoked frequency-following responses. *In Proceedings of Meetings on Acoustics*, *19*(1), *Acoustical Society of America*.
6. Allen, J. B., & Berkley, D. A. (1979). Image method for efficiently simulating small-room acoustics. *The Journal of the Acoustical Society of America*, *65*(4), 943-950.
7. Al Osman, R., Dajani, H., Giguère, C. (2015). Encoding of reverberated speech using auditory brainstem response. *8th Annual CAA Conference and Exhibition*, *2* (16).

8. Al Osman, R., Giguere, C., Dajani, H. (2016). Effects of stimulus rate and noise on speech-evoked auditory brainstem responses. *Canadian Journal of Speech-Language Pathology and Audiology*, 40(2), 104-119.
9. 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.
10. Anderson S., Kraus N. (2010) Objective neural indices of speech-in-noise perception. *Trends in Amplification*, 14:73-83.
11. Anderson S., Kraus N. (2012). cABR: a neural probe of speech-in-noise processing. Proceedings of ISAAR 2011: Speech perception and auditory disorders. *3rd International Symposium on Auditory and Audiological Research*.
12. Anderson, S., Parbery-Clark, A., White-Schwoch, T., & Kraus, N. (2012). Aging affects neural precision of speech encoding. *The Journal of Neuroscience*, 32(41), 14156-14164.
13. Anderson S., Kraus N. (2013). The potential role of the cABR in assessment and management of hearing impairment. *International Journal of Otolaryngology*.
14. Anderson, S., Parbery-Clark, A., White-Schwoch, T., Drehobl, 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.
15. Anderson S., Skoe E., Chandrasekaran B., Zecker S., Kraus N. (2010). Brainstem correlates of speech-in-noise perception in children. *Hearing Research*, 270(1): 151-157.

16. Architectural institute of Japan. Sound material in living environment “SMILE 2004”. (2004). Tokyo: Gilhodo Shuppan Co. Ltd.
17. Arweiler, I., & Buchholz, J. M. (2011). The influence of spectral characteristics of early reflections on speech intelligibility. *The Journal of the Acoustical Society of America*, *130*(2), 996-1005.
18. Arweiler, I., & Buchholz, J. M. (2011). The influence of spectral characteristics of early reflections on speech intelligibility. *The Journal of the Acoustical Society of America*, *130*(2), 996-1005.
19. Assmann, P., & Summerfield, Q. (2004). The perception of speech under adverse conditions. In *Speech Processing in the Auditory System* (pp. 231–308). Springer New York.
20. Banai, K., Hornickel, J., Skoe, E., Nicol, T., Zecker, S., & Kraus, N. (2009). Reading and subcortical auditory function. *Cerebral cortex*, *19*(11), 2699-2707.
21. Bidelman, G.M., & Krishnan, A. (2010). Effects of reverberation on brainstem representation of speech in musicians and non-musicians. *Brain Research*, *1355*, 112-25.
22. Blauert, J. (1997). Spatial hearing: the psychophysics of human sound localization. MIT press.
23. Boothroyd, A. (1988). Perception of speech pattern contrasts from auditory presentation of voice fundamental frequency. *Ear and Hearing*, *9*(6), 313-321.

24. Boothroyd, A. (2008). *The acoustic speech signal* (pp. 159-167). J. R. Madell, & C. Flexer (Eds.). New York: Thieme Medical Publishers, Inc.
25. Bradley, J. S. (1986). Speech intelligibility studies in classrooms. *The Journal of the Acoustical Society of America*, *80*(3), 846-854.
26. Bradley, J. S., Sato, H., & Picard, M. (2003). On the importance of early reflections for speech in rooms. *The Journal of the Acoustical Society of America*, *113*(6), 3233-3244.
27. 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.
28. Campbell, T., Kerlin, J. R., Bishop, C. W., & Miller, L. M. (2012). Methods to eliminate stimulus transduction artifact from insert earphones during electroencephalography. *Ear and Hearing*, *33*(1), 144.
29. 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.
30. 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.

31. Chandrasekaran B., & Kraus N. (2010). The scalp-recorded brainstem response to speech: neural origins. *Psychophysiology*, 47(2), 236-246.
 32. Crandell, C. C., & Smaldino, J. J. (2000). Classroom acoustics for children with normal hearing and with hearing impairment. *Language, Speech, and Hearing Services in Schools*, 31(4), 362-370.
 33. 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.
 34. Dajani, H. R., Purcell, D., Wong, W., Kunov, H., & Picton, T. W. (2005). Recording human evoked potentials that follow the pitch contour of a natural vowel. *IEEE Transactions on Biomedical Engineering*, 52(9), 1614-1618.
 35. Dajani, H. R., & Giguère, C. (2010). Tonotopic variation in the correlation between the shimmer of a natural vowel and that of the evoked. *Proceedings of 20th International Congress on Acoustics*.
- Dame M.A., Patra H., Vaz P.C., & Ray B. (2013). Effects of speakers' language background on speech perception in adults. *The Journal of the Acoustical Society of America*, 134(5):4063.

36. Devore, S., & Shinn-Cunningham, B. (2003). Perceptual consequences of including reverberation in spatial auditory displays. *Proc of International Conference on Auditory Displays*.
37. Ding N., & Simon J.Z. (2013). Adaptive temporal encoding leads to a background-insensitive cortical representation of speech. *Journal of Neuroscience*, 33:5728–5735.
38. 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.
39. Dubno, J. R., Ahlstrom, J. B., & Horwitz, A. R. (2002). Spectral contributions to the benefit from spatial separation of speech and noise. *Journal of Speech, Language, and Hearing Research*, 45(6), 1297-1310.
40. Fant, G. (1960). *Acoustic theory of speech production*. Copenhagen: Mouton and Co.
41. Festen, J. M., & Plomp, R. (1990). Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *The Journal of the Acoustical Society of America*, 88(4), 1725-1736.
42. Fischetti, A., (2003). *Initiation à l'acoustique*. Belin, Paris.
43. Fritz, J. B., Elhilali, M., & Shamma, S. A. (2007). Adaptive changes in cortical receptive fields induced by attention to complex sounds. *Journal of Neurophysiology*, 98(4), 2337-2346.

44. Fujihira, H., & Shiraishi, K. (2015). Correlations between word intelligibility under reverberation and speech auditory brainstem responses in elderly listeners. *Clinical Neurophysiology*, 126(1), 96-102.
45. Galbraith, G. C., Bhuta, S. M., Choate, A. K., et al. (1998). Brain stem frequency-following response to dichotic vowels during attention. *Neuroreport*, 9(8), 1889 –1893.
46. Galbraith, .G. C., Olfman, D.M., Huffman, T.M. (2003). Selective attention affects human brain stem frequency-following response. *Neuroreport*, 14: 735–738.
47. Giguère, C. (2013). Chapter 2: Understanding hearing loss: Implications for speech perception. In *Audiologic Rehabilitation for Children with Hearing Loss: From Infancy to Adolescence*, 18-28. Thieme Medical Publishers. New York.
48. Glyde, H., Hickson, L., Cameron, S., & Dillon, H. (2011). Problems Hearing in Noise in Older Adults A Review of Spatial Processing Disorder. *Trends in amplification*, 15(3), 116-126.
49. 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.
50. Hall, J. W. (1992). *Handbook of Auditory Evoked Responses*. Boston, Allyn & Bacon.

51. Hazrati, O., & Loizou, P. C. (2012). The combined effects of reverberation and noise on speech intelligibility by cochlear implant listeners. *International journal of audiology*, 51(6), 437-443.
52. 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.
53. Hornickel J, Skoe E, Nicol T, Zecker S, Kraus,N. (2009). Subcortical differentiation of stop consonants relates to reading and speech-in-noise perception. *Proceedings of the National Academy of Sciences*, 106(31): 13022-13027.
54. 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.
55. Houtgast, T., & Steeneken, H. J. (1985). A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria. *The Journal of the Acoustical Society of America*, 77(3), 1069-1077.
56. Hu, Y., & Kokkinakis, K. (2014). Effects of early and late reflections on intelligibility of reverberated speech by cochlear implant listeners. *The Journal of the Acoustical Society of America*, 135, 22-28.
57. International Electrotechnical Commission. (2010). Electroacoustics—Simulators of Human Head and Ear—Part 4: Occluded-Ear Simulator for the Measurement of

Earphones Coupled to the Ear by Means of Ear Inserts (60318–4: 2010). *Geneva, Switzerland: International Electrotechnical Commission.*

58. 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.
59. 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.
60. Johnson, K.L., Nicol, T., & Kraus N. (2005). The brainstem response to speech: a biological marker. *Ear and Hearing*, *26*(5), 424-434.
61. Kahrs, M., & Brandenburg, K. (1998). *Applications of digital signal processing to audio and acoustics* (Vol. 437). Springer Science & Business Media
62. Kenyon, E. L., Leidenheim, S. E., & Zwillenberg, S. (1998). Speech discrimination in the sensorineural hearing loss patient: how is it affected by background noise? *Military Medicine*, *163*(9), 647.
63. Klatt, H.D. (1980). Software for a cascade/parallel formant synthesizer. *The Journal of the Acoustical Society of America*, *67*(3), 971-995.
64. Kokkinakis, K., & Loizou, P. C. (2011). The impact of reverberant self-masking and overlap-masking effects on speech intelligibility by cochlear implant listeners (L). *The Journal of the Acoustical Society of America*, *130*(3), 1099-1102.

65. Kraus, N. (1999). Speech sound perception, neurophysiology, and plasticity. *International Journal of Pediatric Otorhinolaryngology*, 47(2), 123-129.
66. Kraus, N., & Nicol, T. (2005). Brainstem origins for cortical "what" and "where" pathways in the auditory system. *Trends in Neurosciences*, 28(4), 176-181.
67. Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews Neuroscience*, 11(8), 599-605.
68. Krishnan, A. (2002). Human frequency-following responses: representation of steady-state synthetic vowels. *Hearing research*, 166(1), 192-201.
69. Krishnan, A., Xu, Y., Gandour, J., & Cariani, P. (2005). Encoding of pitch in the human brainstem is sensitive to language experience. *Cognitive Brain Research*, 25(1), 161-168.
70. Krishnan, A., Gandour, J. T., Bidelman, G. M., & Swaminathan, J. (2009). Experience dependent neural representation of dynamic pitch in the brainstem. *Neuroreport*, 20(4), 408.
71. Krizman, J., Skoe, E., & Kraus N. (2010). Stimulus rate and subcortical auditory processing of speech. *Audiology & Neurotology*, 15(5), 332-342.
72. Lehmann, A., & Schönwiesner, M. (2014). Selective attention modulates human auditory brainstem responses: relative contributions of frequency and spatial cues. *PloS One*, 9(1), e85442.

73. 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.
74. Mayo, L. H., Florentine, M., and Buus, S. (1997). Age of second-language acquisition and perception of speech in noise. *Journal Speech Language Hearing. Research*, 40, 686–69.
75. McGovern, S. G. (2003). A model for room acoustics.
76. Meador, D., Flege, J.E., & MacKay, I.R.A. (2000). Factors affecting the recognition of words in a second language. *Bilingualism: Language and Cognition* 3, 55–67.
77. Mukari, S. Z. M. S., Wahat, N. H. A., & Mazlan, R. (2014). Effects of ageing and hearing thresholds on speech perception in quiet and in noise perceived in different locations. *Korean Journal of Audiology*, 18(3), 112-118.
78. Nábělek, A. K., & Donahue, A. M. (1984). Perception of consonants in reverberation by native and non-native listeners. *The Journal of the Acoustical Society of America*, 75(2), 632-634.
79. Nábělek A. K., & Letowski T. R. (1989). Similarities of vowels in nonreverberant and reverberant fields. *The Journal of the Acoustical Society of America*, 83, 1891–9.

80. Neuman, A. C., Wróblewski, M., Hajicek, J., & Rubinstein, A. (2010). Combined effects of noise and reverberation on speech recognition performance of normal-hearing children and adults. *Ear and Hearing, 31*(3), 336-344.
81. 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.
82. Pichora-Fuller, M. K. (2003). Processing speed and timing in aging adults: psychoacoustics, speech perception, and comprehension. *International Journal of Audiology, 42*, S59-S67.
83. Picton, T. W., & Hillyard, S. A. (1974). Human auditory evoked potentials. II: Effects of attention. *Electroencephalography and Clinical Neurophysiology, 36*, 191-200.
84. Picton, T. W., Stapells, D. R., & Campbell, K. B. (1981). Auditory evoked potentials from the human cochlea and brainstem. *The Journal of Otolaryngology. Supplement, 9*, 1-41.
85. Prévost, F., Laroche, M., Marcoux, A.M., & Dajani HR. (2013). Objective measurement of physiological signal-to-noise gain in the brainstem response to a synthetic vowel. *Clinical Neurophysiology, 124*(1), 52-60.

86. 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.
87. 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.
88. Roman, N., & Woodruff, J. (2011). Intelligibility of reverberant noisy speech with ideal binary masking. *The Journal of the Acoustical Society of America*, *130*(4), 2153-2161.
89. Roman, N., & Woodruff, J. (2013). Speech intelligibility in reverberation with ideal binary masking: Effects of early reflections and signal-to-noise ratio threshold. *The Journal of the Acoustical Society of America*, *133*(3), 1707-1717.
90. Ruggles, D., Bharadwaj, H., & Shinn-Cunningham, B. G. (2011). Normal hearing is not enough to guarantee robust encoding of suprathreshold features important in everyday communication. *Proceedings of the National Academy of Sciences*, *108*(37), 15516-15521.
91. Russo, N., Nicol, T., Musacchia, G., & Kraus, N. (2004). Brainstem responses to speech syllables. *Clinical Neurophysiology*, *115*(9), 2021-2030.

92. Russo, N. M., Nicol, T. G., Zecker, S. G., Hayes, E. A., & Kraus, N. (2005). Auditory training improves neural timing in the human brainstem. *Behavioural Brain Research*, 156(1), 95-103.
93. Sadjadi, S. O., & Hansen, J. H. (2014). Blind spectral weighting for robust speaker identification under reverberation mismatch. *Audio, Speech, and Language Processing, IEEE/ACM Transactions on*, 22(5), 937-945.
94. Sayles, M., & Winter, I. M. (2008). Reverberation challenges the temporal representation of the pitch of complex sounds. *Neuron*, 58(5), 789-801.
95. 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.
96. Shinn-Cunningham, B. (2000). Learning reverberation: Considerations for spatial auditory displays.
97. Skoe, E., Krizman, J., Spitzer, E., & Kraus, N. (2013). The auditory brainstem is a barometer of rapid auditory learning. *Neuroscience*, 243, 104-114.
98. Song, J. H., Banai, K., Russo, N. M., & Kraus, N. (2006). On the relationship between speech-and nonspeech-evoked auditory brainstem responses. *Audiology and Neurotology*, 11(4), 233-241.

99. Song, J. H., Skoe, E., Wong, P. C., & Kraus, N. (2008). Plasticity in the adult human auditory brainstem following short-term linguistic training. *Journal of Cognitive Neuroscience*, *20*(10), 1892-1902.
100. Song, J. H., Skoe, E., Banai, K., & Kraus, N. (2011). Perception of speech in noise: neural correlates. *Journal of Cognitive Neuroscience*, *23*(9), 2268-2279.
101. Song, J. H., Skoe, E., Banai, K., & Kraus, N. (2012). Training to improve hearing speech in noise: Biological mechanisms. *Cerebral Cortex*, *22*, 1180-1190.
102. Sorqvist, P., Halin, N., & Hygge, S. (2010). Individual differences in susceptibility to the effects of speech on reading comprehension. *Applied Cognitive Psychology*, *24*(1), 67-76.
103. Strait, D. L., Kraus, N., Parbery-Clark, A., & Ashley, R. (2010). Musical experience shapes top-down auditory mechanisms: evidence from masking and auditory attention performance. *Hearing Research*, *261*(1), 22-29.
104. Strait, D. L., & Kraus, N. (2011). Can you hear me now? Musical training shapes functional brain networks for selective auditory attention and hearing speech in noise. *Frontiers in Psychology*, *2*.
105. 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.

106. Tun, P.A. (1998). Fast Noisy Speech: Age Differences in Processing Rapid Speech with Background Noise. *Psychology and Aging*, 13(3), 424-434.
107. Tzounopoulos, T., & Kraus, N. (2009). Learning to encode timing: mechanisms of plasticity in the auditory brainstem. *Neuron*, 62(4), 463-469.
108. Vermiglio, A. J. (2008). The American English hearing in noise test. *International Journal of Audiology*, 47(6), 386-387.
109. Wagener, K. C., & Brand, T. (2005). Sentence intelligibility in noise for listeners with normal hearing and hearing impairment: Influence of measurement procedure and masking parameters. *International Journal of Audiology*, 44(3), 144-156.
110. 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.
111. Woods, D. L., & Hillyard, S. A. (1978). Attention at the cocktail party: Brainstem evoked responses reveal no peripheral gating. *Multidisciplinary Perspectives in Event-related Brain Potential Research*, EPA-600/9-77, 43.
112. Wong, P. C., 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.

113. Wróblewski, M., Lewis, D. E., Valente, D. L., & Stelmachowicz, P. G. (2012). Effects of reverberation on speech recognition in stationary and modulated noise by school-aged children and young adults. *Ear and Hearing, 33*(6), 731.
114. Yang, W., & Bradley, J. S. (2009). Effects of room acoustics on the intelligibility of speech in classrooms for young children. *The Journal of the Acoustical Society of America, 125*, 922–933.

Appendix A: Informed Consent Form

Measurement of Speech-Evoked ABR in Normal Hearing Subjects

You are invited to participate in this study. This informed consent form is to make sure that you understand the purpose of this study and the nature of your involvement. The research project is being conducted by Dr. Hilmi Dajani, Associate Professor at the School of Information Technology and Engineering of the University of Ottawa (tel: 613-562-5800, ext. 6217).

Purpose of the study: The purpose of this study is to investigate the brain's electrical response to speech stimuli. This electrical signal, called the auditory-evoked potential, allows clinicians to diagnose hearing problems and localize their origin in the brain. The sounds that clinicians most commonly use for this test are brief clicks or other artificial sounds. These types of sounds do not provide sufficient information to assess the ability to hear speech. As a result, we are interested in developing tests that record the brain's electrical response to speech, or the Speech-Evoked Potential, with a diverse set of speech samples and in different acoustic environments. The knowledge gained is expected to help in the development of better diagnostic tests of hearing loss and tools for evaluating hearing aids.

Procedure: First of all, we will start by measuring the activity from your brainstem, using small electrodes gently placed on your scalp and earlobes. You will then listen to sounds at a comfortable level for the duration of approximately 90 minutes.

Subject description: Adults with normal hearing and without any external or middle ear disease will be included.

Risks to participating: There is no danger or risk to health associated with this study. All procedures have been pre-tested and they have been used routinely for many years in hospitals, clinics, and laboratories.

Withdrawing from the study: Your participation in this study is voluntary. You may withdraw from the study at any time, by verbally informing any of the investigators, even after signing this form.

If you have any concerns with regards to the ethical conduct of the study, you may contact the Protocol Officer for Ethics in Research, University of Ottawa, Tabaret Hall, 550 Cumberland Street, Room 159, Ottawa, ON K1N 6N5, tel.: 613-562-5841, email: ethics@uottawa.ca.

Compensation: You will be receiving monetary compensation for this study.

Confidentiality: Any information about you collected during this study will be kept strictly confidential. Your name will not be associated with the collected data in any way. While the results will appear in students' dissertations and may also be published, you will not be identified.

The data collection will be conducted by Rida Al Osman. The records will be kept on computer files. The files will be password-protected. The data will be kept in Dr. Dajani's or Dr. Giguère office when not in use and will be conserved for a maximum period of 10 years, after which the computer files will be deleted.

In closing: With your participation, you will be given a copy of this consent form. At the conclusion of the study, should you wish, you will be provided with a summary of the results. You may ask questions at any time, even after signing this consent form.

Signatures: I have read the above description of the study and understand the conditions of participation. My signature indicates that I agree to participate in the study.

Name of the participant (please print):

Participant's signature:

Date: _____

Researcher's signature:

Date: _____