SPEECH PERCEPTION PERFORMANCE IN ECOLOGICAL NOISE

Bhanu Shukla

Follow this and additional works at: https://digitalcommons.memphis.edu/etd

Recommended Citation
Shukla, Bhanu, "SPEECH PERCEPTION PERFORMANCE IN ECOLOGICAL NOISE" (2022). Electronic Theses and Dissertations. 3261.
https://digitalcommons.memphis.edu/etd/3261

This Dissertation is brought to you for free and open access by University of Memphis Digital Commons. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of University of Memphis Digital Commons. For more information, please contact khggerty@memphis.edu.
SPEECH PERCEPTION PERFORMANCE IN ECOLOGICAL NOISE

by

Bhanu Shukla

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

Major: Communication Sciences and Disorders

Concentration: Hearing Science and Disorders

The University of Memphis

August 2022
Dedication

This dissertation is dedicated to the "CREATOR." A special feeling of gratitude to my beloved parents', family, and Dr. Lisa Lucks Mendel, who has gone above and beyond to help me and challenge me to achieve things, and without whom I would not have made it.
Acknowledgements

This project would not have been possible without the support of many people. I am deeply grateful to my supervisor, Dr. Lisa Lucks Mendel, for her patience, guidance, and support. I have benefited greatly from her wealth of knowledge, meticulous editing, numerous revisions, and helped make some sense of the confusion. I am extremely grateful that she took me on as a student and continued to have faith in me over the years. I would like to express my deepest gratitude to my committee members, Dr. Gavin Bidelman, Dr. Eugene Buder, and Dr. Meredith Ray. Your encouraging words and thoughtful, detailed feedback have been very important to me.

I would also like to extend my deepest thanks to Ms. Deling He for her unwavering guidance and unparalleled technical support. When the lab atmosphere is conducive, research becomes enjoyable. I appreciate the members of SPAL (Kara Sander, Hailey Wilson, and Monique Pousson) with whom I had the pleasure of working together; we have worked on several projects together, and without your efforts, my job would have surely been more difficult. Finally, but most importantly, I would like to thank our beloved Professor George Relyea for his enormous support during my Ph.D. I would also like to extend my gratitude to Dr. Udit Saxena and Mr. Himanshu Verma for their profound belief in my work and continuous support.
Preface

Chapter 2, Ecological Noise as a Predictor of Realistic Speech Perception Performance, has been submitted to the *Journal of Speech, Language, and Hearing Research*. Its authors are Shukla, B., Mendel, L.L., & Buder, E. (2022).

Chapter 3, Influence of Multi-talker Babble on Speech Perception Performance in Ecological Noise will be submitted to an appropriate journal. Its authors are Shukla, B., Mendel, L.L., & Buder, E. (2022).
Abstract

A variety of noises, including multi talker babble (MTB), speech spectrum noise (SSN), and several other environmental auditory scenes that had MTB as the dominant noise, have been used in research in the field of audiology to evaluate speech perception in individuals with normal hearing and hearing loss. These different noises and environmental auditory scenes have been employed independently to evaluate speech perception performance. Given that there are several sounds in the real world that might impair communication, the evaluation of speech perception accuracy should encompass all noises that a listener may encounter in typical conversation. Combining several auditory scenes (including MTB) into a single noise for speech perception testing would be efficient and valuable.

The studies presented here aimed to address this problem by integrating multiple realistic auditory settings that have previously been used to assess speech perception into a single noise (including MTB). In the first study, an ecological noise (EN) was produced by combining various natural noises with MTB. The impact of this EN on speech perception performance was then evaluated and compared to that of other noises (MTB & SSN). In the second study, the influence of the MTB in the EN was assessed by producing three different ENs: one without babble, one with babble, and one with reverse babble. In both studies, sentences from the QuickSIN, AzBio, and HINT were used to measure speech perception. Behaviorally, the findings of the current work showed that the speech perception performance for the EN used here was poorer compared to the other noises and the presence of MTB affected speech perception. Acoustically, the spectrum of the EN was similar to the other noises, however the EN was highly modulated in the low frequencies.
Ecological noise creates a more difficult listening environment than other noises alone. Speech perception scores obtained using this EN may provide a more realistic expectation of communication in the natural world compared to MTB or SSN, which would be useful for counseling those with hearing loss. Findings from the current work also showed that the QuickSIN was the most sensitive test to measure speech perception in this EN.
# Table of Contents

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Figures</td>
<td>ix</td>
</tr>
<tr>
<td>1. General Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Noise</td>
<td>1</td>
</tr>
<tr>
<td>Effect of noise on speech perception</td>
<td>2</td>
</tr>
<tr>
<td>Speech Spectrum Noise</td>
<td>3</td>
</tr>
<tr>
<td>Multi-Talker Babble</td>
<td>4</td>
</tr>
<tr>
<td>Ecological Noise</td>
<td>7</td>
</tr>
<tr>
<td>2. Ecological noise as a predictor of realistic speech perception</td>
<td>10</td>
</tr>
<tr>
<td>Introduction</td>
<td>15</td>
</tr>
<tr>
<td>Method</td>
<td>16</td>
</tr>
<tr>
<td>Generation of the Ecological Noise</td>
<td>16</td>
</tr>
<tr>
<td>Assessment of Speech Perception Performance</td>
<td>17</td>
</tr>
<tr>
<td>Participants</td>
<td>17</td>
</tr>
<tr>
<td>Materials</td>
<td>17</td>
</tr>
<tr>
<td>Procedure</td>
<td>18</td>
</tr>
<tr>
<td>Acoustical Analysis</td>
<td>21</td>
</tr>
<tr>
<td>Results</td>
<td>22</td>
</tr>
<tr>
<td>QuickSIN 4-Talker Babble vs. Ecological Noise</td>
<td>23</td>
</tr>
<tr>
<td>AzBio 10-Talker Babble vs. Ecological Noise</td>
<td>24</td>
</tr>
<tr>
<td>HINT Speech Spectrum Noise vs. Ecological Noise</td>
<td>25</td>
</tr>
<tr>
<td>Discussion</td>
<td>26</td>
</tr>
<tr>
<td>Energetic Masking and Informational Masking</td>
<td>27</td>
</tr>
<tr>
<td>Effect of SNR</td>
<td>28</td>
</tr>
<tr>
<td>Acoustical Characteristics</td>
<td>29</td>
</tr>
<tr>
<td>Limitations</td>
<td>29</td>
</tr>
<tr>
<td>Conclusion</td>
<td>29</td>
</tr>
</tbody>
</table>
3. Influence of MTB on Speech Perception Performance in Ecological Noise
   Introduction 31
   Method 35
      Creation of a New Ecological Noise 35
      Evaluation of Speech Perception 36
         Participants 36
         Stimuli 37
         Procedure 37
         Acoustical Analysis 39
   Results 40
      Acoustical Analysis 40
      Speech Perception Performance 42
         Speech Perception Scores Across the Noise Conditions within Each Test 43
         Speech Perception Scores Within Each Condition Across Tests 46
   Discussion 49
      Speech Perception Performance Within Each Test Across the Noise Conditions 50
      Speech Perception Performance Within Noise Conditions Across Tests 51
      Acoustical Analysis 52
   Limitations 54
   Conclusion 54

4. General Conclusion 56

5. References 62
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Spectrum of ecological noise, 4-talker babble, 10-talker babble, and speech</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>spectrum noise</td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>Envelope spectrum of ecological noise, 4-talker babble, 10-talker babble,</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>and speech spectrum noise</td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>Mean percent correct scores for 4-talker babble and ecological noise at 0 dB</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>and 5 dB SNR</td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>Mean percent correct scores for 10-talker babble and ecological noise at 0 dB</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>and 5 dB SNR</td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>Mean percent correct scores for SSN noise and ecological noise at 0 dB</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>and 5 dB SNR</td>
<td></td>
</tr>
<tr>
<td>6.</td>
<td>Spectrum of ecological noise with no babble, ecological noise with babble,</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td>and ecological noise with reverse babble</td>
<td></td>
</tr>
<tr>
<td>7.</td>
<td>Envelope spectrum of ecological noise with no babble, ecological noise with</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>babble, and ecological noise with reverse babble</td>
<td></td>
</tr>
<tr>
<td>8.</td>
<td>Mean speech perception scores for QuickSIN sentences using ecological noise</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>with no babble, ecological noise with babble, and ecological noise with</td>
<td></td>
</tr>
<tr>
<td></td>
<td>reverse babble</td>
<td></td>
</tr>
<tr>
<td>9.</td>
<td>Mean speech perception scores for AzBio sentences using ecological noise</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>with no babble, ecological noise with babble, and ecological noise with</td>
<td></td>
</tr>
<tr>
<td></td>
<td>reverse babble</td>
<td></td>
</tr>
<tr>
<td>10.</td>
<td>Mean speech perception scores for HINT sentences under ecological noise with</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td>no babble, ecological noise with babble, and ecological noise with reverse</td>
<td></td>
</tr>
<tr>
<td></td>
<td>babble</td>
<td></td>
</tr>
<tr>
<td>11.</td>
<td>Mean speech perception scores for QuickSIN and HINT sentences under</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td>ecological noise with no babble</td>
<td></td>
</tr>
<tr>
<td>12.</td>
<td>Mean speech perception scores for QuickSIN, AzBio, and HINT sentences under</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>ecological noise with babble</td>
<td></td>
</tr>
<tr>
<td>13.</td>
<td>Mean speech perception scores for QuickSIN, AzBio, and HINT sentences under</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>ecological noise with reverse babble</td>
<td></td>
</tr>
</tbody>
</table>
Chapter I

General Introduction

Though speech perception is described as the process by which the sounds of a language are heard, interpreted, and understood, it is not as simple as it is described. Speech perception is a highly complex phenomenon where speech sounds from the environment ascend to the auditory cortex where they are analyzed as meaningful units. Several factors such as anatomical, physiological, cognitive, background noise, etc. affect one’s speech perception ability (van Rooij & Plomp, 1990). Speech perception gets adversely affected in the presence of a background noise, not just for those with hearing loss, but also for people with normal hearing (Best et al., 2013; Lee et al., 2015; McArdle et al., 2005; Wilson et al., 2007). There are several studies in the literature that have found that background noise is the most common factor that affects speech perception (Compton-Conley et al., 2004; Florentine, 1985; Hutchinson, 1989; Lee et al., 2015; Revit et al., 2002; Sperry et al., 1997; Yoho et al., 2019).

Noise

Noise can be considered to have its own beauty, but any sound that is unpleasant, disruptive, or perceived as loud is typically considered as noise whether it is speech, music, or environmental sound. The difference between non-meaningful sound and speech depends on the listener and how they perceive the sound. For example, western music might be unpleasant for people who do not like it hence it would be considered noise, whereas it would be perceived as pleasurable for people who like it. Noise not only can cause damage to hearing, it can also adversely affect speech perception. Individuals with normal hearing require a +2 dB signal-to-noise ratio (SNR) to understand 50% of speech (Killion et al., 2004), whereas individuals with
mild-to-moderate hearing loss require a much more favorable SNR to understand 50% of speech compared to individuals with normal hearing (Beattie et al., 1997). In general, when we communicate, speech sounds are frequently accompanied by a variety of noises. These noises mask the acoustical and phonetic cues of speech that are important for its recognition/perception. The masking effects caused by different noises depend on their relative energy to the speech sound, frequency spectrum, and intensity over time (Sahoo et al., 2020; Taitelbaum-Swead & Fostick, 2016).

**Effect of noise on speech perception**

Communication often takes place where some sort of noise is present such as traffic, environmental noise, machinery, animals, birds, multi-talker babble (MTB), etc. The masking effect of all these noises depends on the spectrum of the noise, energy (intensity) over time, and average intensity relative to the intensity level of the speech (Sahoo et al., 2020; Taitelbaum-Swead & Fostick, 2016). There are two main masking effects on speech perception that are caused by noise: (1) energetic masking (EM) and (2) informational masking (IM). EM occurs at the peripheral level when the neural excitation evoked by the competing noise exceeds the excitation produced by the target speech. EM is largely a matter of the relative energies in the two signals. Whereas IM occurs at the central and cognitive level where information present in the noise interferes with the intelligibility of speech sounds. IM has a stronger effect than EM because it contains linguistic information that interferes with the intelligibility of the speech signal more than EM which does not contain any linguistic information. The EM effect is mainly seen in steady-state noise such as speech spectrum noise (SSN), whereas the IM effect is seen in MTB and cafeteria noise (Brungart, 2001; Carhart & Tillman, 1970; Taitelbaum-Swead & Fostick, 2016; Vander Werff et al., 2021; Vermiglio et al., 2019).
Carhart and Tillman (1971) first explained the importance of speech perception testing in the presence of noise in a routine audiological battery. They suggested an audiogram cannot reliably predict speech perception in quiet, and similarly, speech perception in noise cannot be predicted from speech perception in quiet (Holmes & Griffiths, 2019; Le Prell & Clavier, 2017; Meyer et al., 2013). Measurement of speech in noise (SIN) is very important because it documents difficulty in understanding speech in the daily listening environment, it helps in designing audiological rehabilitation plans and the selection of amplification devices, and it also provides information for counseling patients about realistic expectations from rehabilitation (Wilson, 2004).

There are several tests that have been designed to assess speech perception utilizing various types of noise. A test such as Hearing in Noise Test (HINT, Nilsson et al., 1994) uses SSN, whereas tests like the Quick Speech in Noise test (QuickSIN, Killion et al., 2004) and the AzBio Sentences test (AzBio, Spahr et al., 2012) utilize MTB. Each of these noises has a different masking effect on speech perception as they differ in their physical characteristics.

**Speech Spectrum Noise (SSN)**

SSN is a stationary noise that has the same long-term average speech spectrum (LTASS) as speech but no actual speech sounds. SSN is generated by filtering white noise to match the LTASS of speech sounds (Van Engen et al., 2014). The masking effect caused by SSN is EM which interferes at the peripheral level by exciting the basilar membrane at locations similar to those activated by important energy regions in the speech sound (Van Engen et al., 2014). For the past eight decades, many studies have been performed using speech-shaped noise and they have found that speech perception is poor in speech spectrum noise compared to quiet (Kalaiah et al., 2019; Sperry et al., 1997; Van Engen et al., 2014; Wilson, 2003; Wilson et al., 2007).
Kalaiah et al. (2019) investigated consonant perception in SSN among young and middle-aged adults with normal hearing sensitivity in quiet and SSN at three different SNRs (+8, 0, and −8 dB SNR). Their results showed that mean consonant scores for young adults were better than middle-aged adults for both quiet and noise conditions. For both the groups, as the SNR decreased the mean consonant scores also decreased. From their findings, they concluded a significant decrease in consonant perception scores for middle-aged adults especially in the presence of SSN. In another study, Sperry et al. (1997) measured word recognition performance in different noises. They included 6-talker babble, the same 6-talker babble presented in reverse, and amplitude modulated SSN that had the same LTASS as the 6-talker babble used in the study. Results showed that the speech perception performance was the worst under 6-talker babble compared to reverse babble followed by SSN.

In the past, SSN was considered to be the best representation of the natural environment and was used in audiometric testing and is even still used today. However, though it has the same average spectrum as speech, it does not have any actual speech sounds which made scientists question how well it represents the natural environment. Since the natural environment consists of several types of noise, different types of speech noises (babble noise or MTB) were developed to assess speech perception.

**Multi-Talker Babble**

Communication frequently happens in an environment where other talkers are active which can interfere with the perception of speech. This is the reason MTB is used for speech perception testing because scientists believe that it is a better representation of the natural environment than SSN (Fikret-Pasa, 1994; Killion et al., 2004). MTB is a noise that is generated when multiple individuals are talking at the same time, i.e., the summed waveform of several
simultaneous talkers. The masking effect caused by MTB is informational masking because it interferes with speech processing at central and cognitive levels (Doll & Hanna, 1997; Kidd Jr et al., 1994; Van Engen et al., 2014). The linguistic information present in the MTB interferes with the intelligibility of speech sounds, which masks the acoustic and phonetic cues of speech sounds – hence greater masking. Although, babble has meaningful linguistic information, the masking effect of babble is heavily dependent on the number of simultaneous talkers in the mixture. An increase in number of talkers makes the babble less informational because less specific speech information is audible when more people are talking.

Over the last three decades, many studies have been performed using babble to assess speech perception ability. These studies have found some very fascinating results suggesting that speech intelligibility in MTB decreases exponentially for both positive and negative SNRs. The effect of 1-talker babble on speech perception is somewhat similar to SSN (response under 1-talker babble is better than SSN). This is because of the listening in the dips phenomenon (Vélez & Bee, 2011). Listening in the dips refers to the ability to catch brief "acoustic glimpses" of speech and other sounds when the intensity of fluctuating background noise momentarily decreases. Speech is highly modulated, and these modulation characteristics create an “acoustic glimpse” (Vélez & Bee, 2011).

Miller (1947) was the first to investigate the number of talkers that influence speech perception ability in his classical study of masking. He investigated the perception of words in 2-, 3-, 4-, 6-, and 8-talker babble and found that speech recognition threshold (SRT- which is defined as the lowest intensity level (dB HL) at which individuals can repeat spondee words 50% of the time) for 1- and 2-talker babble was about 8 dB. Whereas 6- and 8- talker babble produced an additional 3-4 dB of masking compared to 2-talker babble. Miller’s results showed a
monotonic decrease in intelligibility as the number of talkers in the babble increased. Similarly, Festen and Plomp (1990) suggested that 4- and 8- talker babble are more effective maskers than speech-shaped noise. Other studies also support this hypothesis (Cooke, 2006; Killion et al., 2004; Mi et al., 2013; Wilson, 2003; Wu et al., 2021).

Danhauer and Leppler (1979) observed that consonants in a background of babble with 4 and 9 talkers were recognized more poorly than in white noise at SNRs below 5 dB. In another study, Freyman et al. (2004) found that speech perception performance under 2-talker babble was poor but it typically improved as additional talkers were added. Miller’s (1947) own data suggest that 4 and 6 talkers provide marginally more masking than 8 talkers for SNRs of 3 dB and below. Interestingly, the case for word recognition is different. Words have fewer contextual and linguistic cues than sentences. The most effective babble to mask words is a 9-talker babble (Simpson & Cooke, 2005). A 1-, 2-, and 4-talker babble creates a more challenging situation for sentence perception in noise. Babble noise with fewer talkers has more IM effect. As the number of talkers in the babble increases, the IM effect decreases and the EM effect increases (Rosen et al., 2013; McArdle et al., 2005).

According to the findings of the aforementioned investigations, MTB has a more challenging effect on speech perception than SSN, suggesting that MTB reflects a more natural environment than SSN. However, this is not entirely true, because a natural environment consists of several other noises beyond speech, which is the reason scientists started examining the use of natural ecological noise (EN) to evaluate speech perception in noise in a more realistic way.
Ecological Noise

Defining ecological noise is difficult because it includes both natural sounds such as birds chirping, rustling of leaves, wind noise, animal sounds, etc. as well as non-natural sounds such as traffic, crowd noise, cafeteria noise, babble, machinery, etc. The beauty of EN is that it has both IM and EM. EM comes from the non-speech sounds (birds chirping, the rustling of leaves, wind noise, animal sounds, traffic, machinery noise, etc.) and IM comes from the speech sounds (crowds, cafeteria noise, babble noise). There are a few studies that have studied speech perception using EN to simulate a realistic environment.

Lee et al. (2015) investigated speech perception using realistic background noises (MTB, vacuum, and subway noise) and found that speech perception performance was poorer in MTB, followed by vacuum noise and subway noise. Similarly, Best et al. (2013) found the psychometric function to be shallower when speech perception was evaluated in a more complex and realistic environment than in a standard, laboratory condition, and this change in psychometric function elevated speech reception thresholds. Revit et al. (2002) developed a system called R-SPACE that can record and simulate real-world listening situations to evaluate hearing aid performance. Using R-SPACE, Compton-Conley and colleagues (2004) found that performance in the R-SPACE condition was not significantly different from the real-life condition. These studies tried to develop a natural environment in which to measure realistic speech perception performance, keeping MTB as the dominant noise within their EN.

Conversely, Weisser and Buchholz (2019) used 13 different actual environmental scenes to simulate a realistic environment. These auditory scenes mostly reflected a realistic environment, but still, all the scenes had babble as the dominant noise. Though it is true that babble noise is also considered to be ecologically relevant (Fikret-Pasa, 1994; Killion et al.,
2004), there are various additional noises in the real world that may also play a significant role in impeding communication. Natural noise consists of all available noises like a bird chirping, rustling of leaves, wind noise, traffic noise, etc. in addition to babble.

When evaluating someone’s ability to accurately perceive speech, it is logical to take into account any and all background noises they could encounter in everyday communication. It would be most effective and beneficial to combine these scenarios into a single noise (including MTB) that can be used for audiological assessment. The current study aimed to address this issue by integrating multiple realistic auditory scenarios that have previously been used to assess speech perception into a single ecological noise (including MTB).

**Purpose**

Communication in noisy situations is a common complaint from individuals with hearing loss as well as individuals with normal hearing. So far, little research has been done to assess speech perception abilities in a typical (natural) noisy situation, because there is currently no available noise that can functionally be used in the audiologic clinical setting that simulates a real-life listening situation. It is important to have a speech perception assessment that is efficacious that assesses the actual performance of the listener in the real world. Creating such an assessment tool is needed to accurately reflect performance.

Therefore, the purpose of this dissertation was to assess speech perception in individuals with normal hearing using an ecological noise (EN) developed to simulate a natural environment. The EN developed here includes several environmental noises and MTB that can mimic a significantly more realistic environment than those used previously in the literature.
In the first study, we constructed an EN, investigated how it affected people's ability to understand speech, and compared it to several other types of noise. In the second study, we investigated whether the addition of MTB within the EN had a significant effect on speech perception scores.

The specific purposes of the proposed dissertation were:

1. To create an EN that contains several environmental noises and MTB that can mimic a significantly more realistic environment than those used previously in the literature.
2. To measure the effect of EN on speech perception performance and compare it to MTB and SSN on individuals with normal hearing.
3. To conduct a detailed acoustical analysis of the EN (with and without babble), MTB and SSN used in these studies and to determine the frequency and temporal characteristics that make these noises challenging for speech perception.
4. To investigate the effect of the EN with and without MTB on speech perception ability of listeners with normal hearing.
Chapter II

Ecological Noise as a Predictor of Realistic Speech Perception Performance

Introduction

Speech perception is a complex phenomenon of the human auditory system. Anatomical or physiological problems along the auditory pathway may affect the perception of speech, including cognitive factors and the presence of environmental background noise. In 1971, Carhart and Tillman explained the importance of speech perception testing in the presence of noise in a routine audiological battery. They suggested an audiogram cannot reliably predict speech perception in quiet, and similarly, speech perception in noise cannot be predicted from speech perception in quiet. Many years later, other researchers continue to emphasize the importance of speech-in-noise (SIN) testing (Eisenberg et al., 2005; Heinrich et al., 2015), yet audiologists have been resistant to add these tests to the speech audiometry test battery.

Assessment of speech perception in noise is very important because it measures difficulty in understanding speech in the daily listening environment, it helps in designing audiologic rehabilitation plans and the selection of amplification devices, and it provides information for counseling patients about realistic expectations from rehabilitation (Levine et al., 2010; Wilson, 2004). These are the reasons many researchers today continue to stress the importance of SIN testing. Perhaps some audiologists are reluctant to perform SIN testing because the typical noises used in these tests consist of steady-state noises or multi-talker babble and do not include a wider range of sounds in a more typical listening environment. The focus of this study was to measure speech perception in the presence of an environmental noise that can simulate a natural listening environment for more ecologically relevant speech perception testing.
Several tests have been developed to address SIN issues, such as the Connected Speech Test ([CST], Cox et al., 1987), the Hearing in Noise Test ([HINT], Nilsson et al., 1994), the Words in Noise test ([WIN], Wilson et al., 2003), the Quick Speech in Noise test ([QuickSIN], Killion et al., 2004), the Bamford-Kowal-Bench Speech in Noise test ([BKB-SIN], Etymotic, 2005), the Listening in Spatialized Noise-Sentences test ([LiSN-S], Cameron & Dillon, 2007) and the AzBio Sentences (Spahr et al., 2012). Some of these tests use words as stimuli (WIN), while others use sentences (e.g., QuickSIN, HINT). Furthermore, tests like the HINT use speech spectrum noise (SSN), while tests like the QuickSIN and WIN use multi-talker babble (MTB). Development of these tests began to change how the field of audiology addressed the need to assess speech perception in noise, and several studies have used these tests to help understand the mechanisms of speech perception in noise. For example, some studies have investigated how speech perception differs in individuals with normal hearing vs. individuals with hearing loss, while others have examined the different factors which affect speech perception (Best et al., 2013; Compton-Conley et al., 2004; Florentine, 1985; Hutchinson, 1989; Lee et al., 2015; Revit et al., 2002; van Rooij & Plomp, 1990).

As the technical sophistication of research in SIN testing improved over time, researchers began to use different noises in which to measure speech perception (Best et al., 2013, 2015; Compton-Conley et al., 2004; Florentine, 1985; Hutcherson et al., 1979; Lee et al., 2015; Revit et al., 2002; Sperry et al., 1997; Yoho et al., 2019) as well as different signal-to-noise ratios ([SNR], Best et al., 2015; Bidelman, 2016; McArdle et al., 2005; Revit et al., 2002; Wilson, 2003; Wilson et al., 2012; Wong et al., 2008). These researchers found that noises used in SIN testing have differing effects on speech understanding. Masking caused by interference of physical characteristics of the noise is defined as energetic masking (EM), whereas masking
caused by the information present in the noise is defined as informational masking (IM). EM occurs when the energy in both the speech signal and the background noise falls on the same critical band at the same time, impairing intelligibility of portions of the speech signals at the periphery. IM is a higher-level phenomenon that occurs when both the signal and the masker are heard, but the listener cannot distinguish the elements of the target signal from the competing noise because they are so similar in content (Doll & Hanna, 1997; Kidd Jr et al., 1994). For example, SSN and white noise have an energetic masking effect which has less of a masking effect than competing speech such as MTB which has both energetic and informational masking effects (Brungart, 2001; Carhart & Tillman, 1970; Vander Werff et al., 2021; Vermiglio et al., 2019).

In addition to EM and IM effects, decreasing the SNR can also reduce speech perception ability (Shojaei et al., 2016; Song et al., 2011), and increasing the noise level decreases the intelligibility of speech in different ways. As SNRs decrease from positive to negative, the presence of temporal modulations and linguistic information causes speech intelligibility in MTB to decrease drastically, whereas the absence of temporal modulation or cognitively relevant background signals causes speech intelligibility in SSN to decrease more gradually (Shojaei et al., 2016).

There are a few studies that have examined speech perception in a realistic environment. Lee et al. (2015) investigated speech perception ability using words as stimuli presented in a background of real-life noise at different SNRs. The noises were vacuum noise, subway noise, and 40-talker MTB presented at -5 dB, 0 dB, and +5 dB SNR. They found that speech perception scores were poorer in MTB, followed by vacuum noise and subway noise. In addition, scores at -5 dB SNR were the poorest, followed by 0 dB and +5 dB. Best et al. (2013) used a multi-
loudspeaker approach and tried to create a realistic environment and measured speech perception scores using sentences in MTB (8-talkers). They found the psychometric function to be shallower when speech perception was evaluated in this more complex and realistic environment than in a standard laboratory condition, and this change in psychometric function elevated speech reception thresholds.

Revit et al. (2002) developed a system called R-SPACE that can record and simulate real-world listening situations to evaluate hearing aid performance. This system consists of eight shotgun microphones, which can be placed horizontally in a circular fashion in an environment to record and later reproduce the same environment using eight speakers placed in the same positions that were used while recording. Compton-Conley and colleagues (2004) assessed whether directional microphone benefits could be measured in the laboratory setting. The performance of directional microphones was measured in the R-SPACE with a single noise source behind the listener and a single noise source above the listener. Performance in the R-SPACE was compared to performance in a natural, real-life listening environment. They found that performance in the R-SPACE condition was not significantly different from the real-life condition. These studies tried to develop a natural environment in which to measure realistic speech perception performance using MTB. They used different speaker arrays to make it more realistic, but the dominant noise was always MTB.

Weisser and Buchholz (2019) assessed speech perception using 13 different environmental auditory scenes that mostly reflected a realistic environment, and all the scenes had babble as the dominant noise. Each of the 13 situations was used independently to assess speech perception skills. Though it is true that babble noise is also considered to be ecologically
relevant (Fikret-Pasa, 1994; Killion et al., 2004), there are various additional noises in the real world that may also play a significant role in impeding communication.

Natural noise consists of all available noises like a bird chirping, rustling of leaves, wind noise, traffic noise, etc. in addition to babble. It would make sense that all possible noises that a listener experiences in everyday communication be included in the assessment of accurate speech perception performance. Combining these scenes into a single noise (including MTB) that can be utilized for audiometric testing might be more effective and helpful. The current study attempted to address this issue by combining several naturalistic auditory scenes that have been utilized in the past into a single noise (including MTB) to test speech perception.

Communication in noisy situations is a common complaint from individuals with hearing loss as well as individuals with normal hearing. So far, little research has been done to assess speech perception abilities in a typical and natural noisy situation because there is currently no available noise that can functionally be used in the audiologic clinical setting to simulate a real-life listening situation. It is important to have a speech perception assessment that is efficacious that assesses the actual performance of the listener in the real world. Creating such an assessment tool is needed to accurately reflect performance. Therefore, the purpose of this study was to assess speech perception in individuals with normal hearing using an ecological noise (EN) developed to simulate a natural environment. The EN developed here includes several environmental noises and MTB that can mimic a significantly more realistic environment than those used previously in the literature.
Method

Generation of the Ecological Noise (EN)

To measure speech perception performance in a realistic environment, a unique ecological noise (EN) was created. To make it most realistic, many environmental sounds were either downloaded from the internet (Koenig, n.d.) or recorded from the environment. The downloaded noises were carnival noise, a dog barking, traffic, a car engine, rain, a waterfall, a railway station, birds chirping, crowd applause, and stadium noise; the recorded noises were party crowd noise, live music, club and crowd noise, and peak hour traffic noise. These noises were chosen because they are typical noises encountered by individuals on a daily basis during communication. A Zoom H6 portable audio recorder was used to record the live noises using a built-in omnidirectional microphone. The distance between the sound source and the microphone was less than 10 feet. The recorded noise samples had 32-bit resolution with a 44,100 Hz sampling rate.

All the noises were distributed in 8 different tracks, as shown in Table 1. The various environmental sounds were arranged in this way to avoid any gaps of silence when transitioning between noises. For example, a dog barking is a sound that is short in duration, so a few sections of traffic and car engine noise were added in the second track to fill those gaps.

Once all noises were laid in tracks, the files/tracks were added together, and the entire session was normalized to –3 dB range, and one .wav file mixing all the tracks was generated. Normalization is a process by which the loudest part of the waveform is set to a specific amplitude, which means raising or lowering the amplitude of the rest of the signal to that specific level. To make it even more ecologically valid, the 4-talker babble from the QuickSIN (Killion et al., 2004) was added in addition to other environmental noises, and a new sound file was
generated. The produced noise was digitized with a 32-bit resolution using a sampling frequency of 44,100 Hz. The generated noise was 41 seconds in length and was then looped to create 10 minutes of noise, long enough to complete one condition of speech perception testing.

Table 1. Distribution of different environmental noises across all eight tracks of the original ecological noise (EN).

<table>
<thead>
<tr>
<th>Track No.</th>
<th>Noise</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Carnival noise followed by a dog barking</td>
</tr>
<tr>
<td>2</td>
<td>Traffic followed by a car engine noise</td>
</tr>
<tr>
<td>3</td>
<td>Rain followed by a waterfall noise</td>
</tr>
<tr>
<td>4</td>
<td>Party crowd followed by railway station noise</td>
</tr>
<tr>
<td>5 and 6</td>
<td>Birds chirping followed by crowd applause</td>
</tr>
<tr>
<td>7</td>
<td>Live music, party crowd, club noise followed by stadium noise</td>
</tr>
<tr>
<td>8</td>
<td>Peak hour traffic noise</td>
</tr>
</tbody>
</table>

**Assessment of Speech Perception Performance**

**Participants**

A total of 27 participants (aged 18 to 40 years; $M = 25.5$, SD = 5.43) were recruited for the study based on a power analysis. Participants were graduate students from The University of Memphis, and individuals from the city of Memphis. All participants were native English speakers with audiometric thresholds better than 20 dB HL at 500, 1000, 2000, and 4000 Hz in both ears. Participants had normal middle ear functioning and normal cognitive abilities. Participants completed a detailed audiological evaluation which included otoscopic examination, tympanometry (Maico-MI34), pure tone audiometry (GSI-61) and a cognitive screening using
the Montreal Cognitive Assessment ([MoCA], Nasreddine et al., 2005). The MoCA is an easy and quick cognitive screener that measures mild cognitive dysfunction. The test includes domains like visuospatial, naming, language, memory, attention, abstraction, delayed recall, and orientation. All participants were required to have a normal otoscopic examination, normal hearing thresholds (American-Speech-Language-Hearing-Association [ASHA], 2005), normal middle ear function as evidenced by a Type A tympanogram (ASHA, 1988), and no history of any neurologic disorder. Once the participants met these criteria, the cognitive screening was performed to assess their cognitive functioning. If the participants did not meet any of the criteria mentioned above, they were excluded from the study.

**Materials**

Sentence lists from the QuickSIN, AzBio, and HINT were used as the speech stimuli to measure speech perception performance in the EN. Speech perception performance was also determined using the respective noises developed with each sentence test: 4-talker babble from the QuickSIN, 10-talker babble from the AzBio, and speech spectrum noise from the HINT for comparison with the EN created for this study. The scores were measured and compared at two different SNRs of 0 dB and +5 dB. There was a total of 24 conditions: 2 SNRs (0 and +5 dB) × 3 sentence tests (QuickSIN, AzBio, and HINT) × 4 noises (EN, QuickSIN 4-talker babble, AzBio 10-talker babble, and HINT SSN).

**Procedure**

Participants signed an informed consent form approved by the University of Memphis Institutional Review Board. Following the audiological evaluation, speech perception performance was measured for the conditions listed above using the different tests, noises, and SNRs. The test conditions were randomly presented, and lists were counterbalanced to minimize
any learning and practice effects. The testing was carried out in an air-conditioned sound-treated double-walled room meeting ANSI S3.1-1999 (R2008) specifications, and the presentation level of the noise was kept constant at 50 dB HL, while the presentation level of the speech varied between 50- and 55-dB HL. All test materials were calibrated using root mean square (RMS) before presentation to ensure all stimuli and noises were presented at the same level. The speech signal and noise were presented in the sound field at 0° azimuth with the participant sitting at a 1-meter distance from the sound source. Participants were instructed to listen to the speech stimulus and repeat what they heard. They were also asked to ignore the background noise and focus on the speech. Each word in each sentence was worth one point, and based on the number of correct words repeated by the participants, the percent correct response was calculated. Scoring leniency was given only to singulars or plurals: For example, if a participant repeated *beds instead of bed* or vice versa, it was considered correct.

The test materials were presented using compact disks (CDs) through two different CD players, which were routed through an audiometer (GSI 61) to the loudspeaker in the sound booth. Speech stimuli for the EN condition were presented through the Sony (RCD-W500C) CD player, which was routed to the audiometer, and the EN was presented through the Onkyo (DX-C606) CD player, which was also routed to the audiometer. For all other test and noise conditions, speech stimuli and noises were presented through the same Sony CD player.

**Acoustical Analysis**

The study also investigated the acoustical characteristics of EN and compared it to other noises used in this study in addition to speech perception measurement. We characterized the EN by performing spectral and temporal modulation analyses. For the spectral analysis, acoustical energy at each frequency for each noise was extracted using PRAAT software (Boersma &
Weenink 1992-2022). All signals had identical sampling rates and equivalent durations, and 1,048,576 positive frequency bins were obtained from each by FFT analysis (this high resolution is due to the large number of samples in each 41 second sound file plus zero-padding as needed to yield a power of 2 for FFT processing). These magnitudes were then averaged into Bark-scaled bins corresponding to the first 24 critical bands of hearing (Zwicker, 1961), and the spectra were normalized to focus on frequency differences rather than overall level differences.

For the temporal analysis, we modified the technique described in Tilsen and Johnson (2008) to analyze modulation of the full frequency band instead of only the 700-1300 Hz band. For this procedure, all noises were low-pass filtered using a fourth-order Butterworth filter with a 10 Hz cutoff to get the envelope information, down sampled to 80 Hz, and corrected for phase delay. Finally, a Fast Fourier transform was applied to obtain the frequency domain representation of the time-domain amplitude envelope, showing the magnitude of frequency-specific envelope modulations. The resulting modulation spectrum therefore displays the magnitudes of low frequency amplitude modulations, not acoustic frequency magnitudes as in traditional sound spectra.
Figure 1. Spectrum of ecological noise ([EN] red line, 4-talker babble (black line), 10-talker babble (blue line), and speech spectrum noise ([SSN], yellow line). The X axis represents critical bands (Bark scale) and the Y axis represents normalized intensity in dB. The spectral characteristics of the EN were relatively similar to the others.
Figure 2. Envelope spectrum of ecological noise ([EN] red line, 4-talker babble (black line), 10-talker babble (blue line), and speech spectrum noise ([SSN] yellow line). The X axis represents frequency in Hz and the Y axis represents normalized power. The EN is relatively strongly modulated in the low frequency range ~0.625 Hz (a 1.6 s period). As expected, the unmodulated SSN yields a flat modulation spectrum; the datapoints are included as validation of the technique.

Results

The speech perception scores for each test using its original noise (QuickSIN 4-talker noise, AzBio 10-talker noise, and HINT SSN) were compared to the EN within each SNR (0 and +5 dB). Speech perception scores for each test were also compared between the two SNRs. A two-way Analysis of Variance (ANOVA) and Tukey pairwise comparisons were performed to model speech perception performance as a function of noise and SNR and their interaction. This approach allowed us to examine the effect of noise on SNR and speech perception performance. The two-way ANOVA revealed a statistically significant interaction effect between the noise
condition and SNR (F (5, 11) = 23.52, p < .0001). Post hoc Tukey comparisons are described below.

**QuickSIN 4-Talker Babble vs. Ecological Noise**

Figure 3 shows the speech perception scores for the 4-talker babble and the EN condition using the QuickSIN stimuli. Mean percent correct scores for the EN and the 4-talker-babble at 0 dB SNR were 11.73 and 40.24, respectively, whereas the mean percent correct scores at +5 dB SNR were 65.17 and 88.05. Post hoc Tukey comparisons revealed that speech perception scores using the EN were significantly poorer than in the 4-talker babble at 0 dB (p < .0001) and +5 dB (p < .0001) SNR for the QuickSIN sentences. Also, all speech perception scores for 0 dB SNR were significantly poorer compared to +5 dB SNR for both noises.

![Figure 3. Mean percent correct scores for 4-talker babble (black bars) and ecological noise ([EN] red bars) at 0 dB and 5 dB SNR. Vertical lines represent ± 1 standard deviation (SD) and asterisks represent significant differences (*p < 0.05).](image-url)
AzBio 10-Talker Babble vs. Ecological Noise

Figure 4 shows the mean speech perception scores using AzBio sentences in 10-talker babble and EN. Mean percent correct scores for the EN and the 10-talker babble at 0 dB SNR were 58.11 and 71.19, respectively, whereas the mean percent correct scores at +5 dB SNR were 92.81 and 96.93. Post hoc Tukey comparisons revealed that speech perception scores using the EN were significantly poorer than in 10-talker babble noise at 0 dB SNR ($p < 0.0001$), but there was no statistically significant difference at +5 dB SNR ($p > 0.26$) between the two noises for the AzBio sentences. Again, all speech perception scores for 0 dB SNR were significantly poorer compared to 5 dB SNR for both types of noise.

![10-Talker Babble vs. EN](image)

Figure 4. Mean percent correct scores for 10-talker babble (blue bars) and ecological noise ([EN] red bars) at 0 dB and 5 dB SNR. Vertical lines represent $\pm 1$ standard deviation (SD) and asterisks represent significant differences (*$p < 0.05$).
HINT Speech Spectrum Noise vs. Ecological Noise

Figure 5 shows the HINT speech perception scores for the SSN and the EN. Mean percent correct scores for the EN and the SSN at 0 dB SNR were 66.06 and 96.11, respectively, whereas the mean percent correct scores at +5 dB SNR were 96.43 and 99.65. Post hoc Tukey comparisons revealed that speech perception scores using the EN were significantly poorer than in speech spectrum noise at 0 dB ($p < .0001$) but there was no statistically significant difference at +5 dB SNR ($p > 0.3792$) for the HINT sentences. For the EN, speech perception scores at 0 dB SNR were significantly lower than those at +5 dB SNR, but scores at 0 dB and +5 dB SNR were not statistically different for SSN.

![SSN vs. EN](image)

**Figure 5.** Mean percent correct scores for SSN noise (yellow bars) and ecological noise ([EN] red bars) at 0 dB and 5 dB SNR. Vertical lines represent ± 1 standard deviation (SD) and asterisks represent significant differences (*$p < 0.05$).
Discussion

In this study, speech perception performance of individuals with normal hearing using an ecological noise was compared to their performance with other types of noise. To assess speech perception in a more realistic context than created by traditional noise types, the EN was designed using a variety of environmental sounds (stadium noise, live music, traffic noise, birds chirping, etc.) that were either downloaded from the internet or recorded from natural environments. All noises were added in different tracks using Adobe Audition software with the addition of 4-talker babble, and a single sound file was created. The results showed that speech perception scores using EN were significantly poorer than 4-talker babble, 10-talker babble, or SSN at 0 dB SNR. However, at +5 dB SNR, speech perception scores were significantly lower only for the EN compared to 4-talker babble, while the EN scores were statistically equivalent to those for 10-talker babble, and SSN.

It is likely that the EN was more difficult because of its combination of linguistic and non-linguistic information with fluctuating frequency and intensity components from the different environmental noises used. MTB noises (4-talker and 10-talker babble) contain only linguistic information, and this linguistic information likely interferes with the intelligibility of the target stimuli. In contrast, SSN does not have any linguistic information, and its intensity across all frequencies is constant, steady, and unmodulated, which does not interfere as much with the intelligibility of target stimuli (Sperry et al., 1997). Our results also showed poorer performance with the QuickSIN 4-talker babble and AzBio 10-talker babble compared to the HINT (SSN). These results are in agreement with Sperry et al. (1997) who measured speech recognition ability using 6-talker babble, the same 6-talker babble in reverse order, and SSN. They found that meaningful information in the babble noise had a more deleterious effect on
speech recognition scores than when no meaningful information was present in the other two noise conditions. In the current study, an informal verbal assessment of the naturalness of the noises from some participants after finishing their session supported the premise that the EN sounded more realistic than MTB and SSN.

**Energetic Masking (EM) and Informational Masking (IM) Effects**

Despite the fact that babble contains useful linguistic information, as the number of talkers increases, the spectrum of the babble flattens, making it more like SSN, with more EM and less IM. As a result, the most effective babble noises are one-talker, two-talkers, and four-talkers. Babble with fewer talkers has a greater IM effect, while SSN has a greater EM effect. As the number of talkers in the babble increases, the IM effect decreases and the EM effect increases (Rosen et al., 2013). The same effect was observed in the present study as well. The speech perception scores for the 4-talker babble were worse than the 10-talker babble. In addition, our EN has both IM (from 4-talker babble) and EM (from different environmental noises) effects, resulting in a combination of both types of masking effects, making it more realistic. This likely accounts for the greater decrease in the intelligibility of the target sentences observed here when using EN. In comparison to EM induced by SSN, the IM caused by the babble not only impairs speech perception but also increases cognitive load. At the cognitive level, the information contained in the babble interferes by activating phonetic, semantic, and linguistic systems similar to speech signals. When the SNR decreases, the activation of these systems at the cognitive level increases the mismatch between the speech information in long-term memory and the speech signal itself, resulting in even poorer speech perception (Rönnberg et al., 2010; Schneider et al., 2007; Zhang et al., 2014).
Effect of SNR

A decrease in the SNR decreased speech perception scores. All scores at 0 dB SNR were poorer than at 5 dB SNR for EN, 4-talker, and 10-talker babble but not for SSN. The mean speech perception scores for all three tests (see Figures 1, 2, and 3) clearly showed that the effect of SNR was greatest for the EN, followed by the 4-talker babble, the 10-talker babble, and the SSN, respectively. This result is in agreement with previous studies in the speech perception literature using different SNRs (McArdle et al., 2005; Wilson, 2003; Wong et al., 2008). Findings from these studies postulated that the presence of a poorer SNR causes more stress on the auditory system for encoding the auditory information available in the sentences presented as the stimuli. Additionally, an increase in the level of noise increases the cognitive load, which results in a phonological mismatch between the phonological elements perceived and the phonological representations in long-term memory, resulting in impaired speech perception (Rönnberg et al., 2010). The presence of a poorer SNR not only causes more stress on the auditory system to encode the auditory information available in the sentences presented as the stimuli but also shows an increased activation in other parts of the brain (Bidelman & Howell, 2016; Wong et al., 2008). For these reasons speech perception performance was poorer at 0 dB SNR compared to +5 dB SNR in the current study.

We found it interesting that the only condition where speech perception scores under EN were significantly different from the other noises at both 0 and +5 dB SNR was for the 4-talker babble of the QuickSIN. The EN was significantly different from the 10-talker babble and SSN for the AzBio and HINT, respectively, at 0 dB SNR but not at +5 dB SNR. This finding could be due to the fact that the QuickSIN has been shown to be more sensitive in assessing speech perception performance in the presence of noise compared to the AzBio and the HINT (Holder et
It is also possible that at higher SNRs, the IM effect of 4-talker babble in the EN begins to decrease because the speech is louder than the noise, while the EM effect from the environmental sounds in the ecological noise begins to increase, similar to 10-talker babble and SSN. As a result, the EN becomes easier, eliciting a similar behavioral response as 10-talker babble and SSN at +5 dB SNR.

**Acoustical Characteristics**

Spectral analysis of the EN (Figure 1) revealed that the frequency-domain characteristics of the EN appeared to be relatively similar to the other noises used in this study. The main difference was that the EN had lower energy between the 2nd to 5th band (200 Hz to 510 Hz), and 19th to 21st band (5300 Hz to 7700 Hz) compared to the other noises, while frequency bands typically containing speech cues were not different. This would suggest that the masking effect of the EN is not attributable to a simple energy effect. Temporal modulation analysis (Figure 2), however, revealed that the EN is relatively strongly modulated in the low frequency range ~0.625 Hz (a 1.6 s period), while the 4-talker babble modulation was also relatively modulated at ~0.468 Hz compared to other noises. Based on these results, the slow modulation of the EN might have provided glimpses on some sentences, again suggesting reduced masking. Alternatively, listeners might have been distracted due to the nature of the modulation in the EN. This is especially true as this modulation rate is somewhat similar to the rate of the sentence presentation, and this might have created an attentional distraction.
Limitations

There were a few limitations present in this study. It is possible that the additional environmental sounds together with the MTB in the EN made the EN more challenging than the other noises used in this study. Because the EN included several different sounds with varying intensity, timing, and frequency components, it is possible that some speech signals were affected differently by the EN based on when the speech signal occurred during the noise track. In addition, there are a number of other environmental sounds that could have been included in the EN. SNRs were varied in 5 dB steps; smaller dB step sizes might have provided more precise results. Further, only individuals with normal hearing participated in this study, therefore there is limited generalizability of these findings to those who have hearing loss. Finally, only adult participants were recruited for this study. Including younger adults and older adults would have contributed to more generalizability of the results from this study.

Conclusion

In the present study, speech perception performance of individuals with normal hearing was evaluated using a newly developed EN to simulate a realistic environment of noise and was compared with results obtained under 4-talker babble, 10-talker babble, and SSN. Behaviorally, speech perception scores using the EN were significantly poorer compared to when MTB or SSN was used as the noise. This could be because the EN consists of both informational masking effects from the MTB and energetic masking effects from the environmental sounds, whereas the MTB only has the IM effect and SSN only has the EM effect.

Acoustically, the spectral characteristics of the EN were relatively similar to the other noises, whereas the temporal analysis of the EN showed a stronger low-frequency modulation at ~0.625 Hz (a 1.6 s period). The spectral characteristics of the EN were effectively quite similar
to the other noises, so this aspect does not seem to explain the behavioral results. Contrastively, it is possible that the addition of the 4-talker babble with the environmental noises in the EC caused the strong low-frequency modulation and may have interfered with the participants’ attention to the rate of sentence presentation.

The ecological noise used here presents a more challenging listening situation than the use of MTB or SSN by themselves. Results from the current study showed that the speech perception scores measured using this EN may give a more realistic expectation about communication in the natural environment compared to MTB or SSN and would be helpful for counselling people with hearing impairment.

The fact that the MTB was included in the EN made it difficult to determine its influence on these results. Future studies are underway that focus on the effect of the EN on speech perception with and without MTB included to help answer this question. Though it seems clear from this study that the spectral characteristics of the EN alone did not explain why speech perception was more challenging in that condition compared to the other noises, further investigations involving various combinations of noise and MTB are needed in order to distinguish between different cognitive factors such as attention from effects due to basic signal characteristics such as those presented here.
Chapter III

Influence of Multi-talker Babble on Speech Perception Performance in Ecological Noise

Introduction

The ability of the human auditory system to perceive speech is a complicated process. Speech perception may be affected by anatomical or physiological issues in the auditory system, as well as cognitive variables and the presence of external background noise. Even for people with normal hearing sensitivity, understanding speech in the presence of background noise can be challenging. Most people with hearing loss find it particularly difficult to recognize speech in noise (Best et al., 2013; Lee et al., 2015; McArdle et al., 2005; Smoorenburg, 1992; Wilson et al., 2007). Previous research suggests that someone's ability to understand speech in noisy environments cannot be predicted based on their ability to understand speech in quiet (Carhart & Tillman, 1970; Eisenberg et al., 2005; Heinrich et al., 2015). Assessment of speech perception in noise is crucially important in audiology because it plays an extremely important part in the design of audiological rehabilitation and the selection of amplification devices. Additionally, it provides information that can be used to counsel patients about realistic expectations from rehabilitation (Levine et al., 2010; Wilson, 2004).

Speech perception tests mainly utilize two types of noise to measure speech perception. Speech spectrum noise (SSN) is a type of stationary background noise that has the same long-term average speech spectrum (LTASS) as speech but does not actually include any speech sounds. SSN is produced by filtering white noise to match the LTASS of spoken sounds (Van Engen et al., 2014). Multi-talker babble (MTB) is a noise that is produced when multiple people speak simultaneously, i.e., the sum of the waveforms of multiple simultaneous talkers (Van
Engen et al., 2014). A test like the Hearing in Noise Test ([HINT], Nilsson et al., 1994) uses SSN, whereas tests like the Quick Speech in Noise test ([QuickSIN], Killion et al., 2004), and the AzBio Sentence test (Spahr et al., 2012) utilize MTB. Though the speech perception scores obtained using these noises provide an accurate estimate of someone’s speech perception in the presence of noise, the results are far from realistic because the noises used in these tests are generated in a laboratory set up and are highly controlled in terms of their physical characteristics.

Traditional laboratory tests, for example, attempt to predict speech perception in a noisy setting where the physical characteristics of noise (intensity and frequency) are stable or controlled, yet such situations are not very natural compared to a subway, for example, where the characteristics of the noise vary considerably (Best et al., 2013; Lee et al., 2015; Revit et al., 2002). There have been a few studies in the literature that have utilized more ecologically relevant noise to measure speech perception in an attempt to create a more controlled environment, and the majority of these noises were different forms of MTB, such as 4-talker babble, 8-talker babble, 40-talker babble, etc., along with several environmental auditory scenes (e.g., church noise, vacuum noise, street noise, dinner party, etc.) (Best et al., 2013; Lee et al., 2015; Weisser & Buchholz, 2019). These MTB noises and auditory scenes were used independently to measure speech perception; however, combining these noises together might yield more accurate speech perception scores. While babble noise is regarded as environmentally appropriate (Fikret-Pasa, 1993; Killion et al., 2004), there are a variety of other sounds in the physical realm that may also play a significant role in inhibiting communication. Natural noise includes all accessible sounds, such as chirping of birds, rustling leaves, wind, road noise, and so on, in addition to MTB. The impact that these noises may have on speech understanding justifies
the need to include all possible noises that a listener might experience in everyday communication in the assessment of speech perception performance.

Shukla et al. (2022) developed an ecological noise (EN), which included several environmental noises and MTB, that was designed to simulate a significantly more realistic environment than other ENs used previously in the literature. They argued that it would make sense to incorporate all possible sounds encountered by a listener in everyday communication in the evaluation of speech perception to enhance accuracy, and it would be more effective and beneficial to combine these scenes into a single noise (including MTB) that can be used in the audiologic test battery (Shukla et al., 2022). The EN was comprised of a variety of different environmental sounds, including those from a carnival, a dog barking, traffic, a car engine, rain, a waterfall, a railway station, birds chirping, crowd applause, and stadium noise, as well as party crowd noise, live music, club and crowd noise, and peak hour traffic noise. These noises were selected because they are representative of the everyday noises that people experience when attempting to communicate. All the noises were dispersed across eight distinct tracks, the entire session was normalized to a -3 dB range, and one sound file was produced. To make it more ecologically valid, 4-talker babble from the QuickSIN (Killion et al., 2004) was combined with additional environmental noises to create a new sound file.

Shukla et al. (2022) evaluated the speech perception performance of individuals with normal hearing using this EN and compared it to other types of noise. Sentences from the AzBio, QuickSIN, and HINT tests were utilized with their corresponding noises (4-talker babble, 10-talker babble, and SSN, respectively) along with the EN to evaluate speech perception. Their results showed that speech perception scores for the EN were statistically different from scores
obtained with the other noises. Speech perception performance was poorest for EN, followed by 4-talker babble, 10-talker babble, and SSN. Additionally, speech perception performance at 0 dB SNR was considerably poorer than at +5 dB SNR for all noises.

These findings suggest that the EN developed by Shukla et al. (2022) created a more challenging listening environment compared to MTB or SSN. Such an ecological noise could be useful as part of the basic audiological battery for speech perception testing. Speech perception scores evaluated with this ecological noise could offer a more realistic expectation of communication in the natural world compared to MTB or SSN, which is beneficial for counseling individuals with hearing loss. The outcomes of that study also raised several critical questions about what impact the presence or absence of babble in the ecological noise had on speech perception performance. What made this EN more challenging than the other noises? Was it the MTB or the environmental sounds that made it more challenging?

To address these questions, the objective of the present study was to determine whether the addition of MTB within the EN had a significant effect on speech perception scores. In the current study, speech perception performance of individuals with normal hearing was assessed using EN noise with no babble (ENNB), EN noise with babble (ENWB), and EN with reverse babble (ENRB). In addition, the spectral and temporal distribution of these different forms of the EN were also evaluated.
Method

Creation of a New Ecological (EN) noise

The original EN used by Shukla et al. (2022) was generated as a single sound file that contained all the environmental sounds and MTB together, so it was impossible to remove the MTB to reverse it. Therefore, a new EN was created following the same procedure utilized in Shukla et al. (2022). To create the new EN, all the environmental sounds that were downloaded and originally recorded were laid on eight different tracks as shown in Table 2. Once all the noises were set in tracks, the files/tracks were combined, and the entire session was normalized to the -3 dB range, and a .wav file was generated, which was referred to as ecological noise with no babble (ENNB). Normalization is a process of setting the loudest component of the waveform to a certain amplitude by raising or lowering the rest of the signal's amplitude to that level. To generate an EN with babble, the 4-talker babble from the QuickSIN was added to the ENNB and a new sound file was created (ecological noise with babble [ENWB]). To generate ecological noise with reverse babble, the same 4-talker babble in the ENWB was reversed and another sound file was created (ecological noise with reverse babble [ENRB]). The noises were sampled at 44,100 Hz with a 32-bit resolution. All the created noises were 41 seconds long in duration and were looped for 10 minutes, which was enough time to test speech perception in one condition.
Table 2. Distribution of different environmental noises across all eight tracks of the new ecological noise (EN).

<table>
<thead>
<tr>
<th>Track No.</th>
<th>Noise</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Carnival noise followed by a dog barking</td>
</tr>
<tr>
<td>2</td>
<td>Traffic followed by a car engine noise</td>
</tr>
<tr>
<td>3</td>
<td>Rain followed by a waterfall noise</td>
</tr>
<tr>
<td>4</td>
<td>Party crowd followed by railway station noise</td>
</tr>
<tr>
<td>5 and 6</td>
<td>Birds chirping followed by crowd applause</td>
</tr>
<tr>
<td>7</td>
<td>Live music, party crowd, club noise followed by stadium noise</td>
</tr>
<tr>
<td>8</td>
<td>Peak hour traffic noise</td>
</tr>
</tbody>
</table>

**Evaluation of Speech Perception**

**Participants**

A total of 20 individuals (aged 18 to 40 years; M = 25.5, SD = 5.43) were recruited for the study. Participants were graduate students from The University of Memphis as well as individuals from the city of Memphis. All subjects were native English speakers with audiometric thresholds better than 20 dB HL in both ears at 500, 1000, 2000, and 4000 Hz. Participants had normal middle ear functioning and normal cognitive abilities. To participate in the study, all participants needed to have a normal otoscopic examination, normal hearing thresholds (American-Speech-Language-Hearing-Association [ASHA], 2005), normal middle ear function as demonstrated by a Type A tympanogram (ASHA, 1988), and no history of any neurologic condition. Once the subjects satisfied these requirements, a cognitive screening was
conducted to evaluate their cognitive performance. Participants were excluded from the study if they did not match any of the aforementioned criteria.

**Stimuli**

Sentence lists from the QuickSIN, AzBio, and HINT were used as the speech stimuli to measure speech perception performance in the presence of three different ecological noises created for this study. The QuickSIN is composed of IEEE (Rothauser, 1969) sentences presented in four-talker babble. QuickSIN lists have six sentences, and each sentence has five key words that the listener must repeat correctly. The AzBio sentence test assesses speech perception abilities in people with hearing loss and cochlear implant users. The AzBio sentence test is comprised of 15 different sentence lists, with 20 sentences in each list. Ten sentences in each list are spoken by two male speakers, while the remaining sentences are spoken by two female speakers in the presence of 10-talker babble. The HINT is a sentence recognition test that determines how effectively a person can understand speech in both quiet and noisy environments. The HINT uses Bamford-Kowal-Bench (BKB) sentences (Etymotic, 2005), and it consists of 25 equivalent lists of ten sentences that are phonetically balanced at a first grade reading level. The sentences from the QuickSIN, AzBio, and HINT were presented using the three noises created for this study (ecological noise with no babble (ENNB), ecological noise with babble (ENWB), and ecological noise with reverse babble (ENRB)).

**Procedure**

The participants signed an informed consent that was approved by the University of Memphis Institutional Review Board. They also completed a comprehensive audiological evaluation, which included an otoscopic examination, tympanometry (Maico-MI34), pure tone audiometry (GSI-61), and a cognitive screening using the Montreal Cognitive Assessment
([MoCA], Nasreddine et al., 2005). The MoCA is a cognitive screening tool that is simple and quick to use, and it evaluates moderate cognitive disorders. The test evaluates areas such as visual-spatial, naming, language, memory, attention, abstraction, and delayed recall.

Following the complete audiological assessment, speech perception performance was evaluated using the various tests and noises for the conditions listed above. The test materials were presented using compact disks (CDs) via two different CD players that were connected to the speaker in the sound booth via an audiometer (GSI 61). All speech stimuli were delivered using the Sony (RCD-W500C) CD player, while all ecological noises were delivered via the Onkyo (DX-C606) CD player. There was a total of 18 conditions: 2 lists from each test x 3 sentence tests x 3 EN conditions. The test conditions were presented in random order, and the lists were counterbalanced to reduce any learning and practice effects.

The study was conducted in an air-conditioned, sound-treated, double-walled room that met ANSI S3.1-1999 (R2008) criteria, and the presentation level of the noise was 50 dB HL, and the presentation level of the speech was 55 dB HL (i.e., +5 dB SNR). To ensure that all stimuli and noises were delivered at the same level, all test materials were calibrated using root mean square (RMS). The speech signal and the noise were both presented in the sound field at an azimuth of 0 degrees, while the participant was seated at a distance of 1 meter away from the sound source. The participants were instructed to listen to the speech stimulus and repeat it, and they were also advised to ignore the background noise and concentrate on the speech. Each word in each sentence was worth one point, and the responses were calculated based on the percentage of words repeated correctly by the participants. Scoring leniency was given to only singulars or plurals. For example, a point was awarded if a participant repeated "beds" instead of "bed" or vice versa. Intra-judge scoring reliability was tested on 30% of the data using the following...
formula: \[\text{[agreements/ (agreements + disagreements)] \times 100\%}\], and intra-judge scoring reliability was 99%. The scoring was performed in a quiet room using a laptop with the volume set to a comfortable level using Bose headphones (Bose QuietComfort 35 II).

**Acoustical analysis**

In addition to evaluating speech perception, this study also investigated the acoustical characteristics of all the ecological noises (ENNB, ENWB, and ENRB). To characterize each of the ENs, both spectral and temporal analyses were carried out. For the spectral analysis, the PRAAT program was utilized to extract the acoustical energy present at each frequency for each noise (Boersma, 2006). Since each signal had the same sampling rate and duration, 1,048,576 positive frequency bins were produced, and thereafter, these magnitudes were transformed into Bark-scaled bins, which corresponded to the first 24 critical bands of hearing (Zwicker, 1961). The means were then normalized to correspond with the presentation conditions. For the temporal analysis, we modified the approach provided by Tilson and Johnson (2008) to investigate the modulation of the entire frequency range, rather than just 700-1,300 Hz. To obtain the envelope information, each noise was low-pass filtered using a fourth-order Butterworth filter with a cutoff of 10 Hz, down sampled to 80 Hz, and adjusted for phase delay. Finally, the frequency domain representation of the time-domain amplitude envelope was obtained using a Fast Fourier transform, which revealed the magnitude of envelope modulation at each sampled frequency.
Results

Acoustical analysis

Figure 6 shows the spectral distribution of all three noises (ENNB, ENWB, and ENRB), revealing that the spectral characteristics of the ENWB and ENRB were the same, however the spectral distribution of the ENNB was different from the spectra for the ENWB and ENRB. Spectral energy for ENNB was lower between the 2\textsuperscript{nd} to 7\textsuperscript{th} critical bands and higher after the 7\textsuperscript{th} to 24\textsuperscript{th} critical bands compared to ENWB and ENRB. Temporal modulation analysis (Figure 7), however, revealed that ENWB and ENRB were highly modulated at low frequencies compared to ENNB.
Figure 6. Spectrum of ecological noise with no babble ([ENNB], green line), ecological noise with babble ([ENWB], red line), and ecological noise with reverse babble ([ENRB], black line). The X axis represents critical bands (Bark scale), and the Y axis represents normalized intensity in dB. The Spectral characteristics of the ENWB and ENRB were the same, however the spectral characteristics of the ENNB were different. The spectra of the ENWB (red line) and ENRB (black line) were identical because they contained the same input.
Figure 7. Envelope spectrum of ecological noise with no babble ([ENNB], green line), ecological noise with babble ([ENWB], red line), and ecological noise with reverse babble ([ENRB] black line). The Y axis represents power, and the X axis represents frequency in Hz. ENWB and ENRB were highly modulated at low frequencies compared to ENNB.

**Speech perception performance**

The speech perception scores using the three ecological noises (ENNB, ENWB, and ENRB) were compared for all three tests (QuickSIN, AzBio, and HINT). Speech perception performance was compared across the noise conditions within each test as well as across tests within each type of noise. Considering the data were not normally distributed, Kruskal-Wallis one-way Analysis of Variance (ANOVA) on Ranks and Tukey pairwise comparisons were applied to determine if there were significant differences in speech perception performance.
among the noise and test conditions. Speech perception scores were the dependent variables, and
the three different noise conditions were the independent variables. Kruskal-Wallis ANOVA
results revealed a significant main effect (Chi Square = 103.28, $p < .0007$, df=8). The post hoc
Tukey pairwise comparisons evaluated the differences in speech perception scores across the
individual ENs and the three speech perception tests and are described below.

**Speech Perception Scores Across the Noise Conditions Within Each Test**

*QuickSIN.* Figure 8 shows the speech perception scores for QuickSIN sentences for all
the EN conditions. The mean speech perception scores for QuickSIN sentences were 88.30,
81.62, and 85.92 for ENNB, ENWB, and ENRB, respectively. Post hoc Tukey comparisons
revealed that speech perception scores using the ENNB were significantly different from the
ENWB ($p < .05$); however, speech perception scores between ENNB and ENRB were not
significantly different. Similarly, speech perception scores between ENRB and ENWB were not
significantly different.
Figure 8. Mean speech perception scores for QuickSIN sentences using ecological noise with no babble ([ENNB], green bar), ecological noise with babble ([ENWB], red bar), and ecological noise with reverse babble ([ENRB], black bar). Vertical lines on each bar represent ±1 standard deviation (SD) and the asterisk represents a significant difference (*p < .05).

**AzBio.** Figure 9 shows the speech perception scores for AzBio sentences under all the EN conditions. The mean speech perception scores for AzBio sentences were 96.31, 96.17, and 97.08 for ENNB, ENWB, and ENRB, respectively. Post hoc Tukey comparisons revealed that there were no significant differences in speech perception scores obtained between the ENs for this test.
Figure 9. Mean speech perception scores for AzBio sentences using ecological noise with no babble ([ENNB], green bar), ecological noise with babble ([ENWB], red bar), and ecological noise with reverse babble ([ENRB], black bar). Vertical lines on each bar represent ±1 standard deviation (SD).

**HINT.** Figure 10 shows the speech perception scores for HINT sentences under all the EN conditions. The mean speech perception scores for HINT sentences were 98.67, 98.72, and 99.12 for ENNB, ENWB, and ENRB, respectively. Post hoc Tukey comparisons revealed that there were no significant differences in speech perception scores obtained between all the ENs for the HINT.
Figure 10. Mean speech perception scores for HINT sentences under ecological noise with no babble ([ENNB], green bar), ecological noise with babble ([ENWB], red bar), and ecological noise with reverse babble ([ENRB], black bar). Vertical lines on each bar represent ±1 standard deviation (SD).

**Speech Perception Scores Within Each Noise Condition Across Tests**

We further analyzed speech perception scores for each test within each EN. Figure 11 shows speech perception scores for the QuickSIN, AzBio, and HINT sentences under ENNB, indicating significant differences. Speech perception scores for QuickSIN sentences were significantly different from AzBio and HINT sentences ($p < .05$). Similarly, speech perception scores for AzBio sentences were significantly different from HINT sentences ($p < .05$). Figures 12, and 13 show similar results were found for the other two noise conditions (ENWB and ENRB), where significant differences were also found among all tests.
Figure 11. Mean speech perception scores for QuickSIN (green bar), AzBio (red bar), and HINT (black bar) sentences under ecological noise with no babble (ENNB). Vertical lines on each bar represent ±1 standard deviation (SD) and asterisks represent significant differences (*p < .05).
Figure 12. Mean speech perception scores for QuickSIN (green bar), AzBio (red bar), and HINT (black bar) sentences under ecological noise with babble (ENWB). Vertical lines on each bar represent ±1 standard deviation (SD) and asterisks represent significant differences (*p < .05).
Figure 13. Mean speech perception scores for QuickSIN (green bar), AzBio (red bar), and HINT (black bar) sentences under ecological noise with reverse babble (ENRB). Vertical lines on each bar represent ±1 standard deviation (SD) and asterisks represent significant differences (*p < .05).

Discussion

The present study investigated the influence that the presence of multi-talker babble had on assessing speech perception performance using ecological noise generated by Shukla et al. (2022). There were three different ecological noises used in this study: (1) ecological noise with no babble, (2) ecological noise with babble, and (3) ecological noise with reverse babble. Speech perception was measured using three tests - namely QuickSIN, AzBio, and HINT.
Speech Perception Performance Within Each Test Across the Noise Conditions

Our findings demonstrated that there was a significant difference in the speech perception scores for the ENNB and ENWB conditions only for the QuickSIN. In addition, the speech perception scores under ENWB were poorer compared to ENRB and ENNB for the QuickSIN. These findings suggest that the presence of multi-talker babble clearly had an influence on speech perception. Because ENWB combines linguistic and non-linguistic information together with the varying frequency and intensity components of the various environmental noises, speech perception scores were affected. Further, the fact that speech perception scores were not affected in the ENNB or the ENRB conditions reinforces the fact that the addition of linguistic information that is considered meaningful had an effect on speech perception. These findings are consistent with those of Sperry et al. (1997), who assessed speech recognition ability using a 6-talker babble, a 6-talker babble in reverse order, and SSN. They found that the presence of meaningful information in the MTB had a greater detrimental influence on speech recognition scores compared to the absence of meaningful information in the other two noise situations. Similar results were also found by Shukla et. (2022), who measured speech perception in ecological noise and compared it with 4-talker babble, 10-talker babble, and SSN. They found that speech perception scores using EN with MTB were significantly poorer than when MTB or SSN were used alone (Shukla et al., 2022).

In the current study, the scores for the AzBio and the HINT were always higher than those for the QuickSIN in each noise condition. These very high scores made it difficult to see any additional performance differences across the noise conditions, as no other significant results were found for the remaining noise conditions. It is possible that the lack of significant findings for the other noise conditions was due to the fact that the participants' speech perception scores
reached the ceiling. The ceiling effect is seen when an independent variable (noise conditions) no longer affects a dependent variable (speech perception scores). When a ceiling effect occurs, speech perception scores reach at or near the possible upper limit of the test, and any differences that might occur between conditions cannot be seen due to the lack of variance. A similar effect was also seen in the current study where speech perception scores for all the noise conditions (ENNB, ENWB, and ENRB) reached maximum for the AzBio and HINT. Similarly, scores for ENRB and ENNB for the QuickSIN also reached ceiling and that likely affected our ability to measure any significant differences.

Speech Perception Performance Within Noise Conditions Across Tests

The results of the current study also showed that speech perception scores for the QuickSIN were always poorer than the AzBio, followed by the HINT under ENWB, followed by ENRB and ENNB. The reason why QuickSIN scores were poorer in all noise conditions could be related to the sensitivity of the test. The sensitivity of a test is defined by its validity, and reliability, and its ability to accurately measure speech perception performance. A test is considered to have face validity if it is found to measure what it was designed to measure. Whereas the degree to which a test produces consistent and stable results with repeated administrations is referred to as reliability. The greater the sensitivity of the test, the better its ability to accurately assess performance or identify a disorder. Past research has shown that the QuickSIN is more sensitive to assessing speech perception in noise than the AzBio or HINT (Holder et al., 2018; Shukla et al., 2022; Sultan et al., 2020; Wilson et al., 2007). A similar pattern was also observed by Shukla et al. (2022), where speech perception scores were always poorer for the QuickSIN compared to the AzBio and HINT (Shukla et al., 2022). The findings of the present study are in agreement with Sultan et al. (2020), who found that the sensitivity of the
QuickSIN was higher than the HINT. They plotted receiver operating characteristic (ROC) curves showing the sensitivity of the QuickSIN was 88.9% compared to the sensitivity of the HINT (72.7%). Similarly, Wilson et al. (2007) found that tests such as the QuickSIN and the Words in Noise (WIN) test are more sensitive measures to determine speech perception performance in background noise compared to the BKB-SIN and HINT materials. In another study, Holder et al. (2018) found that the AzBio sentence test reached ceiling at SNRs greater than 0 dB compared to the QuickSIN and BKB-SIN (Etymotic, 2005).

The research reviewed above suggests the QuickSIN is the most sensitive of the three tests used here for measuring speech perception in the presence of noise. Perhaps because the QuickSIN employs 4-talker babble, which was the babble used in Shukla et al. (2022) and in the current study, it is more sensitive to that particular babble. Given the consistency of QuickSIN results found in Shukla et al. (2022) and in the present study, the current findings suggest that the QuickSIN is the most sensitive test for measuring speech perception in the presence of this EN as well.

**Acoustical Analysis**

Acoustical analysis of the ENs used here showed that the spectral characteristics of the ENWB and ENRB were the same given that the same babble was used in both noises but just reversed. However, as expected, the spectral distribution of the ENNB was different from the ENWB and ENRB. Spectrally, the limited energy between critical bands 2–7, which corresponds to the 150 Hz -700 Hz range, might have been the reason why the scores were better under ENNB compared to ENWB. Previous studies have shown that low-frequency noises at moderate
levels cause more masking than high-frequency noises (Egan & Hake, 1950; Kalafata & Persson Waye, 2017).

Temporal modulation analysis, however, revealed that ENWB and ENRB were highly modulated compared to ENNB. Temporally, ENWB and ENRB showed modulation in the low frequencies compared to ENNB, which had a flat spectrum at those frequencies. This low-frequency modulation might have distracted the listeners because the rate of modulation was similar to the rate of sentence presentation in the speech signals. This might have increased the modulation interference (MI), which could be the reason that the speech perception scores were poorer in ENWB (Yost & Sheft, 1994). MI occurs when the modulation of a masker and target signal is similar, resulting in a loss of sensitivity in detection and discrimination of amplitude modulation of that signal (Yost & Sheft, 1994). The results of the current study are also in agreement with Kwon and Turner (2000), who found higher MI when the rate of masker modulation was within the range of speech modulations. Even though ENWB and ENRB were highly modulated at lower frequencies, the presence of linguistic information from the 4-talker babble in the ENWB may have produced an additional distraction, which could have interfered with the intelligibility of the speech signals. This may explain why ENWB was even more challenging than ENRB. Despite the fact that ENRB was modulated in the low frequencies compared to ENNB, speech perception scores for both the noises were similar. Further, the scores reached ceiling which could be why no significant differences were found. Additional studies using varying SNRs are needed to reach to a clearer distinction between the effects of ENRB and ENNB on perception.
Limitations

This study had a few drawbacks. Speech perception was only examined at an SNR of +5 dB, and the responses hit ceiling, limiting the generalizability of the results. The ceiling effect caused speech perception scores to reach at or near the possible upper limit, making it difficult to see any statistically significant differences if they existed. More challenging SNRs (e.g., +2 dB, 0 dB, etc.) could possibly eliminate the ceiling effect and may provide more accurate findings. In addition, only individuals with normal hearing participated in this study, limiting the applicability of these results to those with hearing loss. Assessing performance from listeners with hearing loss should minimize the ceiling effects which will potentially provide additional information regarding the effect of EN.

Conclusion

The present study investigated whether the addition of MTB within the EN had a significant effect on speech perception scores in individuals with normal hearing. The influence of MTB was evaluated using QuickSIN, AzBio, and HINT using three different ENs (ENNB, ENWB, and ENRB). Behaviorally, results from the current study showed that there was an influence of MTB on speech perception only for the QuickSIN. It is probable that the combined linguistic and non-linguistic information in the ENWB, together with the varying frequency and intensity components of the various environmental noises, is likely what interfered with the intelligibility of the QuickSIN sentences. The scores on the AzBio and HINT reached the ceiling, which limited our ability to determine any other significant results. The current study also found that speech perception scores for QuickSIN were significantly poorer compared to the AzBio, followed by the HINT under ENWB, followed by ENRB, and ENNB. These findings suggest the
QuickSIN is more sensitive to measuring speech perception in noisy environments than the AzBio and HINT.

Acoustically, ENNB showed less spectral energy in the low frequencies, whereas the temporal analysis revealed that the ENWB and ENRB had stronger low frequency modulation. There is a possibility that the higher spectral energy of ENWB was responsible for a greater masking effect than ENNB, leading to poor behavioral performance. The addition of the 4-talker babble to the environmental sounds in the EN, on the other hand, may have created the strong low-frequency modulation and interfered with the participants' attention to the rate of sentence presentation. Because the scores reached the ceiling, generalizability of the findings is difficult. Future research should focus on incorporating several more challenging SNRs with smaller step sizes in order to have a better understanding of the influence that MTB in EN has on speech perception.
Chapter IV

General Conclusion

Previous research in the field of audiology has evaluated speech perception in individuals with normal hearing and hearing loss using a variety of noises such as MTB, SSN, and several other environmental auditory scenes that had MTB as the dominant noise. These MTB noises and environmental auditory scenes have been utilized independently to measure speech perception performance. Babble is considered ecologically important due to its varying frequency and intensity over time (which is similar to many environmental noises) because it is reflective of restaurant and social settings where multiple people speak at the same time. Given that there are a number of noises in the real world that may impede communication, all noises that a listener can experience in everyday communication should be included in the evaluation of speech perception accuracy. It would be efficient and beneficial to combine various auditory scenes into a single noise (including MTB) for audiologic testing of speech perception. The current work attempted to address the aforementioned method by combining several naturalistic auditory scenes that have been utilized in the past into a single noise (including MTB) to test speech perception.

The goal of this dissertation was to evaluate speech perception in individuals with normal hearing using an ecological noise that was created to simulate a natural setting. The EN produced here includes a variety of environmental noises and MTB that can simulate a much more realistic environment than those previously utilized in the literature. These studies also measured the spectral and temporal properties of the EN as well as other noises, including 4-talker babble, 10-talker babble, and SSN.
This dissertation was divided into two parts. In the first study, we created an EN that contained natural noises and the 4-talker babble from the QuickSIN. We then assessed speech perception performance of individuals with normal hearing using this EN and compared it to several other types of noise (4-talker babble alone, 10-talker babble from the AzBio, and SSN from the HINT). Behaviorally, speech perception scores using the EN were significantly poorer compared to the 4-talker (QuickSIN), and 10-talker babble (AzBio) as well as the SSN (HINT). We concluded that these findings occurred because the EN had both informational masking effects from the 4-talker babble and energetic masking effects from the environmental sounds, while the 10-talker babble only had the IM effect, and the SSN only had the EM effect.

Acoustically, the spectral characteristics of the EN were relatively similar to the other noises, whereas the temporal analysis of the EN showed a stronger low-frequency modulation. The spectral characteristics of the EN were effectively quite similar to the other noises, so this aspect does not seem to explain the behavioral results. Contrastingly, it is possible that the addition of the 4-talker babble with the environmental noises in the EN caused the strong low-frequency modulation and may have interfered with the participants’ attention to the rate of sentence presentation.

The first study concluded that the ecological noise developed for this research presents a more challenging listening situation than the use of MTB or SSN alone. Speech perception scores measured using this EN may give a more realistic expectation of communication in the natural environment compared to MTB or SSN and would be helpful for counseling people with hearing loss. The outcomes from the first study raised several critical questions about what impact the presence or absence of babble in the ecological noise had on speech perception.
performance. The results made us question what made this EN more challenging than the other noises. To address this question, the objective of the second study was to determine whether the addition of MTB within the EN had a significant effect on speech perception scores.

In the second study, the speech perception performance of individuals with normal hearing was assessed using the EN with no babble (ENNB), with babble (ENWB), and with reverse babble (ENRB). In addition, the spectral distribution of these different forms of the EN was also evaluated. The MTB in the EN utilized in the first study (original EN) could not be removed to reverse it because the original EN sound file was created as a single .wav file containing both natural noises and babble together. Hence, to meet the requirements of the second study, a new EN was created by employing the exact procedures used originally. First, a sound file that contained only environmental noise was created (ENNB), then a 4-talker babble was added to ENNB, and a new sound file was generated (ENWB). To produce ENRB, the same babble in ENWB was reversed and another sound file was created. The influence of MTB was evaluated using QuickSIN, AzBio, and HINT sentences using the three different ENs (ENNB, ENWB, and ENRB).

Behaviorally, results from the second study showed that there was an influence of MTB on speech perception for the QuickSIN sentences. It is probable that the combined linguistic and non-linguistic information in the ENWB, together with the varying frequency and intensity components of the various environmental noises, is likely what interfered with the intelligibility of the QuickSIN sentences. The current study also found that speech perception scores for the QuickSIN were significantly poorer compared to the AzBio, followed by the HINT under ENWB, followed by ENRB and ENNB. These results are in agreement with the literature that
the QuickSIN is more sensitive than the AzBio and HINT at measuring how well people perceive speech in noisy environments.

Acoustically, the ENNB showed less spectral energy in the low frequencies, whereas the temporal analysis revealed that the ENWB and ENRB had stronger low frequency modulation. There is a possibility that the higher spectral energy of the ENWB was responsible for a greater masking effect than ENNB, leading to poor behavioral performance. The addition of the 4-talker babble to the environmental sounds in the ENNW, on the other hand, may have created the strong low-frequency modulation and interfered with the participants' attention to the rate of sentence presentation.

The EN generated for the second study was a little different than the EN developed in the first study. Spectrally, both the ENs were similar. However, temporally, the original EN showed low frequency modulation with strong modulation at 0.625 Hz. The new EN was also highly modulated at low frequencies, but no peak was observed at 0.625 Hz. A possible explanation for the difference is that the MTB in the newly generated EN was at a lower intensity level than the original EN causing the new EN to be a little softer overall than the original EN. In addition, the bird noise in the new EN was not always heard at the same place as in the original EN, while the MTB was time aligned in both the ENs. This misalignment of bird noise might have filled the temporal gaps in the new EN and possibly reduced the challenging nature of the new EN.

The scores for EN in both the studies reached ceiling for AzBio and HINT at positive SNRs, which limited our ability to discern any other significant differences. At more favorable SNRs, it is likely that the IM effects from the 4-talker babble in the EN started to decline since speech was louder, and at the same time, the EM effect started to increase. As a consequence of
this, the difficulty of the EN may have diminished at the higher SNRs and become similar to the
difficulty of the other noises.

The studies conducted in this dissertation had some limitations. The physical and
affective properties of the sounds incorporated in the EN were not taken into account while
choosing the environmental sounds to include (such as calming, annoying etc.). In the current
investigations, the choice of environmental sounds was not made on the basis of these
characteristics; rather, they were essentially idiosyncratic. In addition, the specific recording
characteristics of the downloaded environmental noises (such as sampling rate, microphone
positioning in relation to the sound source, original recording medium, and storage media) all
had a significant effect in the selection of environmental noise. Yet, it was not possible to obtain
the specific recording settings for the environmental noises that were utilized in these studies.
Further, the scoring technique used in the current studies was percent correct response, which is
limiting in the information it can provide. It is more prone to ceiling and floor effects, and does
not consider other factors that are internal to the listener such as listening effort, reaction time,
etc.

The findings from both studies showed that speech perception scores obtained using both
the ENs (original and new) were poorer than the other noises. The consistent findings in both the
studies presented in this dissertation indicate that the EN used here presented a more challenging
situation than other noises, and speech perception assessment in such an ecologically relevant
auditory scene yielded realistic speech perception scores. Future research should modify the EN
based on the physical and affective characteristics of environmental sounds to make it even more
ecologically relevant and investigate the use of EN with individuals who have different degrees
of hearing loss. In addition, future research should investigate whether the modified EN can also be effective in evaluating hearing aid performance and assist hearing aid fittings to obtain better outcomes. In addition, future research should investigate

Temporal properties of competing noise play an important role in speech perception. If the temporal modulation of the background noise is similar to the speech signal, it causes modulation interference. Whereas when the temporal characteristics of noise are different than the speech signal, it causes a release from masking. Further investigations should also focus on quantifying the spectral and temporal distribution of different noises used in different tests and how those differences affect the perception of the speech stimulus used in those tests. Additionally, further investigations should measure speech perception using a variety of noises that are present in the natural environment that have different temporal properties from the speech signal.
References


https://doi.org/10.1001/archotol.1970.00770040379010


68


