Impact of Speaking Styles on the Accuracy of Predicted Speech Intelligibility

by

Pitchulee Uayporn

B.S. in Communication Science and Disorder, Mahidol University, 2009

Submitted to the Graduate Faculty of the

School of Health and Rehabilitation Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2020

UNIVERSITY OF PITTSBURGH

SCHOOL OF HEALTH AND REHABILITATION SCIENCES

This dissertation was presented

by

Pitchulee Uayporn

It was defended on

November 18, 2020

and approved by

Sheila Pratt, Ph.D., Professor, Department of Communication Science and Disorder

Christopher Brown, Ph.D., Associate Professor, Department of Communication Science and Disorder

Lori L. Holt, Ph.D., Professor, Department of Psychology, Carnegie Mellon University

Dissertation Director: Catherine V. Palmer, Ph.D., Professor, Department of Communication Science and Disorder; Professor, Department of Otolaryngology; Director of Audiology and Hearing Aids, University of Pittsburgh Medical Center (UPMC)

Impact of Speaking Styles on the Accuracy of Predicted Speech Intelligibility

Pitchulee Uayporn, Ph.D.

University of Pittsburgh, 2020

Conversational speech used in research studies is not true conversational speech that individuals use in day-to-day communication. Laboratory-created speech materials are read or memorized and repeated and have different acoustic characteristic compared to true conversational speech. It is of interest to investigate how speaking styles (clear speech, lab conversational speech, and natural conversational speech) impact actual (measured) and predicted speech intelligibility in young adults with normal hearing. Two experiments were conducted in the current study. Speech stimuli were created using the contents of the Story Retelling Procedure (SRP) (Doyle et al., 2000; McNeil et al., 2007) produced by a male talker to create stimuli for each of the three speaking styles. Speech recordings were rated by thirty individuals with normal hearing based on how natural speech sounded. There was a strong, positive correlation between the speech recordings and how listeners perceived the naturalness of speaking styles therefore allowing comparison of materials considered clear speech, lab conversational speech, and natural conversational speech.

Experiment 1 was designed to investigate if there was any significant difference among speech intelligibility for clear, laboratory conversational, and natural conversational speech in five listening conditions (quiet, +3, 0, -3, and -6 dB SNR). The dependent variable (DV) was proportion correct of identified keywords (i.e., speech intelligibility). The results showed that speaking styles and listening conditions impact measured speech intelligibility. Specifically, there were significant differences in the speech intelligibility between lab conversational speech and natural conversational speech. Moreover, the clear speech speaking style can be used to improve listening performance in challenging listening conditions.

Experiment 2 was designed to investigate if the STMI model can accurately predict speech intelligibility for different speaking styles and to evaluate the overall ability of the STMI model to capture speech intelligibility in multi-talker babble noise conditions. The results demonstrated that the current version of the STMI may not be sensitive enough to predict speech intelligibility for different speaking styles when embedded into multi-talker babble.

Table of Contents

List of Tables

List of Figures

Preface

I would like to express my deepest appreciation and sincere gratitude to my mentor Dr. Catherine V. Palmer for guiding me through my PhD journey and for providing help and everything I have ever needed. She is a wonderful person who I always look up to and wish I could be like her someday. Thanks to my dissertation committee, Dr. Sheila Pratt, Dr. Christopher Brown, and Dr. Lori L. Holt for the support and guidance with my work. Special thanks to Dr. Elaine Mormer for her support, mentally and academically, and all of the honest feedback she has given me to improve myself in many aspects. Moreover, I would like to sincerely thank Dr. Susan Shaiman for giving me the opportunity to work as her TA for the Anatomy & Physiology of Speech and Speech Science classes. During my time as the TA, she treated me so well and she was a wonderful and decent boss who was always caring and reasonable. She made my teaching experience more enjoyable and pleasant. Thank you everyone for being such wonderful role models and examples. I also would like to thank the undergraduate-student volunteers - Kamryn James and Gillian Pietrowski - for their assistance in the study.

In addition, I want to specially thank my supportive parents - Mr. Panom and Mrs. Sumalee Uayporn - and my beloved sisters - Pimsuree, Piwsuwaree, and Prawwatcharee - who always encourage and support me to pursue my goal and are always there whenever I need them. Thank you to the Vicuna family in San Antonio, Texas for still opening their home to me and taking me as part of the family. I also would like to thank Jackie Harden, Suchart Suksamran, and Yimwan & Mike O’Brien, my dearest friends who were like sisters and brother to me. Thank you for staying with me during many difficult times, letting me vent, and providing me emotional support, encouragement, and much useful advice. Lastly, thank you to my PhD fellows and lab members for your encouragement and friendship. However, there are many wonderful people who crossed my path during my time at the University of Pittsburgh, who I cannot mention enough; I would like to thank you all for encouraging me and providing me with the motivation to be a better person and complete my dissertation.

I could not imagine myself standing where I am without all of you. You all have played a part in helping me get to where I am today. I could not have accomplished this achievement without any of you. You all are wonderful and amazing human beings who have made my time in the United States very meaningful and memorable. I appreciate everyone’s contribution. Thank you.

1.0 Introduction

Communication is essential to the human condition. We communicate with one another in several different modes such as talking, writing, reading, listening, and gesturing. However, speech is the most common means of communication in humans. Speech is produced by a talker, sound waves are transmitted through the air, and eventually perceived by a listener. Speech is a highly modulated signal across time in frequency regions and as a function of energy. A large amount of informational value is concentrated in relatively few spectro-temporal regions. Speech also is a highly redundant signal allowing communication in complex sound environments. Some spectro-temporal regions survive noise well and, if they can be detected and integrated, often contain enough information to allow successful communication in degraded listening environments (Mattys, Brooks, & Cooke, 2009). There are numerous factors that can affect the transmission of speech such as background noise, room reverberation, hearing loss, distortions in hearing aids or other communication devices, speech material, and speaker. As a result of interrupted transmission of speech; speech audibility, speech intelligibility, and speech quality will be impacted in varying degrees depending on the strength of those factors.

Generally, the amount of sound that is audible is determined from a physiological point of view. The terms speech intelligibility and speech quality often are used interchangeably but, in fact, have different meanings. Speech intelligibility refers to the ability of the listener to identify the word or set of words that were produced. Speech understanding refers to how well speech conveys the meaning to a listener, i.e., the amount of speech items that are recognized correctly. Speech quality refers to the quality of the reproduced speech signal with respect to the amount of audible distortions.

In audiology research and clinical practice, most speech materials used are read-speech material which have been deliberately clearly spoken. These materials are referred to as clear speech. Clear speech typically is produced when a talker believes that a listener has difficulties in perceiving the speech due to the presence of background noise, a hearing loss, or having a different native language. Clear speech is distinct from conversational speech which is produced during everyday communication. In general, clear speech is louder, slower, carefully pronounced, and has longer pauses between words than conversational speech.

Speech materials may consist of single words and in these cases the significance of clearly spoken material versus a conversational presentation will not be particularly relevant. In sentence-length or longer materials, the difference between clear speech and conversational speech will be relevant and performance may impact research findings and clinical recommendations. In the field of Audiology, speech materials were created for diagnostic purposes. The use of clear speech created a controlled condition that could be used to measure relative differences between ears or between time points to identify specific pathological conditions in patients.

The use of these original materials has been extended to use in assessing treatment choices (e.g., signal processing in amplification systems), assessing treatment (e.g., auditory training), and are used in research in the development and assessment of new signal processing to be implemented in amplification systems. In these applications, the face validity of using clear speech comes into question. The treatments being applied or technologies being developed are for use in understanding conversational speech. The rationale that relative differences are still what are of interest may not be supported since the features of conversational speech (lower fundamental formant frequency values, relatively less high frequency energy, shorter pause intervals, faster speaking rates, and co-articulation) may interact with various signal processing strategies differently from clear speech.

There are several studies examining acoustic comparisons of clear and conversational speech samples in many languages. Clear speech generally is slower and more precisely articulated than conversational speech. In the English language, the acoustic information of clear speech is different from that of conversational speech in terms of static cues (spectral, intensity, and temporal cues) and dynamic cues (spectro-intensity, tempo-intensity, and spectro-temporal cues) (Byrd & Tan, 1996; Krause & Braida, 2002, 2004; Picheny, Durlach, & Braida, 1986). Because of these acoustic differences between clear and conversational speech, audibility might be different and, in turn, might impact observed and predicted speech intelligibility. Howell and Kadi-Hanifi (1991) demonstrated that read-speech material cannot represent spontaneous speech acoustically (Howell & Kadi-Hanifi, 1991).

In this document, several topics will be discussed: the characteristics of clear speech, the characteristics of conversational speech, the acoustic differences between them, speech intelligibility prediction indices such as the Articulation Index (AI), Speech Transmission Index (STI), Speech Intelligibility Index (SII), and spectrotemporal modulation index (STMI). This review will generate questions related to the impact of using different speech materials to characterize speech understanding. Specifically, will characteristic differences in speech materials be shown in measured and predicted speech intelligibility?

2.0 Speaking Styles

2.1 Clear Speech

Clear speech is a speaking style that is adopted by a talker to deliberately produce clear speech, and it typically is produced when a talker believes that a listener has difficulties perceiving the speech due to the presence of background noise, a hearing loss, or having a different native language. Read-speech material can be referred to as clear speech; and the majority of speech materials used in audiology research and clinical practice are read-speech. In general, clear speech is slower and more precisely articulated than conversational speech because a talker attempts to reach the optimal position for a speech sound and eliminates the effects of coarticulation where words are linked together and a phoneme starts to possess different characteristics as a result of surrounding phonemes. Therefore, several of the phonological features of a speech sound, especially in sentence length materials, are more precisely produced in clear speech (Ferguson & Kewley-Port, 2007).

For many years, acoustic comparisons of clear speech and conversational speech have been investigated. Clear speech typically has advantages over conversational speech due to a decreased speaking rate, longer and more frequent pauses, an expanded formant frequency range, greater intensity (sound pressure levels), more salient stop releases, increased energy in the 1-3 kHz range of the long-term speech spectrum, increase in modulation depth of low frequency modulations of the intensity envelope, and expanded vowel spaces.

The average intelligibility of clear speech is greater than that of conversational speech across talkers and across groups of listeners including listeners with normal hearing (Ferguson, 2004; Krause & Braida, 2002; Maniwa, Jongman, & Wade, 2008), listeners with hearing loss (Ferguson, 2012; Picheny, Durlach, & Braida, 1985; Schum, 1996; Uchanski, Choi, Braida, Reed, & Durlach, 1996), listeners with simulated hearing loss, and listeners using cochlear-implants (Liu, Del Rio, Bradlow, & Zeng, 2004).

2.2 Conversational Speech

Conversational speech is speech produced during daily communication; therefore, the term conversational speech applies to the type of speech produced under casual or typical circumstances when no special speaking effort or instruction is made (Uchanski, 2005). Conversational speech sometimes is called plain speech because it is the most common and natural verbal communication. Words are produced fluently and are connected together. Coarticulation effects occur frequently due to the short amount of time for articulators to move across so many positions. The term reduced speech sometimes is used to refer to sounds being deleted or produced less clearly than in careful (clear) speech and to speech with syllables or words deleted which mostly occur in casual conversation. Therefore, with this speaking style, speech typically fluctuates within speakers, across speakers, and in different acoustic environments.

2.3 Acoustic Differences Between Clear and Conversational Speech

In this section, the perceptual and acoustic changes that occur when a talker changes from a conversational to a clear speech speaking style will be outlined.

Studies have examined acoustic changes between clear speech and conversational speech not only in the English language, but also in other languages. The focus of this document will primarily be on studies investigating the English language. The intelligibility of clear speech is greater than that of conversational speech in most cases (Payton, Uchanski, & Braida, 1994; Picheny et al., 1985). This finding also is consistent with another study in the Telugu (South Indian) language (Durisala, Prakash, Nambi, & Batra, 2011). The investigators compared characteristics of clear speech and conversational speech in the Telugu and found that clear speech possessed higher fundamental frequency (F0), greater intensity, longer duration, higher consonant-vowel ratio (CVR), and greater temporal energy than conversational speech; and concluded that these acoustic properties contributed to higher intelligibility in clear speech.

The acoustic properties of speech can be sorted into three dimensions: intensity (i.e., amplitude), frequency (i.e., spectral), and time (i.e., temporal). When these dimensions interact with one another, dynamic cues are created: intensity and frequency, spectro-intensity cue; intensity and time, tempo-intensity cue; and frequency and time, spectro-temporal cue.

2.3.1 Intensity Domain

Clear speech is significantly more intense than conversational speech (Picheny et al., 1985, 1986). Picheny and colleagues reported that clear speech usually is produced at a level 5 to 8 dB greater than conversational speech (Picheny et al., 1986). Durisala et al. (2011) also found that the intensity of clear speech is relatively 5.5 dB higher than that of conversational speech. However, in most experiments, the overall RMS levels of clear speech and conversational speech were equated; therefore, the overall level difference cannot be the factor contributing to the higher speech intelligibility in clear speech (Uchanski, 2005).

2.3.2 Spectral Domain

When comparing clear and conversational speech, the spectral domain includes spectral cues resulting from resonances (i.e., formants) of the vocal tract during production of speech that differentiate the two types of speech. Spectral cues help in vowel perception. The vowel perception is dependent on the spectral pattern (i.e., spacing between formants). Spectral cues also are helpful for the perception of fricatives (such as /s/, /z/, /v/, /f/) which usually have more energy in the speech spectrum above 4,000 Hz. Figure 1 shows the spectrogram (the pattern of speech acoustics along three dimensions including time, frequency, and intensity) of /asa/ spoken by a female talker. The frequency range of the /s/ sound is from 5,000 to 10,000 Hz with the frequency energy around 8,000 Hz (darkest area). On the other hand, the surrounding vowel /a/ has the frequency energy below 2,000 Hz.

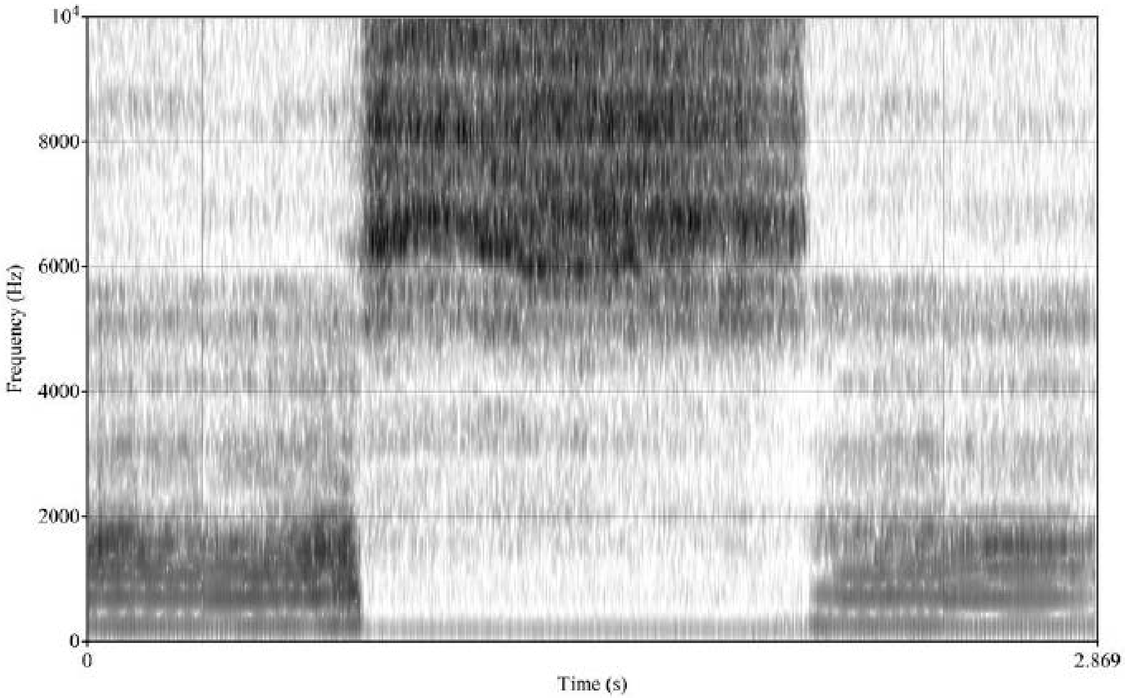


Figure 1 Spectrogram of /asa/

Time is represented on the x-axis, while frequency is represented on the y-axis. Intensity is represented by the darkness of the areas within the spectrogram.

2.3.2.1 Fundamental Frequency (F0) and Formant Frequencies

Fundamental frequency (F0) is a measure of pitch that can be seen in a spectrogram. Bradlow and colleagues (2003) investigated speech intelligibility in sentence-length speech materials of children with learning disability to perceive compared to that of children without learning disability (i.e., control group). Specifically, they tested if listeners could derive substantial benefit from the acoustic-phonetic cue enhancements found in naturally-produced clear speech. In their experiment, they included speaking style (naturally-produced clear speech vs conversational speech) and signal-to-noise ratios (-4 dB vs -8 dB), and talker (male vs female) as variables that varied among participants. They also conducted an acoustic analysis of their stimuli to speculate whether there were distinct characteristics (between the two speaking styles) that might be responsible for speech intelligibility benefits. One of the measures was comparison between the fundamental frequency of clear speech and conversational speech. The result of this analysis was consistent with a previous study by Picheny et al. (1986) that found that clear speech exhibited an increase in mean F0 and F0 range for both talkers relative to conversational speech. For clear speech, the mean F0 was increased by 1.12 and 5.43 semitones, and F0 range was increased by 6.23 and 5.81 semitones, for male and female talkers, respectively (Bradlow, Kraus, & Hayes, 2003). Experimental results also showed that both groups of children received the clear speech benefit (i.e., the children had higher speech intelligibility scores for clear speech than conversational speech), and the benefit became larger when the children listened to a female talker rather than a male talker. It can be summarized that clear speech generally has a relatively higher fundamental frequency (F0) value and a wider range of F0 than conversational speech (in other words, conversational speech has relatively lower F0 values than clear speech). However, it is unlikely that a simple increase in F0 in clear speech completely explains the speech intelligibility advantage over conversational speech. The increase in F0 in clear speech occurs due to an increase in vocal effort that attempts to increase relative intensities of higher frequency components in the speech spectrum (Uchanski, 2005). Clear speech additionally shows expanded vowel formant frequencies (Bradlow et al., 2003; Picheny et al., 1986). Formant frequencies also are different in clear speech compared to conversational speech (Ferguson & Kewley-Port, 2007). Clear speech exhibited an increase in F1 and F2 ranges relative to conversational speech (Bradlow et al., 2003).

2.3.2.2 Long-Term Average Speech Spectrum (LTASS)

Compared to conversational speech, clear speech has higher energy at higher frequencies for the long-term speech spectrum (LTASS) which leads to a decrease in spectral balance (Krause & Braida, 2004). In other words, conversational speech has less relative energy above 1000 Hz. The authors concluded that poorer speech perception in conversational speech was due to this lack of high frequency information inherent in this type of speech production. This finding also is consistent with another study that investigated the acoustic characteristic differences between spontaneous speech and read speech and how they affected speech recognition (Nakamura, Iwano, & Furui, 2008). The researchers used a larger-scale speech corpora consisting of speech with various speaking styles. They found that the spontaneous speech showed a reduction of spectral space compared to that of read speech (meaning that the spontaneous speech has a shorter speech spectrum than the read speech), and they concluded that spectral space is one of the major contributing factors for speech recognition (Nakamura et al., 2008). The importance of the frequency spectrum has been verified by the speech intelligibility index (SII) which is an index that can be used to quantify the relationship between speech audibility and speech intelligibility. Each frequency band contributes different amounts of speech intelligibility.

2.3.3 Temporal Domain

Speech is a temporally distributed signal meaning that the cues to individual phonetic contrasts in speech are distributed in time. Within this domain, clear and conversational speech are drastically different. When speech is spoken in different styles, durational cues also are inherently changed. These cues include silent gaps between phonemes, onset of the following phoneme of words, and transition between phonemes. Temporal cues are helpful for identifying the presence or absence of voicing in consonants (House, 1961) and voiceless fricatives have longer duration than their voiced counterparts (Baum & Blumstein, 1987). These cues help to differentiate the perception between fricatives and affricates (Kluender & Walsh, 1992; Raphael & Dorman, 1980).

2.3.3.1 Speaking Rate

The role of speaking rate has been examined by several investigators. Clear speech is significantly slower in speaking rate and contains lengthened pauses between words while conversational speech ranges between 160-200 words per minute or 3-4 syllables per second which is twice as fast as clear speech (Picheny et al., 1986). This is because there are fewer pauses during conversational speech and the overall articulation rate increases.

Picheny et al. (1986) reported that clear speech has speaking rates of 90 to 100 words per minute (WPM), while conversational speech has speaking rates of 160 to 200 WPM. The authors concluded that the reason why clear speech had a slower speaking rate was because of the increase in duration of pauses and increase in duration of sound segments. In 2003, Bradlow and colleagues analyzed their sentence speech materials to capture if there were any modifications between clear speech and conversational speech that would account for the speech intelligibility differences between clear speech and conversational speech between the group of children with learning disability and the control group. They analyzed the overall speaking rates of clear speech and conversational speech of each talker (male vs female), and they found that both talkers showed significant decreases in speaking rates when clear speech was produced; however, they speculated that the female talker decreased the speaking rate from conversational speech to clear speech far more than male talker (Bradlow et al., 2003).

Picheny and colleagues (1987) extended their previous studies by focusing on the role of speaking rate. They artificially slowed down the rate of conversational speech until its overall duration was equal to that of clear speech. They also compressed clear speech so that its overall duration was equal to that of conversational speech. The results showed that shortening the duration of clear speech decreased the intelligibility of the sentences, whereas expanding the duration of conversational speech did not improve the intelligibility of the sentences. Uchanski et al. (1996) analyzed durational differences between conversational and clear speech, and found differences in phonemic segmental durations between the conversational speech and clear speech. They used a non-uniform time-scaling technique to artificially slow down the conversational speech duration so that its phonemic segmental durations were equal to that of clear speech. They also used the same time scaling technique to compress the clear speech duration so that it was equal in segmental duration to the conversational speech. Their results showed that slowing down the conversational speech resulted in poorer speech intelligibility. Also, speeding up the clear speech resulted in poorer intelligibility than the non-processed conversational speech. It can be concluded that the underlying benefits of clear speech versus conversational speech is independent of rate.

Talkers can be trained to produce a form of clear speech at normal conversational rates (Krause & Braida, 1995). In 2002, Krause and Braida further investigated the role of speaking rate by training talkers with significant public speaking experience to naturally produce clear speech and conversational speech at slow, normal, and quick rates. These talkers were trained to produce nonsense sentence materials. Each sentence consisted of five to eight words, and key words could be noun, verb, and adjective. As a result, they found that clear speech consistently provided an intelligibility advantage over conversational speech regardless of speaking rates (see Table 1). For example, clear speech with normal speaking rate was intelligible at 59 percent key words correct compared to conversational speech at the same speaking rate which was intelligible at 45 percent correct (i.e., clear speech had a 14-point advantage over conversational speech at a normal speaking rate). Clear speech with slow rate (63% correct) also obtained a 12-point advantage over conversational speech with a slow rate (51% correct). At the quick speaking rate, although clear speech had higher intelligibility than the conversational counterpart (46% and 27%, respectively), they reported that the speaking rate for clear speech (218 WPM) was significantly slower than the speaking rate for conversational speech (269 WPM); therefore, there was no statistical advantage of clear speech over conversational speech for quick speaking rate. The findings suggest that acoustical factors other than reduced speaking rate are responsible for the high intelligibility of clear speech (Krause & Braida, 1995, 2002). Additionally, the study demonstrated that clear speech does not need to be slow. From this evidence, we can conclude that speaking rate is not a contributing factor for clear speech's intelligibility advantage over that of conversational speech.

Table 1 Speech Intelligibility of Each Speaking Mode (Krause & Braida, 2002)

Speaking Styles Speaking Rates	Clear Speech	Conversational Speech
Slow	63%	51%
Normal	59%	45%
Quick	46%	27%

2.3.3.2 Pauses

Clear speech typically has more frequent and longer pauses than conversational speech. A pause usually is defined as any silent interval (or gap) between words. There are some studies examining the role of pauses. In Picheny et al.'s (1986) study, they defined a pause as any silent interval between words that was greater than 10 ms and excluded the silent intervals preceding word-initial plosives. In Bradlow et al.'s (2003) study, they defined a pause as any period of silence of at least 5 ms instead of using the 10 ms criterion because the long-duration cutoff would exclude several periods of silence of an intentional pause. Bradlow et al (2003) excluded any silence prior to word initial stop consonants because it was impractical to separate a true pause from the stop closure. Additionally, the investigators calculated the average pause-to-sentence duration ratio for each sentence. Talkers increased the number of pauses, the average pause duration, and the average pause-to-sentence duration ratio in clear speech relative to conversational speech (Bradlow et al., 2003). This finding was also consistent with other studies reporting that the number of occurrences of pauses and the average duration of pauses increased for clear speech (Krause & Braida, 2004; Picheny et al., 1986). We can summarize that clear speech contains lengthened pauses between words while conversational speech contains fewer pauses. As mentioned above in the speaking rates, one possible reason for conversational speech to have higher speaking rates (about 160-200 WPM or 3-4 syllables per second) is because of shorter and fewer pauses (Picheny et al., 1986).

However, the relationship between occurrences of pauses and speech intelligibility is not certain, and differences in pause structure do not necessary account for differences in speech intelligibility (Picheny et al., 1986; Uchanski et al., 1996). Pauses might help to increase speech intelligibility because they provide some additional time for a listener to process the information, but it is not always the case that the speech with longer pauses would be understood better. Therefore, the existence of pauses seems unlikely to be a contributing factor for speech intelligibility differences between clear speech and conversational speech.

Ronnberg et al. proposed that perception of running speech involves simultaneous processing and storage of auditory information. This idea might be supportive of why clear speech has increased advantage over conversational speech because clear speech allows listeners more time for processing clear auditory information. Similar to vision, you likely perceive a clear image better than a blurred one.

2.3.3.3 Vowel Duration

Some acoustic studies showed that all vowels are enhanced (lengthened) when clear speech is produced, i.e., vowel duration increases. The long vowels are lengthened more than their short counterparts. In other words, clear speech enhances the length contrast between long and short vowels rather than just lengthening all vowels by the same amount (Krause & Braida, 2009).

Table 2 summarizes the differences between clear speech and conversational speech for the domains discussed thus far including intensity, spectral characteristics, and temporal characteristics. In the next section combinations of these speech characteristics will be reviewed.

Table 2 Overview of Acoustic Differences between Clear and Conversational Speech

Cues	Clear Speech	Conversational Speech
Intensity	• Produced at 5-8 dB greater	• Lower fundamental formant frequency
Spectral	• Expanded vowel formant frequencies (decrease in spectral balance)	• Less relative energy above 1kHz • Less plosive bursts intensity (especially in word final position)
Temporal	• Slower speaking rate: 90-100 wpm • Phonemic segmental duration increased • Pause durations increased • Greater temporal amplitude modulation	• Faster speaking rate: 160-200 wpm • Faster overall rate of articulation • Fewer pauses, smaller gaps • Boundaries between syllables are not distinct • Shallow depth of modulation

2.3.4 Joint Domains

2.3.4.1 Spectro-Intensity and Spectro-Tempo-Intensity Cues

Temporal Envelope Modulations and Temporal Fine Structures are examples of joint spectro-intensity and spectro-tempo-intensity domains. Temporal features of speech can be divided into three parts: envelope (2-50 Hz), periodicity (50-500 Hz), and temporal fine structure (600 Hz and above). The temporal envelope's relative amplitude and duration are cues and translate to manner of articulation, voicing, vowel identity and prosody of speech. Periodicity provides information about whether the signal is primarily periodic or aperiodic, e.g., whether the signal is a nasal or a stop phoneme. Temporal fine structure (TFS) is the small variation that occurs between periods of a periodic signal or for short periods in an aperiodic sound and contains information useful to sound identification such as vowel formants. Both temporal envelope and temporal fine structure cues are useful for speech intelligibility (Rosen, 1992).

Krause & Braida (2004) investigated the effect of temporal modulation between speaking styles (clear speech vs conversational speech) with various speaking rates (normal, slow, and fast speaking rates). They found that the temporal envelope of clear speech with a slow speaking rate has a higher peak in the 1-3 Hz region than that of conversational speech with a normal speaking rate, but the increase in modulation is not associated with the change in speaking rate as it showed that both of the envelopes of clear speech with normal and slow speaking rates were similar in how they differed from that of conversational speech with normal speaking rate (Krause & Braida, 2004). When comparing the envelopes between two speaking styles with the normal speaking rate, there was an increase in the modulation's depth for low modulation frequencies which might contribute to speech intelligibility in clear speech (Krause & Braida, 2004). The greater the depth, the greater the speech intelligibility.

Liu and colleagues (2004) also conducted a study investigating global acoustic properties of clear speech and conversational speech and the relationships between speech intelligibility and those properties. Temporal envelope and temporal fine structure were properties that were examined. The results indicated that temporal envelope carries acoustic cues that contribute to increased clear speech intelligibility, while temporal fine structure is important for speech recognition in noise for both speaking styles (Liu et al., 2004).

There have been additional studies confirming that the most prominent cues affecting the speech understanding differences between clear speech and conversational speech are temporal envelope (in high SNR) and temporal fine structure (in low SNR). These two parameters contribute to the advantage of clear speech over conversational speech (Liu & Zeng, 2006; Lorenzi, Gilbert, Carn, Garnier, & Brian, 2006).

2.3.4.2 Spectro-Temporal Cues

Because speech is spontaneous and dynamic, the stream of speech sounds varies dynamically, and each acoustic feature changes simultaneously. There is a reason to believe that not only one acoustic cue is responsible for the clear speech intelligibility advantage. There are some experiments showing that the combinations of multiple features actually are responsible for the speech intelligibility benefit of clear speech over that of conversational speech.

Kain and colleagues (2008) investigated the relationship between combinations of acoustic characteristics and speech intelligibility of clear speech and conversational speech. They introduced the new approach by combining some features of clear speech and conversational speech. They extracted clear speech features and replaced them into the conversational speech material, therefore, creating a new kind of speech material called Hybrid. For hybrid speech, there were two conditions: HYB-DSP, spectral properties, phoneme sequences, and duration were manipulated; and HYB-EFN, speech energy, fundamental frequency (F0), and non-speech features (e.g., pauses) were manipulated. In the analysis, they compared speech intelligibility of clear speech, conversational speech, and hybrid speech. Figure 2 shows the results of the study demonstrating that the intelligibility of HYB-DSP speech is significantly higher than baseline conversational speech, but there is no significant difference between HYB-EFN and baseline conversational speech. Consistent with other literature, clear speech has higher speech intelligibility than baseline conversational speech. They further examined the intelligibility of hybrid speech with other possible sets of clear speech's acoustic features: HYB-D, uses only phoneme durations from clear speech; and HYB-SP, uses a combination of spectral features and phoneme sequences from clear speech, Figure 3 shows that the combination of short term spectrum and duration (HYB-DSP) yielded higher speech intelligibility than other hybrid conditions that replaced only spectrum, and only duration (Kain, Amano-Kusumoto, & Hosom, 2008). The speech intelligibility of this combination also approximately reached the same level of correctness to that of clear speech. We can conclude that combinations of spectral and temporal cues are responsible for increased speech intelligibility when comparing clear speech and conversational speech.

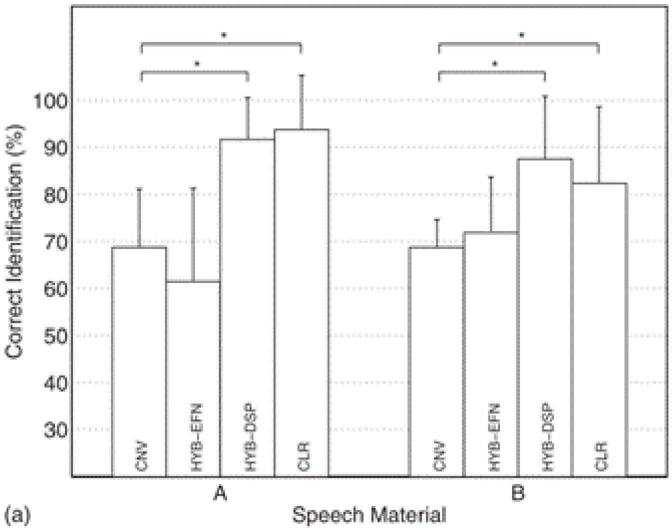


Figure 2 Results from Kain et al., 2008 showing Speech Intelligibility Scores for different conditions: baseline-conversational (CNV), HYB-EFN, HYB-DSP, and clear speech. The significant different results were marked with an asterisk with permission from AIP Publishing.

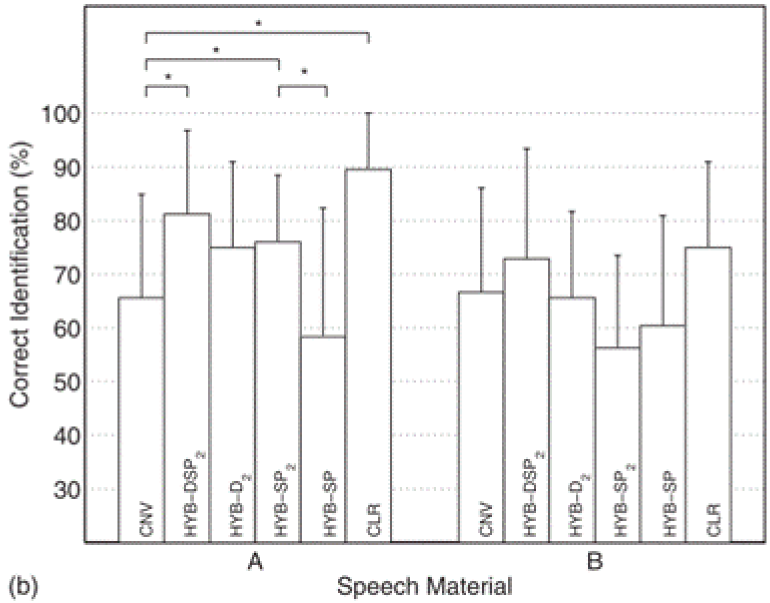


Figure 3 Results from Kain et al., 2008 showing Speech Intelligibility Scores using combination of features baseline-conversational (CNV), HYB-DSP HYB-D, HYB-SP, and clear speech. The significant different results were marked with an asterisk with permission from AIP Publishing.

Recently, a study looked at an entire formant contour (i.e., the pattern of change in formant frequencies across a single word or sentence). The researchers modified the formant contour (preserving duration cues of clear speech) of conversational speech to match it with that of clear speech and found that this manipulation increased speech intelligibility of the modified conversational speech (Amano-Kusumoto, Hosom, Kain, & Aronoff, 2014). They conducted a study examining the acoustic features contributing to speech intelligibility and improving speech intelligibility of conversational speech by approximating clear speech features. They combined acoustic features of clear speech and conversational speech by using a hybridization algorithm, and the results showed that there are significant improvements over conversational speech stimuli of about 11-23% in sentence-level stimuli.

From this evidence, there is a reason to believe that not only spectral cues alone or temporal cues alone are responsible for the clear speech intelligibility advantage, but the combination of these acoustic cues is important for an increase in speech intelligibility of clear speech over that of conversational speech. Table 3 provides a summary of the speech cues discussed in this section and highlights their importance to speech intelligibility.

Table 3 Importance of Acoutic Cues

Domain	**Importance**	**Features**
Intensity	• Perception of vowels • Perception of fricatives	• Amount of energy e.g., plosive burst
Spectral	• Perception of manner of phonemes e.g., voiced and nasality	• Fundamental formant frequencies • Speech energy • Vowel space
Temporal	• Perceptual distinction between affricates and fricatives	• Speaking rates • Numbers and durations of pauses • Articulation rates
Spectro-Intensity	• Identification of voiceless stop consonants	
Spectro-temporal	• Perception of place of articulation	• Formant transitions • Transition durations • Voice onset time
Tempo-intensity	• Identification of stop consonant	• Envelope of amplitude fluctuation • Depth of modulation frequency

Table 3 Importance of Acoutic Cues (continued)

Domain	Importance	Features
Spectro-tempo-intensity	• Ability to understand speech in background noise	• Temporal fine structures

The impaired auditory system has some degree of difficulty using the rapidly changing formant transitions as a cue to speech perception (Zeng & Turner, 1990; Turner, Smith, Aldridge, & Steward, 1997). Normal hearing listeners mostly rely on the formant transition cue to identify a target phoneme. However, there is a study showing that listeners with hearing impairment do not perform differently from listeners with normal hearing when using formant transitions in quiet conditions (Hedrick & Younger, 2007).

Besides formant contours (i.e., formant transitions), voice onset time (VOT) is one of the spectro-temporal cues. VOT is the duration of the of time between the release of a plosive (i.e., a stop consonant) and the beginning of voicing or vocal fold vibration. From Picheny et al.'s (1986) study, conversational speech showed shorter VOT compared to clear speech.

Spectro-temporal cues are useful for both individuals with normal hearing and with hearing loss to understand speech and are salient features for the clear speech intelligibility advantage over conversational speech. Models predicting speech intelligibility that take spectro-temporal properties into account will be reviewed in light of the potential importance of this feature in differentiating clear and conversational speech.

The spectro-tempo-intensity cue (temporal fine structure) is a cue that affects a listener's ability to understand speech in background noise for individuals with normal hearing. The temporal fine structure cue helps listeners maintain speech intelligibility in noisy situations;

however, it appears that listeners with hearing loss are unable to use this temporal fine structure cue to perceive speech in background noise (Qin & Oxenham, 2003; Lorenzi et al., 2006). Therefore, this cue might not be of interest to investigate since this cue is not useful for individuals with hearing loss who are a population of future interest in this area of work.

3.0 Prediction of Speech Intelligibility

The auditory periphery is composed of the outer ear and middle ear which pre-filter acoustic stimuli and attenuate some of the stimuli, then the stimuli is sent to the inner ear where the basilar membrane is situated. There are hair cells along the basilar membrane that are tuned by frequencies (i.e., tonotopic organization). When these hair cells vibrate, they cause an electro-chemical potential difference that innervates an impulse firing an electrostatic signal along the auditory nerve fiber. This spectral-temporal representation of the acoustic stimuli is presented to the central nervous system and is transmitted to the brain which allows us to hear and understand the stimuli.

Several computational models have been created to estimate how we would hear any speech signal and how much we would understand the signal through any signal processing techniques under various listening conditions. These computational models also allow us to test speech-algorithms for hearing aid signal processing development.

Testing for speech intelligibility can be a time-consuming process and uses a lot of resources especially for diagnosis of hearing impairment. In order to make testing more efficient, the audibility index was created several decades ago. In general, we use the term audibility index for any index defining speech intelligibility based on importance-weighted measurements of the audible speech spectrum (i.e., determining the amount of sound available to a listener). The most renowned indices are the Articulation Index (AI) and its successor, the Speech Intelligibility Index (SII).

There are several underlying factors contributing to speech understanding that need to be incorporated into a speech intelligibility predictor's algorithm. Speech styles are one of them. As discussed previously, the listeners' speech intelligibility performance is significantly reduced for conversational speech compared to clear speech. This is due to the fact that conversational speech is significantly different from clear speech acoustically. Therefore, it is important for researchers to carefully select speech materials to be used in testing that can closely represent speech in the daily communication that individuals encounter if the goal is to estimate how the individual will perform in real-world conditions. The majority of speech perception studies use clearly spoken words, read sentences or passages to be able to control for the acoustic characteristics of the stimuli. However, they do not fully represent speech that individuals encounter in their daily communication. In addition, several models that predict or estimate speech intelligibility are developed using clear speech stimuli. Therefore, it is questionable whether those models would be capable of providing an accurate prediction when conversational speech is of interest. In this document, we will discuss some of the main and widely used models. Some of them might be able to account for some of the acoustic differences between clear and conversational speech and be able to predict speech intelligibility of conversational speech accurately. We will examine whether some of the salient acoustic properties that differentiate the speech intelligibility of clear and conversational speech would be captured in predicted speech intelligibility generated by these models.

3.1 Articulation Index (AI)

The Articulation Index is the best-known index for estimating speech intelligibility. It was originally invented by Fletcher in the Bell Laboratories in the 1920s, and it was developed for about thirty years until the Articulation Index was published by Fletcher and Galt in 1950. When

Fletcher retired, research focused on the Articulation Index was discontinued. As a consequence, the Fletcher and Galt version of the Articulation Index never was used in real practice. The first Articulation Index that was used in practice was the American National Standards Institute 1969 version (ANSI, 1969) which was a simpler version of the Articulation Index calculation used in communication research during World War II. In 1947, French and Steinberg published a review of speech intelligibility research and its potential problems, and they identified that the Audibility Index could be used to predict speech intelligibility (French & Steinberg, 1947).

The ANSI S3.5-1969 version of the Articulation Index (ANSI, 1969) calculates the effectiveness of the speech communication channel by providing an index between 0 and 1. The frequency range is between 150 and 8000 Hz, and it is divided into twenty bands whose frequency limits are chosen based on the importance of that frequency to the long-term average speech spectrum (LTASS) (French & Steinberg, 1947). The width of each frequency band is adjusted to make the bands equal in importance. These adjustments were made on the basis of intelligibility tests with low-pass and high-pass filtered speech, which revealed a maximum contribution from the frequency region around 2500 Hz. Auditory masking also is accounted for within each octave band within the Articulation Index. For the Articulation Index, using two key assumptions: 1) each frequency band contributes independently and 2) the contribution of each band is dependent on the effective signal-to-noise ratio (SNR) within that band. Therefore, the speech and noise signals must be defined to obtain an accurate calculation. When conditions are optimal, each frequency band would equally contribute 5% to the Articulation Index’s result of 1 (which means the signal is fully audible to a listener). On the other hand, when conditions are not optimal, only part of the speech signal would be transmitted in each frequency band resulting in the Articulation Index’s result of less than 1 (a less audible signal).

An Articulation Index of 1 is not required for 100% understanding (Killion, Mueller, Pavlovic, & Humes, 1993). To convert the Articulation Index into a predicted speech intelligibility (i.e., speech understanding), an articulation-to-intelligibility transfer function must be applied. The transfer function assumes that the predicted speech intelligibility depends on the proportion of time the speech spectrum exceeds the audibility threshold or the noise. The Articulation Index can provide accurate predictions of average speech intelligibility over a wide range of conditions including broadband noises (Egan & Weiner, 1949; Miller, 1947), high- and low-pass filtering (Fletcher & Galt, 1950) and distortions of the communication (Beranek, 1947). It also has been used to model the loss of speech intelligibility resulting from sensorineural hearing impairments (Fletcher, 1952, 1953; Humes, Dirks, Bell, Ahlstrom, & Kincaid, 1986; Ludvigsen, 1987). However, several studies have found that the Articulation Index overestimates the performance of individuals with hearing loss (Egan & Weiner, 1949; Fletcher & Galt, 1950; Hulsch, 1975). In addition, the Articulation Index is less effective under nonlinear distortions and reverberation conditions.

3.2 Speech Intelligibility Index (SII)

The Speech Intelligibility Index (SII) was created to overcome some of the difficulties found in the Articulation Index. Compared to the Articulation Index, the SII provides a more general framework which allows users to flexibly define some basic input variables such as speech signal level, background noise level, the listener's auditory threshold, and the reference point for the measurement (i.e., at the ear drum level or free-field level). The SII also corrects for upward spread of masking and high presentation levels. Unlike the Articulation Index that used 20 bands of differing sizes, the SII provides some options for frequency bands with equal sizes to use in its calculation (i.e., octave bandwidths, equal bandwidths, 1/3 octave bandwidths, and critical bandwidths). The SII also provides frequency importance functions (FIFs) which are used to determine the contribution to speech intelligibility of each frequency band.

The American National Standard's Methods for the Calculation of the Speech Intelligibility Index (ANSI, 1997) is a mathematical method of quantifying loss of speech audibility and it has been used to predict the speech recognition performance of individuals with normal hearing and hearing loss. It has been developed over time and its complexity has grown to improve accuracy in its prediction. The SII is currently the most widely used audibility index in both clinical audiology and hearing research. It provides a numerical expression of the audibility of an average speech signal based on the intensity of the speech signal of interest, the listener's hearing thresholds, and noise levels. The numerical value ranges from 0 to 1 representing the proportion of speech information available to a listener. The value of 0 indicates that none of the speech information is audible to the listener, while the value of 1 indicates speech information is fully audible to the listener. The general SII formula is:

$$SII = \sum_{i=1}^{n} I_i A_i$$

Where I_i is the band-importance functions of the speech material of interest, A_i is the band audibility, and n is the number of frequency bands used in summation. In order to create a SII value, the band audibility coefficient and band-importance function are multiplied together for each frequency band, and then they are summed up across the total number of frequency bands used in the computational procedure (ANSI, 1997). There are several factors affecting the calculated SII value such as band-importance functions, hearing thresholds, noise levels, computational bands, types of stimuli, etc. The factor with the most impact on the calculation is the frequency-importance functions. They are determined by the amount of degradation in speech recognition or speech understanding that occurs when the target band is filtered out.

The SII provides a way of quantifying the audibility of speech and its effect on intelligibility (ANSI, 1997). The SII is based on the assumption that speech intelligibility can be predicted from the amount of long-term-average speech spectra of the speech and background noise reaching the ear that is above the hearing threshold of the listener (Ching et al., 2013). Since the result of the SII calculation is the audibility, it cannot directly estimate or predict the average speech intelligibility. Therefore, another conversion is needed. A transfer function is used to translate audibility into predicted speech intelligibility. The shape of the appropriate transfer function (Sherbecoe & Studebaker, 2002; Studebaker & Sherbecoe, 1991, 1993; Studebaker, Sherbecoe, McDaniel, & Gwaltney, 1999) depends on the speech material (e.g., word level, sentence level). Several researchers have used the predicted values of the SII to determine the impact of audibility on a speech signal to the resulting speech intelligibility.

Because the SII procedure is a simplified model of the auditory periphery, the procedure can be extended to include some conditions that are not accounted for in the standard procedure. Rhebergen and Versfeld (2004) proposed an extension of the SII called extended speech intelligibility index (ESII) to account for any fluctuations in the masking noise (Rhebergen & Versfeld, 2004) by dividing the SII calculation into short time frames in order to account for fluctuating noise. However, the ESII requires access to the target speech and the interfering noise separately and cannot be used in cases where the mixture is degraded or enhanced by some type of signal processing algorithm. This modification slightly increased the predicted values of the SII. This adapted version of the SII has not been used by other researchers.

Another SII modification was introduced by Kates and Arehart (2005) to improve the accuracy of estimating speech intelligibility under conditions of additive noise and peak-clipping and center clipping distortion. They believed that these conditions could affect speech intelligibility performance of individuals with hearing loss using hearing aids, and there is no metric that could successfully capture the change in speech intelligibility when these conditions take place especially in hearing aid or other communication systems (Kates & Arehart, 2005). Their goal was to accurately predict speech intelligibility under the effects of broadband noise and nonlinear distortion reproduced by hearing aids and other communication systems. They also validated the procedure in groups of individuals with normal hearing and hearing loss. They found that the most effective procedure was to divide the speech signal into three amplitude-level regions: low (between 10 and 30 below the overall RMS), mid (between 0 and 10 dB below the overall RMS), and high (above the overall RMS); then compute the coherence SII separately for the signal segments in each region, and then estimate speech intelligibility from a weighted combination of the three coherence SII values.

3.3 Speech Transmission Index (STI)

Steeneken and Houtgast (1980) proposed the speech transmission index (STI) which is a physical method for evaluating the quality of speech-transmission channels. The STI value ranges from 0 to 1 (bad to excellent). The STI value of 1 indicates that a speech-transmission channel carries out the speech perfectly, i.e., the intact speech remains perfectly intelligible; while the value of 0 indicates that the speech information is completely lost when transferred through a channel. Similar to the SII, the STI is a monaural model that is based on the SNR in a number of frequency bands. For the STI, the SNR in each band is related to the reduction of amplitude modulations caused by the transmission system. The reduction of modulations is determined by the decrease of the modulation index of sinusoidally modulated noise signals in different modulation frequency bands, divided into octave bands. Additionally, STI has transmission index values, similar to frequency-importance functions in SII, for weighting the contribution of each individual band to the STI value. However, there are some differences between the STI and the SII. The signal used in the STI is a speech-like signal (i.e., amplitude-modulated speech-shaped noise) rather than a real speech signal used in SII. The concept of the STI is that the speech intelligibility is related to the preservation of the spectral differences between the consecutive phonemes that is the temporal envelope of the speech. This means a decrease in speech intelligibility is associated with the reduction in the modulation depth of the temporal envelope. However, the STI calculation is simpler than that of SII because the SII accounts for upward spread of masking and hearing acuity (van Wingaarden & Drullman, 2008). Basically, to derive the STI value, the modulation depth of the signal in each frequency band is measured, multiplied by the transmission index value of each frequency band, and then the products are summed across frequency. The STI is highly correlated with speech intelligibility scores when the environment is degraded by noise, reverberation, and hearing loss.

3.4 Speech Intelligibility Prediction Based on Mutual Information (SIMI)

The speech intelligibility prediction based on mutual information (SIMI) (Jensen & Taal, 2014) is a monaural speech intelligibility prediction approach that is based on mutual information between critical-band amplitude envelopes of a clean signal and a noisy/processed signal. This method also predicts speech in noise when noise is not necessarily stationary. The authors expected that if the mutual information is zero, then the predicted intelligibility of the noisy signal will be zero, i.e., none of the noisy signal can be understood. On the other hand, if the noisy signal envelopes provide some information about the clean signal (i.e., mutual information is positive), the intelligibility of the noisy/processed signal would be at some level dependent on the amount of mutual information. The relationship between the mutual information and the intelligibility of the noisy signal was shown to be positively strong. However, there are some concerns about the model. First, the model compares amplitude envelopes between the clean signal and noisy signal. In reality, we do not have access to the same information in both conditions simultaneously, so this method is questionable. Second, the model developers aimed at simplicity by not mentioning the impact of band- importance functions. They assumed that each spectral band contributes equally. There are several studies that have addressed importance of frequency bands and they showed that each band contributed to speech understanding differently (ANSI, 1997; McCreery & Stelmachowicz, 2011; Pavlovic, 1994; Ricketts, Henry, & Hornsby, 2005; Steeneken & Houtgast, 1999). This model has not been used in investigations outside of the investigator's laboratory or in real world application.

3.5 Spectro-Temporal Modulation Index (STMI)

Although the Articulation Index (AI), Speech Intelligibility Index (SII), and Speech Transmission Index (STI) are the primary speech intelligibility prediction models that have been used widely, there is another method proposed for calculating speech intelligibility: the Spectro-Temporal Modulation Index (STMI) (Chi, Gao, Guyton, Ru, & Shamma, 1999; Elhilali, Chi, &

Shamma, 2003). The STMI is a physiologically motivated model of auditory processing that is capable of measuring speech intelligibility and effects of noise, reverberation, and other distortions by assessing the integrity of both spectral and temporal modulations in a speech signal. Generally speaking, the STMI measures the changes in the auditory model output. Like the STI, the STMI has specific weighting functions for the speech spectrum. However, the STMI elaborates the STI in that it explicitly incorporates the joint spectro-temporal modulations of the speech signal into the calculation. The STMI can be derived directly from a speech sample by quantifying the difference between the spectro-temporal modulation content of the clean and noisy speech signals. First, the clean speech is analyzed, and its 4-D output is averaged over the stimulus duration to generate the 3-D template of the speech token. The noisy speech signal is then analyzed in the same way, and the outputs of both speech signals are subtracted (Figure 4). The prediction of STMI was validated by comparing to actual speech intelligibility performance of human subjects (Elhilali et al., 2003). As seen in their results, there is a good relationship between the STMI and the intelligibility scores (Figure 5). Thus, they proposed that the STMI is a method that can accurately measure speech intelligibility, especially in reverberant and noisy conditions.

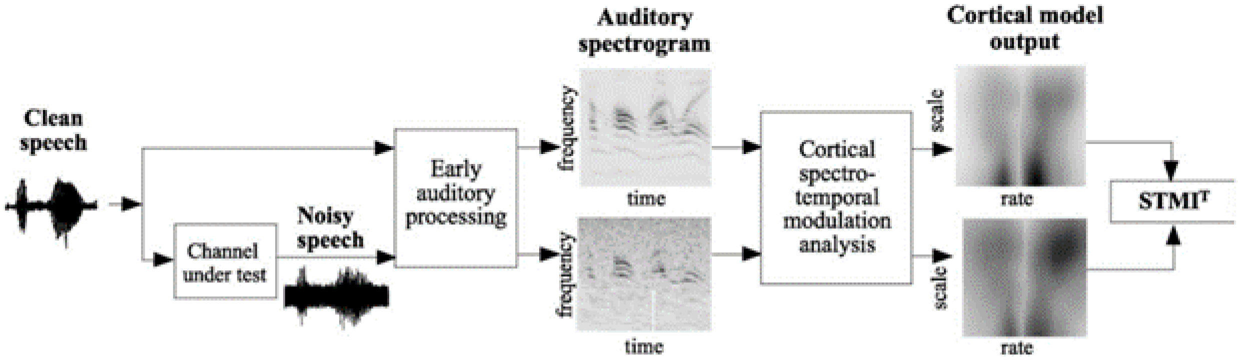


Figure 4 Schematic of the STMI Computation from a Speech Sample (Elhilali et al., 2003) with Permission from Elsevier.

Figure 5 Relationship between STMI and Speech Intelligibility Scores (Elhilali et al., 2003) with Permission from Elsevier.

Although the STMI is not widely used, it is a model based on physiological findings in the primary auditory cortex and on psychoacoustical measurements of human sensitivity to spectral and temporal modulations. The model is validated showing that its prediction matches with the actual speech intelligibility performance. The model is sensitive to the joint modulations between spectral and temporal meaning and accounts for changes in the spectro-temporal content of the speech signals which are salient features contributing to speech intelligibility differences between clear speech and conversational speech although the model was not specifically developed for comparing speech styles. This model might be able to capture the differences in spectro-temporal features between clear speech and conversational speech, and therefore precisely predict the expected speech intelligibility of conversational speech, as well as that of clear speech.

Table 4 provides a summary of the speech intelligibility models reviewed in this section. The type of measure and factors that are included in the measure are highlighted.

Table 4 Comparison between Speech Intelligibility Predictive Models

Models	**Type of measure**	**Factors including in the measure**
Articulation Index (AI)	Static measure	• Effective SNR within a number of bands • each band equally contributes 5% 20 equally spaced bands
Speech Intelligibility Index (SII)	Static measure	• Extend from AI • Various frequency band spacing • Assign weights to each band • Account for masking • Need transfer function to transform audibility to predicted intelligibility
Speech Transmission Index (STI)	Temporal measure	• Indirect measure • Predict speech intelligibility loss due to channel effects
Spectro-Temporal Modulation Index (STMI)	Measure that accounts for physiological effects of the auditory periphery	• Employs auditory model to allow the analysis of joint spectro-temporal modulations in speech to assess the effect of noise, reverberations, and other distortions • Sensitive to non-linear distortion and still works when multiple distortions occur

4.0 Summary and Statement of the Problem

Speech materials used in the audiology clinic and the speech individuals encounter in their daily life are substantially different. It is of interest to examine the relationship between speaking style and predictions of speech intelligibility. The impact of speaking styles on predicting speech intelligibility performance is not well understood. As demonstrated in the literature, there are speech intelligibility performance differences between listening to clear speech and conversational speech. Currently, there are many researchers attempting to identify the salient features of clear speech that contribute to the speech intelligibility advantage over that of conversational speech. The literature points to spectro-temporal features (i.e., formant transition, duration of the formant transition, and voice onset time) of speech as one feature that might be responsible for the clear speech intelligibility advantage over conversational speech. Although the Articulation Index (AI), Speech Intelligibility Index (SII), and Speech Transmission Index (STI) are the primary speech intelligibility prediction models that have been used widely and validated, they only incorporate some of the static cues of speech with inclusion of either spectral or temporal features. They do not account for the dynamic cues like spectro-temporal cues which may be the salient contributing factors for speech intelligibility when comparing conversational and clear speech. In order to investigate the impact of speaking styles comprised of the same speech stimuli on predicting speech intelligibility, the features of the Spectro-temporal Modulation Index (STMI) make it a potentially more preferred model to accurately predict the speech intelligibility of both clear speech and conversational speech. It would be of interest to test the prediction of this model as compared to measured speech intelligibility performance of clear and conversational speech in that the model is designed to account for the feature (e.g., spectro-temporal cues) that investigators have identified as a salient difference between these speech materials.

5.0 Research Question, Specific Aims, and Hypotheses

Clarity of speech signals plays an important role in determining speech intelligibility, especially in noise. The clarity of speech signals can vary among speaking styles e.g., from clearly spoken (hyper-articulated) to conversationally spoken (hypo-articulated). Speaking styles can affect the perception of words and with different context the target word can be perceived differently (Vitela, Warner, & Lotto, 2013). There is flap reduction in conversational speech or reduced speech (Warner, 2005), meaning that the occurrence of brief tapping on the alveolar ridge with the tongue is minimized when conversational (or reduced) speech is produced. Conversational speech can be thought of as a form of distorted speech given that there are omissions and deletions in many sounds. It would be important to be able to accurately estimate the speech intelligibility from a true representation of speech signals that individuals encounter in daily communication to have good face validity of hearing assessment and treatment outcome assessments.

This proposed study aims to investigate the impact of speaking styles (clear speech, natural conversational speech, and conversational speech in lab setting) on the actual performance (i.e., measured speech intelligibility outcomes) and the predicted performance (i.e., the predicted speech intelligibility derived from a selected predictive model, i.e., STMI). This investigation also will examine the accuracy of the prediction of speech intelligibility by examining the correlation between actual performance and predicted performance. This research design will allow capturing the differences in speech intelligibility among these speaking styles and evaluation of the STMI model that incorporates spectro-temporal modulations in its ability to capture the differences in speech intelligibility among these speaking styles.

Research Question: How do various speaking styles (i.e., clear speech, natural conversational speech, and conversational speech in lab setting) impact the accuracy of the prediction of speech intelligibility?

Specific Aim 1: To determine if there is any significant difference among speech intelligibility of clear speech, laboratory conversational speech, and natural conversational speech.

Null Hypothesis for Specific Aim 1: There is no significant difference among speech intelligibility of those speech stimuli.

Alternative Hypothesis for Specific Aim 1: Speech intelligibility of clear speech is greater than that of both laboratory and natural conversational speech; however, the speech intelligibility of natural conversational speech would be less than that of laboratory conversational speech.

Specific Aim 2: To determine if a selected speech intelligibility model, STMI model, can accurately predict speech intelligibility of various speaking styles (i.e., clear speech, natural conversational speech, and conversational speech in lab setting).

Null Hypothesis for Specific Aim 2: There is no relationship between the predicted speech intelligibility outcomes and the actual speech intelligibility performance. The STMI model does not have the ability to predict speech intelligibility correctly for each speaking styles.

Alternative Hypothesis for Specific Aim 2: There is a positive relationship between the predicted speech intelligibility outcomes and the actual speech intelligibility performance. The STMI model has the ability to predict precise speech intelligibility for each speaking styles. Therefore, there will be differences in the predicted speech intelligibility obtained from the STMI model.

Significance: The findings of the proposed study will assist in better treatment recommendations for individuals with hearing loss when considering speech intelligibility within a framework of conversational speech. Since evidence-based prescription formulas for hearing aid fittings are based on clear speech intelligibility and ignore the changes caused by true conversational speech, implementing true conversational speech in a predictive model could help to improve the quality of care of individuals using amplification devices. The proposed study will investigate the accuracy of prediction of speech intelligibility among speaking styles: clear speech (slowly read speech which is a typical type of speech material in audiological assessment), laboratory conversational speech (read speech or sometimes memorized speech that is spoken in a conversational/casual manner by a talker), and natural conversational speech (spontaneous speech that is induced by asking a talker to retell stories to his friends).

6.0 Research Design and Methods

In order to answer the research question, speech materials selection was carefully considered to ensure that the materials represented the real-world listening environment so that outcomes could be generalized to everyday communication and would exhibit performance that is expected to be seen in real life. The study was designed to compare the measured speech intelligibility of clear speech, laboratory conversational speech, and natural conversational speech in various listening conditions obtained from adults with normal hearing. The measured speech intelligibility and the predicted speech intelligibility obtained from the STMI predictive model which is designed to be sensitive to spectro-temporal modulation of speech was compared. Spectro-temporal modulations are believed to contribute to the intelligibility of speech and may differentiate between types of speaking style. To answer the research question, a cross-sectional, within-subject research design with multiple subjects was used.

6.1.1 Research Participants

Power analysis was calculated via G-Power using 3x5 two-factor within-subject repeated measure with alpha of 0.05. A small effect size (0.2) and medium effect size (0.5) were used in the calculation, resulting in a required sample size of 24 and 15 subjects, respectively in order to achieve 95% statistical power. However, data from 36 participants were collected for the current study to ensure that we had sufficient data for the analysis.

Because of the COVID-19 pandemic and the suspension of all in-person research, some of the inclusion and exclusion criteria were modified to make remote data collection possible. For instance, the hearing test was completed using hearing test apps such as uHear (for iOS) and Hearing Test (for Android). This was considered acceptable given that the goal was not to establish specific hearing thresholds, but rather to rule out significant hearing loss. Potential participants were asked to upload their hearing test results when completing the questionnaire via Qualtrics. The link to the questionnaire was available to pre-screen individuals prior to the consent process to ensure that individuals had the necessary computer hardware and system (such as a camera, a microphone, and a pair of headphones) and were in good health required for the research. Inclusion criteria are described in Table 5.

Table 5 Inclusion Criteria

In-person Testing Criteria	**Remote testing Criteria**
- Age between 18 to 35 years old	- Age between 18 to 35 years old
- Native American English speakers	- Native American English speakers
- Conventional pure tone air-conduction hearing thresholds within normal limits for both ears o Testing from 250 – 8000 Hz including 3000 and 6000 Hz o Hearing threshold ≤ 25 dB HL at all testing frequencies	- Hearing's result from an application shows within normal hearing range for both ears o uHear (iOS): from 1000 – 6000 Hz, level could not be specified. (See more detail below) o Hearing Test (Android): from 1000-8000 Hz, hearing level ≤ 25 dB HL.
- Normal, type A tympanogram	- N/A due to inaccessibility

Table 5 Inclusion Criteria (continued)

In-person Testing Criteria	Remote testing Criteria
- Word recognition score at 40 dB SL within the 95% confidence interval according their PTA (the average of thresholds at 500, 1000, and 2000 Hz) on the Northwestern University Auditory Test No. 6 (NU-6) determined by the SPRINT Chart for 25-word lists (See Appendix C).	- NU-6 Word Recognition score is 90% or higher when testing at comfortable level. - 95% CI cannot be determined due to lack of hearing level at 500 Hz and unknown specific hearing level when uHear app is used.
- The Montreal Cognitive Assessment (MoCA) score 26 or higher (out of 30)	- MoCA score is 26 or above (out of 30)

The uHear app does not include the details of hearing level for each testing frequency when hearing test results were generated. However the app classified the test results into six categories: normal hearing (up to 25 dB HL), mild hearing loss (26-40 dB HL), moderate hearing loss (41-55 dB HL), moderately severe hearing loss (56-70 dB HL), severe hearing loss (71-90 dB HL), and profound hearing loss (greater than 90 dB HL). The testing frequencies provided in the app are 500, 1000, 2000, 4000, and 6000 Hz. Therefore, the hearing level at 3000 and 8000 Hz could not be tested via the app.

Although both hearing apps, the uHear and the Hearing Test, can provide test result at 500 Hz, the author decided to exclude the test result at 500 Hz and below due to the fact that low frequency hearing thresholds are often compromised by the participant’s listening environment, headphones, and how headphones are worn.

Exclusion criteria for both testing protocols included:

- Have recent middle ear problems (within the last 3 months of the test date) and/or still in doctor's care
- Have constant ringing or buzzing in his/her ear(s)
- Have ear surgery within the last 3 months.
- Have been diagnosed with any neurological condition or psychological disorder and still in treatment
- Have been diagnosed with motor speech disorder

Exclusion criteria were added when remote protocols were administered to ensure that participants would be able to complete the experiment. The added exclusion criteria included:

- Uncomfortable using a computer
- Not willing to wear circum-aural headphones
- Computer does not have a camera and a microphone

Thirty-eight participants completed the screening tests and signed consents. For screening tests, they were asked a series of case history questions to ensure eligibility for the study (Appendix B). The questionnaire included demographic, medical, and audiologic questions. An otoscopic examination was performed when in-person protocols were administered to ensure that the participant's ear canals were free from occluding wax. The hearing test and word recognition test were administered. Lastly, the Montreal Cognitive Assessment (MoCA) was administered to ensure that participants did not have significant cognitive function and working memory problems.

Two participants did not meet the screening criteria for the study and were excluded from participation. Therefore, thirty-six individuals participated in the current study. Five participants were male and thirty-one were female between 19 and 31 years old (Mean = 23.23). Most participants were recruited from a program of the University of Pittsburgh's Clinical and Translational Science Institute (CTSI) via the Pitt+Me™ Research Registry. While some participants were recruited from classes taught in the Communication Science and Disorder (CSD) programs, flyers, and word of mouth through friends. Participants were paid $20 for their research participation upon completion. If they did not qualify for the main experiment, they were paid $5 for their time spent during the screening tasks. Ten individuals participated in the research activities in person at the HEAR Core Laboratory in the Forbes Tower building at the University of Pittsburgh, while twenty-six participants completed the protocol remotely via online appointments. All of the ten participants who completed the research protocols in lab had clinically normal hearing in both ears (25 dB HL or less) as measured by pure-tone audiometry at octave frequencies from 250 – 8,000 Hz and at half-octave frequencies at 3,000 and 6,000 Hz. Their Northwestern University Auditory Test No. 6 (NU-6) 25-word list (Tillman & Carhart, 1966) word recognition testing scores were between 88% and 100% and were within 95% confidence interval according to their pure tone average as plotted on the sprint chart for 25-word lists (Thibodeau, 2000). Their middle ear status also was assessed by otoscopic examination and tympanometry and showed no sign of middle ear problem. The other twenty-six participants who completed the study remotely took an online hearing test via mobile application (either uHear for iOS users or Hearing Test for Android users) and the results showed that they had normal hearing in both ears defined as having thresholds better than 25 dB HL across 1000 to 6000 Hz for results obtained from uHear app and 1000 to 8000 Hz for results obtained from Hearing Test app.   Their NU-6 word recognition scores at comfortable listening levels as determined by the subject were between 92% and 100% (Mean = 98.31%). To determine comfortable listening levels, participants were asked to adjust their computer volume while listening to a 1000-Hz calibration pure tone used for speech test calibration until they heard the sound comfortably and the volume was kept at that setting throughout the test.

These participants did not have objective middle ear evaluation but were included if they indicated they had not had any middle ear problems. None of the participants in the study reported any recent middle ear problems (within the last three months), neurological disorder, psychological disorder, or motor speech disorder. Their Montreal Cognitive Assessment (MoCA) scores were between 26 and 30 (Mean = 28.08) indicating cognitive function and working memory within normal limits.

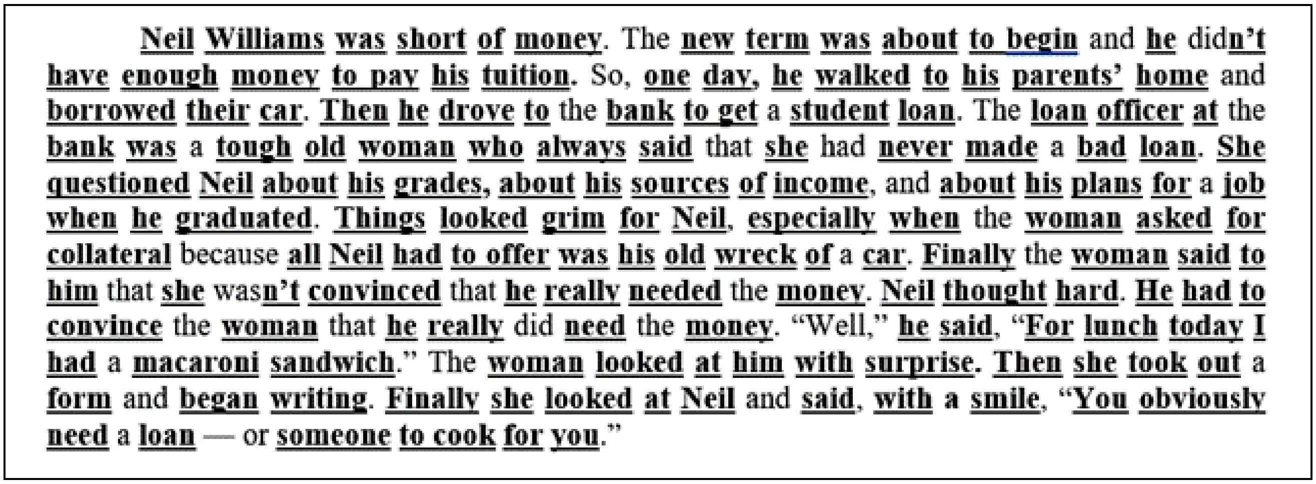
6.1.2 Materials

6.1.2.1 Stimuli

Speech stimuli used in this study were extracted from contents of the Story Retelling Procedure (SRP) (Doyle et al., 2000; McNeil et al., 2007) which is a test for differential diagnosis of persons with aphasia. It elicits connected spoken speech by having a person listen to a passage that ranges from 1 minute to 1 minute and 40 seconds and then retell the story. All retold passages were recorded with a male talker in a clear speaking manner. A listener is asked to listen to the whole story at once, and then retells the story in his/her own words. The correct Information Units (IUs) are scored, then the %IUs is computed. There are four forms containing three passages, so there are total of twelve passages (a sample passage is shown in Figure 6). Each passage contains about 111-162 IUs (or an average of 152 IUs) which is defined as an identified word, phrase, or acceptable alternative from the story stimulus that is intelligible and informative and that conveys accurate and relevant information about the story (McNeil, Doyle, Fossett, Park, & Goda, 2001).

The Percent Information Units (%IUs) is calculated to examine how many correct IUs a listener achieved compared to the total number of words. The %IUs can be used to quantify the informativeness of connected language on story retellings in the same way as using a percent correct from several auditory tests.

The rationale of using the contents of the SRP test is that it allows a standard talker (who produced speech used in the current study) to produce speech in three different speaking styles: clear speech, laboratory conversational speech, and natural conversational speech. Specifically the talker could produce natural speech by retelling the story which results in more conversational-like speaking styles that individuals encounter in day-to-day conversation and there is an option for using other reasonable alternative words that convey the similar meaning which allows for systematic scoring in the main experiment.



Neil Williams was short of money. The **new term was about to begin** and **he** did**n't have enough money to pay his tuition**. So, **one day, he walked to his parents' home** and **borrowed their car**. **Then he drove to** the **bank to get** a **student loan**. The **loan officer at** the **bank was** a **tough old woman who always said** that **she** had **never made** a **bad loan**. **She questioned Neil about his grades, about his sources of income**, and **about his plans for** a **job when he graduated**. **Things looked grim for Neil**, **especially when** the **woman asked for collateral** because **all Neil had to offer was his old wreck of** a **car**. **Finally** the **woman said to him** that **she** was**n't convinced** that **he really needed** the **money**. **Neil thought hard**. **He had to convince** the **woman** that **he really** did **need** the **money**. "Well," **he said**, "**For lunch today I had** a **macaroni sandwich**." The **woman looked at him with surprise. Then she took out** a **form** and **began writing**. **Finally she looked at Neil** and **said**, **with a smile**, "**You obviously need** a **loan** — or **someone to cook for you**."

Figure 6 Loan Passage, one of the twelve SRP passages

When creating clear speech and laboratory conversational speech, the standard talker was asked to read the story aloud in a clear and conversational speaking manner. Natural conversational speech was created by asking the talker to retell the story that he heard as if talking to a friend. Speech materials were created using all twelve stories from the SRP test. The content of each story was produced by a standard male talker in three speaking styles. The talker was instructed with the following instructions to produce speech stimuli for each speaking style accordingly. The speaker was instructed to say a passage three times for each speaking style. The following instructions from Schum (1996) were adapted and given.

Instruction for clear speech:

"I want you to read a story aloud in a clear manner. Imagine that you are speaking to a person with hearing-impairment. I want you to speak as clearly and precisely as possible and try to produce each word as accurately as you can."

Instruction for laboratory conversational speech (conversational speech in lab setting):

"I want you to read a story and make yourself familiar with it. Memorize the story as much as you can. You may look at the story while you are saying it. Keep in mind that I want you to speak clearly and naturally. Conversational speech is different from the clearly spoken speech you used before. For example, you tend to talk faster in conversation."

Instruction for natural conversational speech (naturally produced conversational speech):

"You will hear a story. Listen to it carefully. After that I want you to retell the story in your own words as close as possible to the one you heard. Speak naturally as you would in conversation with your friends and family and imagine that you are telling them a story with details as much as you can."

Initially, there were three male talkers who produced recordings of each speaking style for all twelve passages, but only recordings of one single talker were chosen. In total, a hundred and eight passages were recorded. These were recorded at a rate of 44.1 kHz in a double-wall sound booth. The speaker was seated in the sound booth with his mouth 5 inches away from a microphone (bandwidth 20 – 20,000 Hz). The microphone was routed to a PC digital recorder with settings for a mono recording. The sensitivity of the microphone was adjusted to prevent any peak clipping of the speaker's voice. These recordings were played to thirty young-adult listeners with clinically normal hearing. These recordings were presented binaurally at a comfortable listening level under Sennheiser HD 280 Pro headphones in a random order of speakers and speaking styles. Listeners were asked to listen to recordings and then provided a rating based on how natural the recording sounded to them. For each recording they provided ratings from a scale of 1 (speech is extremely clear) to 7 (speech is extremely natural, sounds like they were talking to people). Spearman Ranked Order Correlation by talkers was conducted between subject ratings and speech production of the recordings. Cohen's standard was used to evaluate the strength of the relationship, where coefficients between .10 and .29 represent a small association, coefficients between .30 and .49 represent a moderate association, and coefficients above 0.50 represent a large association. For Talker 1, there was a significant positive correlation between subject ratings and recordings ($\mathrm{r} = 0.677, \mathrm{p} < .001$); Talker 2, there was a significant positive correlation between subject ratings and recordings ($\mathrm{r} = 0.746, \mathrm{p} < .001$); and Talker 3, there was a significant positive correlation between subject ratings and recordings ($\mathrm{r} = 0.901, \mathrm{p} < .001$). The results of all the talkers indicate a strong relationship between subject ratings and their speech production of the recordings. According to results of the analysis and the boxplot (Figure 7), the recordings from Talker 3 were selected to serve as the stimuli in the main experiment. Acoustic differences among clear, lab conversational, and natural conversational speech analyzed average sentences are displayed in Table 6. Significant differences were found (p < .05) across the three speaking styles for speaking rate, speech rate, articulation rate, ratio of pause duration and total duration, and raw intensity before the processing. Significant difference in Mean F1 and F2 were not observed. For Mean F0, the significant difference (p = .022) was revealed only between clear and lab conversational speech.

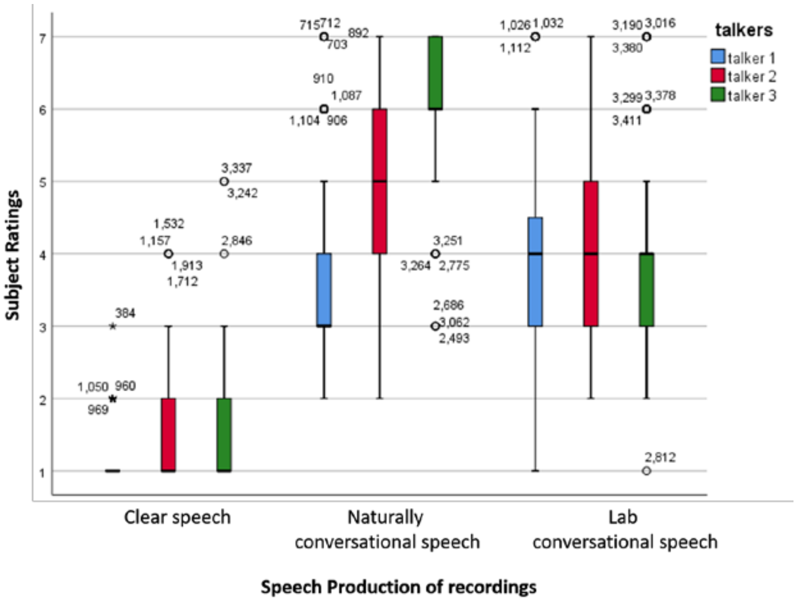


Figure 7 Boxplot showing the relationship between subject ratings and speech production by talkers

Table 6 Acoustic differernces among clear, lab conversational, and natural conversational speech

	Clear		Lab Conversational		Natural Conversational	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Speaking Rate (wpm)	121.38	29.86	208.44	12.44	178.92	14.02
Speech Rate (Syllable per Total duration)	2.83	0.38	3.81	0.25	3.43	0.26
Articulation Rate (Syllables per phonation time)	3.75	0.37	4.90	0.26	4.52	0.23
Pause/Total Duration ratio	0.35	0.11	0.18	0.96	0.23	0.10

Table 6 Acoustic differernces among clear, lab conversational, and natural conversational speech (continued)

	Clear		Lab Conversational		Natural Conversational	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Raw Intensity before processing (dB SPL)	67.81	3.95	65.64	3.34	63.44	4.12
Mean F0 (Hz)	118.24	56.77	148.50	58.50	133.61	86.76
Mean F1 (Hz)	467.56	121.47	506.93	225.55	507.95	111.33
Mean F2 (Hz)	1589.48	418.44	1595.76	426.52	1574.05	249.43

The stimuli used in the current study were created based on the recordings of Talker 3. Recordings from twelve passages in three different speaking styles were cut into sentences with a range of 8 to 15 words; 135 sentences for natural conversational speech and 180 sentences each for clear and lab conversational speech. The list of sentences for each speaking style covered all twelve stories. If all of the sentences were used, a listener might have been able to remember the sentence between conditions. To minimize the redundancy in the content for each speaking style, a total of 183 unique sentences were chosen as speech stimuli for the current study. No sentences were repeated within speaking style or among the different speaking styles to prevent learning effects.

6.1.2.2 Masker

Speech material presented in a quiet condition lacks real-world validity because the majority of listening situations that individuals encounter are in noise. Also, people, particularly individuals with hearing loss, often complain of not hearing well in background noise especially when around multiple talkers.

Instead of testing speech perception in the quiet condition only, examining speech perception in noise is helpful to accurately assess the hearing ability of individuals and to provide realistic consultation for a person with hearing loss. Different maskers affect speech perception performance (e.g., percent correct scores) differently. By definition, noise is an acoustic phenomenon that has random and aperiodic features with a continuous spectrum; psychologically speaking, noise is any undesirable sound or signal (Durrant & Feth, 2012). Any sounds can be counted as a noise depending on a listener's perspection and listening situation. There are different types of maskers implemented in studies in hearing and speech perception such as white noise, pink noise, steady-state noise, speech-shaped noise, and speech babble noise. They are distinct in characteristics, advantages, disadvantages, how they interact with the speech signal, and the impact on speech perception (see Table 7).

Table 7 Comparison among various maskers

Maskers	Characteristic	Masking effects
White noise	Flat spectrum across all frequency, equal power in any given bandwidth	- Greater masking effect on acoustic cues above 1000 Hz - Many acoustic cues will be masked such as cues for fricatives, F2, and F3
Steady-state noise	Modulated noise; Change in level less than 5 dB at any given frequency during a given time	- Variability in noise level is reduced - Produce energetic masking

Table 7 Comparison among various maskers (continued)

Maskers	Characteristic	Masking effects
Speech shaped noise	White noise that its spectrum is shaped as long-term spectrum of the speech signal	- Greater masking effect than white noise on acoustic cues below 1000 Hz - Similar masking effect as multi-speaker speech babble - Less representative of everyday communication
Speech babble	A number of other speakers speaking at the same time	- Similar masking effect as speech shaped noise - Introduce the confound of informational masking - More representative of everyday communication

Theunissen et al (2009) reported that the two most commonly used maskers are steady-state speech-shaped noise and multi-talker babble. Wilson, Carnell, et al (2007) also supported the use of multi-talker babble over steady-state speech-shaped noise due to complaint of difficulty understanding speech in noise.

Besides the types of noise, the amount (level) of noise relative to the level of the speech signal, known as signal-to-noise ratio (SNR), also plays an important role in determining the ability to understand speech in noise. To obtain a straightforward percent correct score, a fixed signal to noise ratio (SNR) should be implemented. Wendt et al (2018) examined the effect of signal-to-noise ratio on listening effort by testing under different SNR conditions ranging from -20 dB to +8 dB (in 4 dB steps) which produced speech intelligibility ranging from 0% to 100% correct for the Hearing in Noise Test (HINT) sentences. As a result of this experiment, they found that high recognition performance (100% correct) was achieved at SNR between +4 and +8 dB. The performance decreased with a decreasing SNR, at SNR -12 dB, the performance was around 5-7% correct, and the performance was impossible (0% correct) with SNR of -16 and -20 dB (Wendt, Koelewijn, Książek, Kramer, & Lunner, 2018). For comparable length of stimuli (i.e., passage-level stimuli), the CST was reported to be sensitive to a small change in SNR that produces a large change in scores (% correct). The researchers found that the range of 9 dB SNR (from 0 to -9) could result in a change in speech performance from 100% intelligibility to essentially 0% intelligibility, so their take-home message was to select the SNR carefully when using the CST (Cox, Alexander, & Gilmore, 1987).

Therefore, in addition to presenting the stimuli in a quiet condition, the stimuli were mixed with a four-talker speech babble noise to create another 4 listening conditions that represented more real-world listening environments because people, particularly individuals with hearing impairment, often complain of not hearing well in background noise consisting of multiple talkers. Therefore, the multi-talker babble was embedded with the speech stimuli at 3-dB SNR increments creating signal-to-noise ratios at +3, 0, -3, and -6 dB SNR.

6.1.3 Instrumentation

For in-person testing, the experiment was done in a double-wall sound-treated booth. The speech stimuli were presented through the MADSEN Astera[2] audiometer using ER-1 insert earphones which deliver the same frequency response to the average eardrum (i.e., normal ear canal resonance) of the open ear listening condition (see Appendix A). Stimuli were presented bilaterally at 65 dB SPL. The order of stimuli was randomized to prevent order effects. Prior to each participant listening, the sound level was verified by measurement of the calibration noise. The same microphone that was used in stimuli recording was used to record participant responses for later transcription and scoring. Participant responses were captured by using Adobe Audition CS 5.5 software. The stimuli presentation was completed via SuperLab 5 software allowing participants to control the experiment at their own pace with an option for breaks.

For remote testing, the experiment was done at the participant's residence in a relatively quiet room environment. Participants were required to measure the sound pressure level of their test surrounding (i.e., ambient noise level) to ensure that they were in an appropriate test setting. The ambient noise level was measured using a smart phone application and was less than 40 dBA. The stimuli presentation including control of listening level and response recording was performed on the participants' computer or laptop through an online experiment hosted and administered via https://gorilla.sc. Sennheiser HD 280 Pro headphones were shipped to participants and were used for stimulus presentation to control for variability from using different models of headphones. Frequency response of these headphones can be found in Appendix A. Attached microphone on participants' computer or laptop was used to record their responses. Prior to the experiment, participants were asked to download a sound level meter application (SoundMeter X or NIOSH SLM) onto their phone. They were required to measure sound level of their test surroundings to ensure that they were in an appropriate test setting. They also were asked to measure the calibration noise at their headphones on the right side by placing the microphone of their phone against the inner aspect of the headphone cup. Then they adjusted their computer's volume as necessary until the sound level meter application read 65 dB (Z). They were asked to wear the headphones and listen to the stimuli at this specific volume setting throughout the experiment. Participant response recording was captured by the recording function through Pitt Zoom meeting and the host server. Participants were asked to speak as clearly as possible when responding and were able to complete the experiment at their own pace in one sitting with an option for breaks.

6.1.4 Procedure

6.1.4.1 Experiment 1

To investigate Specific Aim 1 (i.e., to determine if there is any significant difference among speech intelligibility of clear, laboratory conversational speech, and natural conversational speech), a repeated measure design was used meaning a participant listened to all three speaking styles in all listening conditions (i.e., quiet, +3 dB SNR, 0 dB SNR, -3 dB SNR, and -6 dB SNR). The dependent variables (DVs) are percent correct of identified keyword (i.e., speech intelligibility).

6.1.4.1.1 Procedure for in-person testing

Participants took part in a consent process as approved by the Institutional Review Boards of The University of Pittsburgh. All of the procedures were completed in the HEAR Core in Forbes Tower, School of Health and Rehabilitation Sciences. Pure-tone audiometry and word recognition testing were completed in a double-wall sound booth, completed using ER-3 earphones. Other screening tasks included the case history questionnaire, otoscopic examination, tympanometry, and MoCA test and were completed outside of the sound booth in the HEAR Core. Participants who qualified for participating in the study were then seated in the sound booth in front of the computer screen and instructed that they would carefully listen to sentences from a male talker in both quiet and noise conditions and to repeat exactly what they heard. They also were encouraged to guess if they were not sure what they heard. Each participant completed a practice trial in the quiet condition and in the noise condition at the most favorable SNR (+3 dB) to ensure that the participant understood the instructions and the task for the main experiment. Feedback was provided after the practice trial.

After the completion of the practice trial, participants continued to listen to the stimuli in the experimental trials containing 183 sentences in a random order of speaking styles and listening conditions. Feedback was not provided during the experimental trials. The entire experimental task took approximately 60 to 90 minutes to complete. Participant responses were recorded. After the session, the participant responses were reviewed and transcribed by one of two native English speakers. For 5 participants, computer software was used for transcription. Scoring of keywords was completed based on the transcriptions. Because transcription was completed by two different individuals, the interrater reliability between the two transcribers was calculated based on a random sampling of 5% of the data. The interrater reliability result showed 97.67% agreement between the two transcribers. In addition, the automatic software transcription used for 5 participants was evaluated by comparing the software transcription of a set of test stimuli (i.e., audiofiles) in the quiet condition to the actual stimuli which resulted in 95% accuracy. Finally, a set of participant responses transcribed by the software was compared to human transcription and resulted in 93.82% agreement.

6.1.4.1.2 Procedures for Remote Protocols

Due to the COVID-19 pandemic, some of the in-person protocols were adjusted so that the data collection could be done remotely. Participants were asked to complete the experiment in an optimal test environment that was quiet, free of distractions, with no other activity while doing the experiment. Details of the remote protocols are described below.

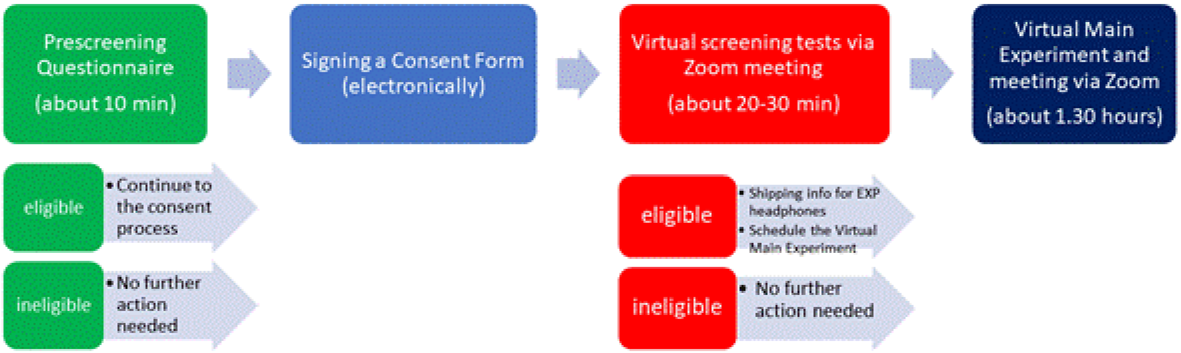

Prescreening Questionnaire
(about 10 min)
Signing a Consent Form (electronically)
Virtual screening tests via Zoom meeting
(about 20-30 min)
Virtual Main Experiment and meeting via Zoom
(about 1.30 hours)
eligible
• Continue to the consent process
ineligible
• No further action needed
eligible
• Shipping info for EXP headphones
• Schedule the Virtual Main Experiment
ineligible
• No further action needed

Figure 8 Diagram showing process for remote protocols

1. A Pre-screening questionnaire via Qualtrics was sent to prospective participants before enrolling them in the study. The questionnaire included demographic, medical, audiologic questions, and uploading a screenshot of his/her hearing test result from one of the specified hearing test mobile applications depending on the participant's phone operating system (i.e., uHear or Hearing Test).
2. Upon review of the answers to the questionnaire: if the answers are consistent with the study's criteria, the participant received an electronic consent form via DocuSign along

with the explanation of the research study including nature of the study, a list of research activities, and risks and benefits of participation.

3. When they decided to participate, they electronically signed their name and date to the consent via DocuSign.
4. Upon receiving the signed consent, the PI scheduled a Virtual Screening Appointment via Pitt's Zoom to perform the screening procedures which included the Word Recognition Test and MoCA Test (version 8.1 via Audio-Visual Conference).
    a. If met all criteria: they were invited into the study.
    b. If not: they were excluded from the study.
5. Proceed to the Virtual Experiment Session via Pitt's Zoom the same day if the participant's headphones were the same model (i.e., Sennheiser HD 280 Pro) that we were supplying to subjects. Otherwise, the Virtual Experiment Session was scheduled for another date and time when the participant received the study-provided headphones.
    a. Participants were asked to download a sound level meter application (SoundMeter X or NIOSH SLM) onto their phone to measure their room ambient noise and stimulus level.
6. A link for the main experiment task (Experiment 1) was sent to the participants so they could control their own pace while completing the task. The experimenter also audio-recorded the participant's response (similar to the task done in laboratory) via Pitt's Zoom meeting for later transcription.
    a. Prior to the experimental task, the participant was required to measure the sound pressure level of their test surrounding to ensure that they were in an appropriate test setting.

b. They also were asked to measure the calibration noise at their headphones on the right side by placing the microphone of their phone against the inner aspect of the headphone cup. Then they adjusted their computer's volume as necessary until the sound level meter application read 65 dB (Z).

c. Following the completion of sound measurement and adjustment, they were presented with the practice trial to ensure that they understood the instructions and how to do the task. Then they proceeded to the experimental trials listening to the stimuli in a random order.

6.1.4.2 Experiment 2

To investigate Specific Aim 2 (i.e., to evaluate the ability of the spectro-temporal modulation index (STMI) model in capturing speech intelligibility of speaking styles at different listening conditions), predicted speech intelligibility was derived via MATLAB script using the STMI (Elhilali, Chi, & Shamma, 2003) constructed based on the auditory model as described by Chi et al. (1999). The clean speech (i.e., speech in quiet condition) was used as a reference, while speech in noise were used to derive the predicted speech intelligibility. Therefore, each speaking style would have four predicted speech intelligibility results (i.e., at +3, 0, -3, and -6 dB SNR) for each participant. Both the predicted speech intelligibility and the measured speech intelligibility obtained from Experiment 1 were used as variables in the correlation analysis.

Experiment 2 was conducted offline (without the need of participants). The spectrogram of the speech signal (input spectrogram) was analyzed by a bank of spectro-temporal modulation selective filters creating the spectro-temporal response field (STRF). Each STRF output from each filter was computed and convolved with the input spectrogram to generate a new spectrogram with a 3-D template in terms of scale, rate, and frequency. The clean speech signal and the noisy speech signals at different SNR levels were analyzed separately to obtain the 3-D output of clean speech $\{T\}$ and the 3-D output of noisy speech signals $\{N\}$. Then, the STMI was computed using the simple Euclidian distance as in the following equation:

$$STMI = \frac{1 - \|T - N\|^2}{\|T\|^2}$$

where the Euclidian distance $\|T - N\|^2$ is the shortest distance between the noisy and clean outputs.

7.0 Results

7.1 Experiment 1

Experiment 1 was designed to investigate Specific Aim 1: To determine if there is any significant difference among speech intelligibility of clear speech, laboratory conversational speech, and natural conversational speech. A participant listened to all speaking styles (3 levels: clear, lab conversational, and natural conversational speech) in all listening conditions (5 levels: quiet, SNR +3, SNR +0, SNR -3, and SNR -6). The dependent variable (DV) was proportion correct of identified keywords (i.e., speech intelligibility).

Prior to statistical analysis, proportion correct speech intelligibility scores were converted to Rationalized Arcsine Units (RAU) to normalize variance across conditions. The data from the participants who were seen in person and the individuals who completed the study remotely were compared across listening conditions and speaking styles, thus the overall speech intelligibility across all conditions was calculated for each participant. The assumption of normality for in-person testing was met, but the assumption was violated for remote testing. Thus, three outliers (poor performers) in the remote testing group were removed allowing the assumption of normality to be met for this group. The researchers felt this was an appropriate action given that they had concern that for some individuals at home, a number of distractions during testing for specific subjects likely impacted their performance. The assumption of homogeneity was met. The independent sample t-test was performed for the mean overall speech intelligibility between the two testing conditions: in-person (n=10) and remote (n=23). Please see Appendix D.1 for a table of all of the results from the statistical analysis. The result indicated that there was no significant difference of mean overall speech intelligibility between in-person and remote testing conditions, $t(31) = 2.064, p = 0.05$, suggesting speech intelligibility performance was not impacted significantly for this set of subjects by the testing conditions whether the participants were seen in-person or completed the experiment remotely. Table 8 shows the mean overall speech intelligibility between in-person and remote testing conditions. Therefore, these data were analyzed as one group of participants.

Table 8 Mean Speech Intelligibility between Testing Conditions

Testing Condition	**Mean** **Overall Speech Intelligibility (RAU)**	**S.D.**
In-Person	64.93	6.86
Remote	59.61	6.78

A two-way repeated measure analysis of variance (ANOVA) was used to evaluate changes in speech intelligibility as a function of speaking styles and listening conditions. Mauchly’s Test indicated that the assumption of sphericity was not met; therefore, the Greenhouse-Geisser correction was used. Please see Appendix D.2 for a table of all of the results from the statistical analysis. The main effect of speaking style was statistically significant, $\mathrm{F}(1.594, 51.022) = 114.810, \mathrm{p} < .001, \eta_{p}^{2} = .782$, indicating that proportion correct speech intelligibility was significantly different across speaking styles. Post hoc comparisons using Bonferroni correction were conducted to evaluate the pattern of significant differences while controlling for Type I error. The speech intelligibility of clear speech was significantly higher than that of lab conversational speech and natural conversational speech, with mean differences of 5.47 RAU and 8.72 RAU, respectively. The speech intelligibility of lab conversational speech also was higher than that of natural conversational speech, with mean difference of 3.25 RAU. The mean speech intelligibility as a function of speaking styles are shown in Table 9.

Table 9 Mean Speech Intelligibility as a Function of Speaking Styles

Speaking Styles	Mean Speech Intelligibility (RAU)	S.E.
Clear speech	65.78	1.39
Lab conversational speech	60.31	1.39
Natural conversational speech	57.06	1.19

The main effect of listening condition was statistically significant, $\mathrm{F}(1.708, 54.651) = 152.705, \mathrm{p} < .001, \eta_p^2 = .827$. Post hoc comparisons using Bonferroni correction indicated that the highest speech intelligibility was observed in the quiet listening condition and speech intelligibility was significantly degraded in subsequent noisy listening conditions with a calculated minimum mean difference of 3.69 RAU. However, there was no significant difference in speech intelligibility between the listening conditions at +3 and 0 dB SNR. The mean speech intelligibility as a function of listening conditions are reported in Table 10.

Table 10 Mean Speech Intelligibility as a Function of Listening Conditions

Listening Conditions	Mean Speech Intelligibility (RAU)	S.E.
Quiet	68.42	1.43
+3 dB SNR	64.73	1.18
0 dB SNR	63.63	1.28
-3 dB SNR	58.79	1.36
-6 dB SNR	49.68	1.63

Moreover, the two-way interaction between speaking styles and listening conditions was statistically significant, $\mathrm{F}(5.275, 168.788) = 17.582, \mathrm{p} < .001, \eta_p^2 = .355$, suggesting that the pattern of decreased speech intelligibility for listening conditions was statistically significant across speaking styles. Post hoc analysis using Bonferroni correction was performed to evaluate the difference in speech intelligibility among the speaking styles and listening conditions; the calculated minimum mean significant difference of 3.79 was observed. However, significant differences in speech intelligibility were not observed: (1) in quiet listening condition between clear and lab conversational speech and (2) at -3 and -6 dB SNR listening condition between natural and lab conversational speech. In general, for the same listening condition, the speech intelligibility decreases greatest for natural conversational speech stimuli compared to clear and lab conversational speech stimuli. Figure 9 shows the speech intelligibility as a function of listening conditions across speaking styles.

For comparison purposes, the analysis was performed with the original data without deleting the outliers (n=36) in order to check whether the results would differ. The results obtained from the original data also showed significant main effect of speaking styles, main effect of listening conditions, and the interaction between speaking styles and listening conditions. The results of 36 participants were generally identical to the results of 33 participants.

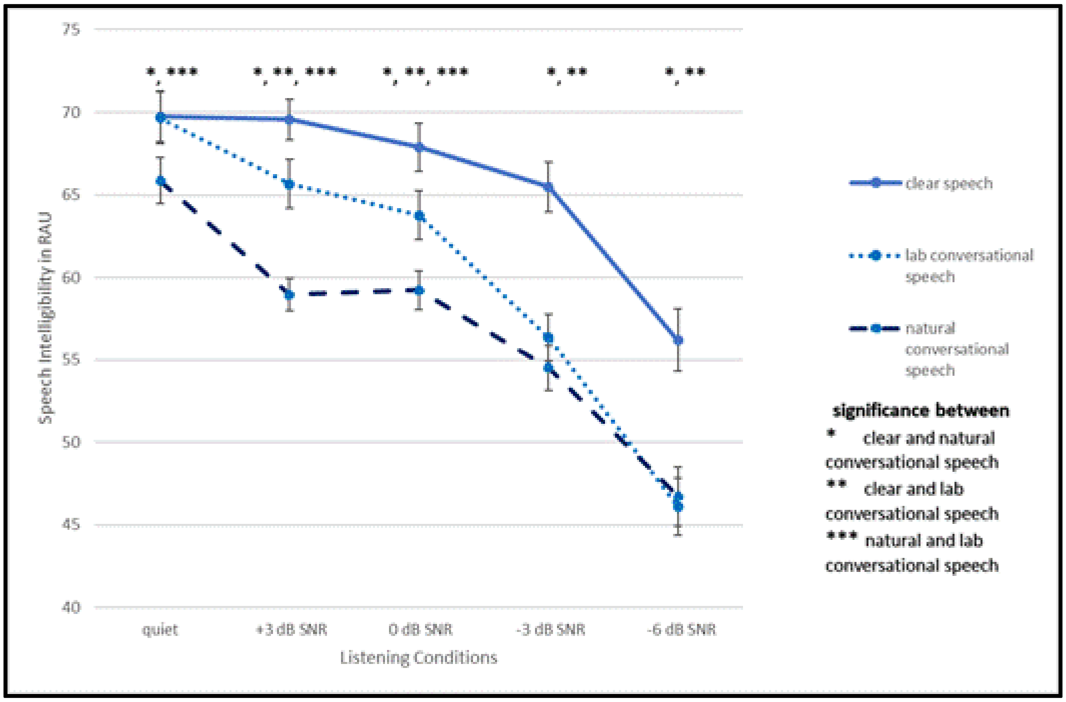


Figure 9 Line graph displayed speech intelligibility as a function of listening conditions across speaking styles

7.2 Experiment 2

Experiment 2 was designed to investigate Specific Aim 2: To determine if the STMI model can accurately predict speech intelligibility of speaking styles (i.e., clear speech, natural conversational speech, and conversational speech in lab setting). The predicted speech intelligibility scores were derived using the STMI model for listening conditions for speaking styles. The STMI values ranged from 0 to 1 (unintelligible to fully intelligible). A sample of derived STMI values can be found in Appendix E.

A Pearson product-moment correlation was conducted to examine the overall relationships between the measured speech intelligibility and the derived STMI values. There was a small correlation between the measured speech intelligibility and the derived STMI values, $r = .346, p < .001$. A scatterplot of the measured speech intelligibility and the derived STMI values is shown in Figure 10.

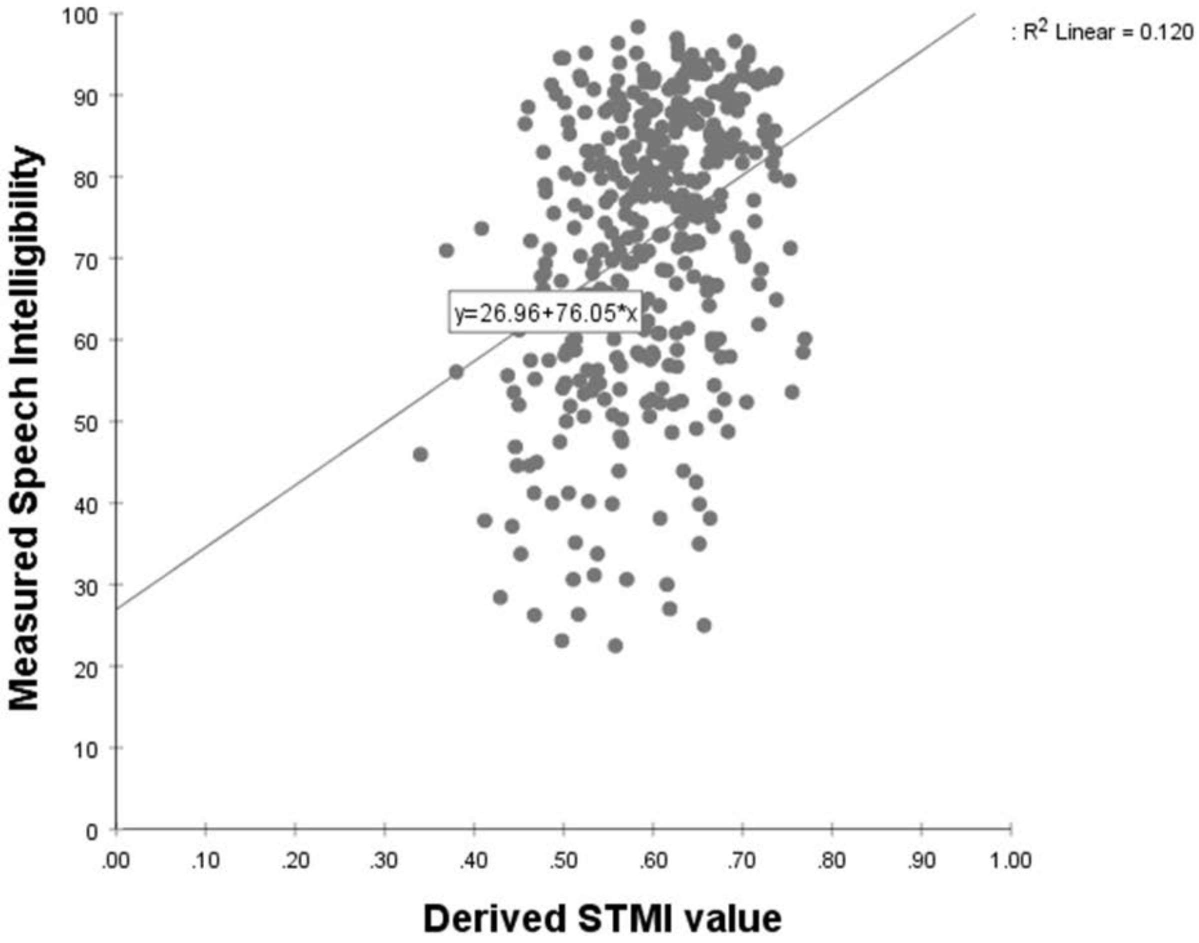


Figure 10 Scatterplot of the measured speech intelligibility and the derived STMI values

Simple linear regression analysis was conducted to predict measured speech intelligibility based on the derived STMI values. All assumptions of the linear regression were met. A significant regression equation was found, $F(1, 394) = 53.572, p < .001$ with $adjusted\ r^2$ of .117. The prediction score (percent correct) of overall measured speech intelligibility is equal to $26.964 + (76.050 * STMI\ value)$. The prediction score increased 7.605 percent correct for each one-tenth of STMI value. Only 11.7% of the variance in prediction score can be explained by the model.

Further, the data were analyzed by speaking styles. A Pearson product-moment correlation revealed significant correlation between measured speech intelligibility and derived STMI values for clear, lab conversational, and natural conversational speech with $r = .347, p < .001, r = .500, p < .001$, and $r = .283, p = .001$, respectively. Simple linear regression of each speaking style was performed. All assumptions of the linear regression were met.

A significant regression equation for clear speech was found, $F(1, 130) = 17.799, p < .001$ with $adjusted\ r^2$ of .114. The prediction score (percent correct) for clear speech intelligibility is equal to $36.100 + (74.496 * STMI\ value)$. The prediction score increased 7.45 percent correct for each one-tenth of STMI value. Only 11.40% of the variance in prediction score can be explained by the model.

A significant regression equation for lab conversational speech was found, $F(1, 130) = 43.404, p < .001$ with $adjusted\ r^2$ of .245. The prediction score (percent correct) for lab conversational speech intelligibility is equal to $-2.303 + (120.528 * STMI\ value)$. The prediction score increased 12.05 percent correct for each one-tenth of STMI value. Only 24.50% of the variance in prediction score can be explained by the model.

A significant regression equation for natural conversational speech was found, $F(1, 130) = 11.342, p = .001$ with $adjusted\ r^2$ of .073. The prediction score (percent correct) for natural conversational speech intelligibility is equal to $37.415 + (48.610 * STMI\ value)$. The prediction score increased 4.861 percent correct for each one-tenth of STMI value. Only 7.30% of the variance in prediction score can be explained by the model.

8.0 Discussion

The purpose of the current study was to evaluate the impact of speaking style on the proportion scores of speech intelligibility in different listening conditions and to evaluate if the STMI values (i.e., the predicted speech intelligibility derived from the STMI model that is sensitive to detecting the spectro-temporal modulations present in speech) can be used to predict the measured speech intelligibility. Overall, speech intelligibility was significantly different among speaking styles (clear, natural conversational, and lab conversational speech). The listening performance when presenting clear speech stimuli was significantly better than when the lab conversational speech and natural conversational speech were presented. The listening performance with lab conversational speech stimuli was significantly higher than when natural conversational speech was presented. This finding indicates that the conversational speech produced in a laboratory setting does not impact listening performance to the same extent as conversational speech spoken naturally. Therefore, using conversational speech produced in a laboratory setting does not represent natural conversational speech in terms of performance. The differences in characteristics of these two speaking styles result in differences in speech intelligibility (e.g., degree of coarticulation, vowel space, formants).

When there is a decrease in signal-to-noise ratio (SNR) producing a more difficult listening environment, the listening performance of the two conversational (lab and natural) speaking styles were relatively similar. Therefore, to overcome listening difficulty in noise, a talker might need to develop a clear speaking style so that a listener can still maintain their listening performance compared to their original performance in the quiet condition.

As shown in Figure 11, decreasing performance was observed when SNR became less favorable (from the quiet condition to -6 dB SNR). As anticipated, clear speech exhibited the highest proportion correct of speech intelligibility scores for all listening conditions while the proportion correct of speech intelligibility scores for the laboratory conversational speech lies in between that of the clear speech and that of the natural conversational speech in any listening conditions except for (1) the quiet condition where the listening performance between clear speech and lab conversational speech were the same, and (2) when the SNR became negative (i.e., -3 and -6 dB SNR) where there is a lack of performance difference between the lab conversational and natural conversational speech. These findings related to performance in noise were consistent with previous literature (Abel, Alberti, Haythornthwaite, & Riko, 1982; Brungart et al., 2020; Kalikow, Stevens, & Elliott, 1977; Li et al., 2011; Wendt et al., 2018).

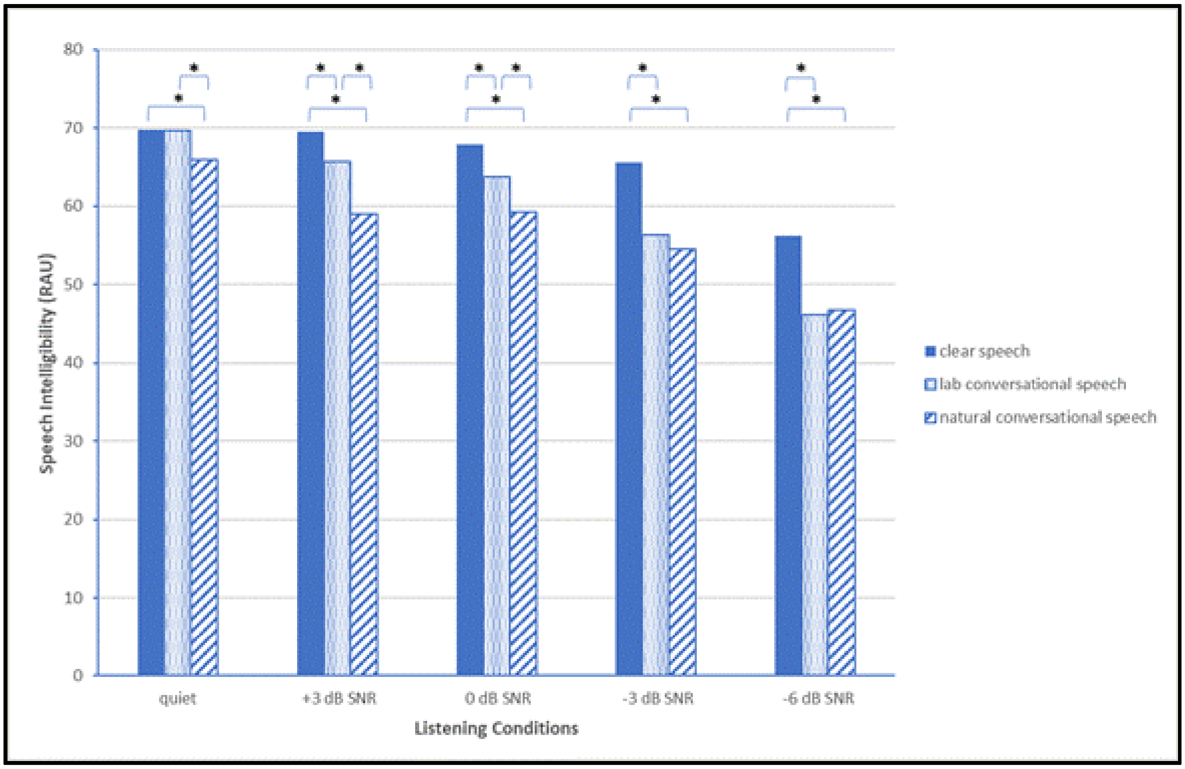


Figure 11 Mean Speech Intelligibility across Listining Conditions and Speaking Styles

The results of Experiment 1 demonstrate that noise adversely effects speech intelligibility in normal-hearing listeners especially when speech becomes less clear (e.g., less clear pronunciation, increased speaking rate). Extensive research has shown that clear speech has benefits over conversational speech, especially in challenging listening conditions or when listeners have perceptual difficulty (Ferguson, 2004, 2012; Krause & Braida, 2002; Liu et al., 2004; Maniwa et al., 2008; Picheny et al., 1985; Schum, 1996; Uchanski et al., 1996).

Figure 12 provides the long-term average speech spectrum (LTASS) of the speech signal of the three speaking styles of one sentence stimuli (normalized). These spectral cues may explain the lack of difference in speech intelligibility in the quiet listening condition between clear speech and lab conversational speech. As seen on the graph, the LTASS of clear and lab conversational speech are similar, while the energy of the speech spectrum of the natural conversational speech is decreased.

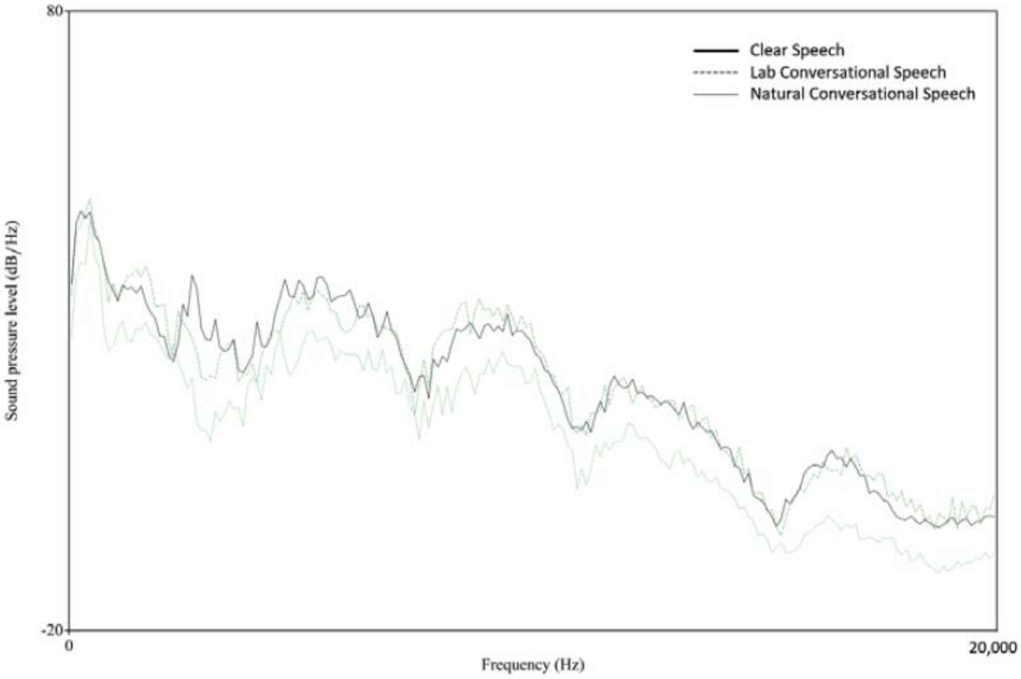


Figure 12 Long-Term Average Speech Spectrum of Speech Signal

When listening in the noise condition, the pattern difference of the speech signal speaking rate and the babble might contribute to speech intelligibility differences among clear, lab conversational, and natural conversational speech in the current study. Babble may help listeners to differentiate different speaking styles. The babble consisted of four babble talkers, including both male and female. The speaking rate of babble was similar to a typical conversational speech in everyday communication and it was held constant across all speaking styles i.e., babble has the same acoustic characteristic across speech stimuli that varied in speaking styles. The observed benefit of clear speech over conversational speech may be due to the fact that there was a significant difference in speaking rate between babble and the clear speech. This difference may assist listeners in differentiating the background noise (i.e., babble) from the speech signal (i.e., clear speech stimuli). On the other hand, the speaking rate between babble and the lab conversational and natural conversational speech signal were quite similar and listeners may not receive the temporal cue difference while listening to conversational speech in babble. Therefore, the difference in temporal cues between the babble and the speech signal could contribute to speech intelligibility. As seen in the results, the speech intelligibility of clear speech remained higher than that of both lab and natural conversational speech, and as expected, the speech intelligibility continued to decline as listening conditions became more challenging (i.e. decreased SNR). The results demonstrated a significant clear speech benefit in any difficult listening conditions.

Moreover, the pattern of fundamental frequency (F0) between the babble and the speech signal were different. As mentioned, the babble consisted of both male and female talkers, while the speech signal consisted of a male talker. The babble may have a higher fundamental frequency as compared to the speech signal of the three speaking styles. The pattern difference in fundamental frequencies also could help listeners segregate the speech signal from the babble. Acoustic analyzes of the babble showed that the mean fundamental frequency was 152.55 Hz which was close to the mean fundamental frequency of lab and natural conversational speech at 148.50 and 133.61 Hz, respectively, while the mean fundamental frequency of clear speech was 118.24 Hz. From this information, we could hypothesize that the pattern of fundamental frequency between babble and speech signal could contribute to speech intelligibility; the larger the difference in pattern, the higher the speech intelligibility. When the F0 of the speech signal was close to F0 of the babble, it could make the listening condition harder for listeners to differentiate between the speech signal and the background noise. As a result, the speech intelligibility of conversational speech decreased as compared to the speech intelligibility of clear speech.

The current findings suggest that performance is not the same among clear speech, lab conversational speech and natural conversational speech. Given that the typical goal of auditory treatment including signal processing development is targeted at individuals who are trying to communicate in real world (natural conversation) communication situations, these differences are meaningful in terms of choosing appropriate speech understanding tasks for research protocols. Using natural conversational speech may provide good face validity for hearing assessment and treatment outcome assessment.

Additionally, accuracy of speech intelligibility prediction by the STMI model was investigated in this study. Although the derived STMI values were computed using the same method described in (Elhilali et al., 2003), the current study used 4-talker babble noise to degrade the signal (speech stimuli) which is different from the original authors' noise stimulus. In the Elhilali et. al. (2003) study, the speech signals were degraded by additive white Gaussian noise and reverberation distortions, and a wide range of noise levels were incorporated. The current study tested speech intelligibility in limited noisy conditions (+3, 0, -3, -6 dB SNR). Unlike what was anticipated, the STMI values derived in the current study cannot be observed at the low and high ends of the scale. However, the correlation analysis indicated that there was a linear positive relationship between proportion scores of measured speech intelligibility and the derived STMI values, indicating that the actual intelligibility performance would increase when there was an increase in STMI value. According to the result of the simple linear regression STMI as a predictor of overall speech intelligibility, an STMI value of zero (i.e., predicted unintelligible) corresponded to a prediction score of a measured intelligibility of approximately 27% correct. While the STMI value equal to one (i.e., predicted fully intelligible) corresponded with the measured speech intelligibility of 103.01% correct. The STMI value only accounted for 11.7% of the variance in actual speech intelligibility. Although there were significant correlations between (1) speech intelligibility across speaking styles, (2) speech intelligibility for clear speech and STMI, (3) speech intelligibility for lab conversational speech and STMI, and (4) speech intelligibility for natural conversational speech and STMI; the results from linear regression indicated small adjusted R square. This suggested a poor fit of the regression models. Hence, the STMI might not be sensitive to predict speech intelligibility of noisy speech when the listening condition does not include white noise and reverberation. The STMI values derived in this study could not accurately predict speech intelligibility of individuals in multi-talker babble noise conditions. Because both signal and noise in the study were speech, there were some interactions between the speech signal and the babble noise when they were mixed. The mixed signal would exhibit different spectro-temporal modulation from the original speech signal and original babble noise alone. This interaction is different from the interaction of speech and white noise; thus, the spectro-temporal modulation of the speech may be preserved when accompanied by white noise allowing for higher predictive ability of the model.

Recently, Venezia, Hickok, & Richards (2016) conducted a study where they introduced a new technique labelled “auditory bubbles” that is more sensitive to capturing spectro-temporal cues of speech. As a result of this work, a modified STMI model may capture variance in speech intelligibility that is not reflected in the original STMI (Elhilali et al., 2003). At this writing, the updated STMI model is not available for widespread use and could not be incorporated into the data analysis. The need for a modified STMI model is consistent with the findings in the current study indicating that the derived STMI values could not explain a significant portion of the variance in speech intelligibility in multi-talker babble. The original STMI model should be adjusted to be able to capture more variance in speech intelligibility and to account for multi-talker babble noise given that this noise type is more representative of real-world listening conditions. This would be helpful in providing speech intelligibility predictions for individuals in real life listening conditions that are often reported as difficult (e.g., the cocktail party effect).

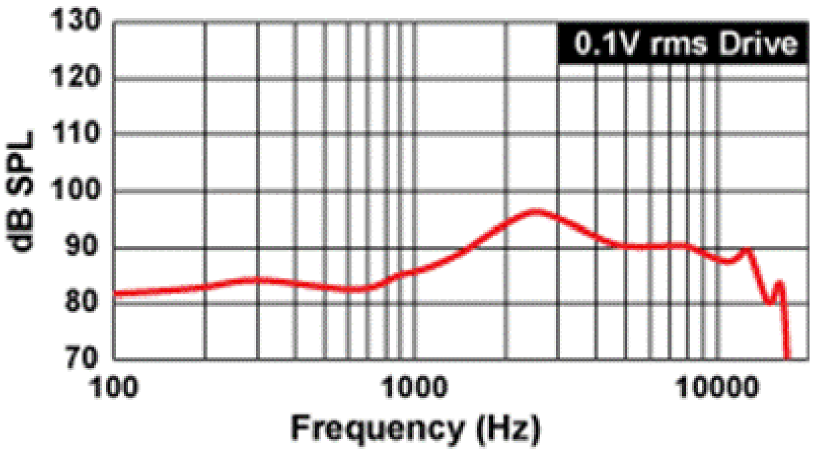
9.0 Limitations and Future Work Directions

The current study has limitations. Some limitations are related to speech stimuli. Test sentences were elicited from the content of the passages. Within each passage, pronouns were used extensively which provide a closed set of options from which to choose which might impact the intelligibility scores when they were key words (scored items). Also keywords within a single sentence may allow a listener to guess what word they would hear next. For example, when a listener hears “He drove to the bank to get a student loan.”, the word “loan” may be guessed correctly if “bank to get a student” is heard correctly. Another sentence sample would be “The third ticket is because you are parked in a no parking zone.”, the word “ticket” and “parking” might influence each other. Since keywords were embedded in a meaningful sentence, they cannot be considered as independent of each other. Therefore, there was the effect of sentence context on word intelligibility in the current study. Some sentences may be more predictable than others. Also, the stimuli were produced by one male talker; participants may adapt to the talker’s way of speaking during more a favorable listening condition (e.g., at +3 dB SNR) and be able to compensate for any information masked by babble at less favorable listening conditions (e.g., -6 dB SNR) despite compromised audibility of the keywords. Therefore, talker familiarity could play a role in speech intelligibility. In a future study, it would be of interest to investigate the talker effect and develop new speech stimuli that includes multi-talkers and the use of low predictability sentences.

Focusing on presenting more realistic stimuli and listening condition, the current study used speech babble over other types of noise making the noise choice different from the study of Elhilali et al. (2003). The results might be more comparable to the original study if white noise was used as the degrader, The STMI model might have been better at predicting speech intelligibility of different speaking styles in the original noise condition, but the purpose of this study was to use a realistic noise type found in real-world settings.

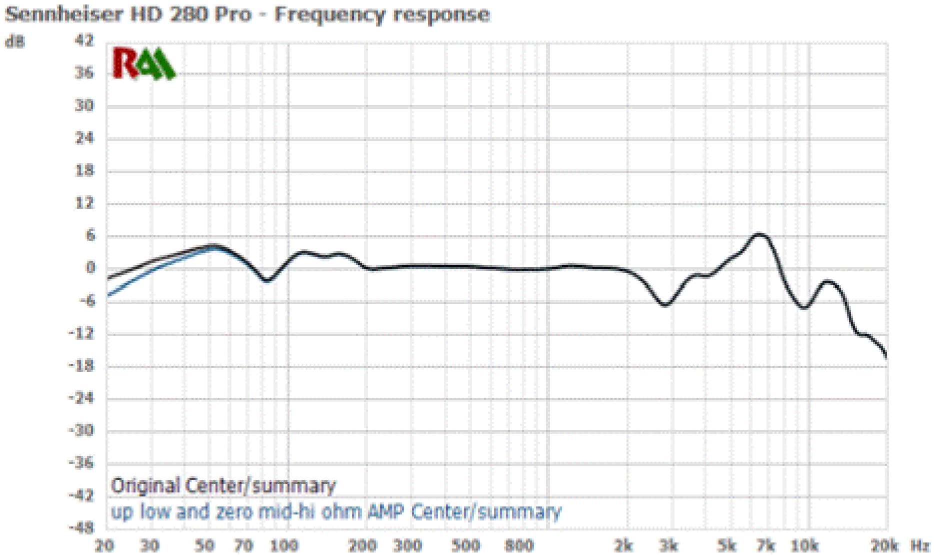
It is anticipated that the results of the current study might influence future research in speech perception by encouraging use of speech stimuli that are a true representation of naturally-produced conversational speech. A standard for true conversational speech may be developed so that appropriate test materials could be administered in a clinical setting with less time consumption to receive information about how individuals would perform in real-world conditions. In addition, these preliminary results may encourage further development of the STMI or other predictive models to better predict performance with natural conversational speech in natural noise conditions (e.g., multi-speaker babble).

Appendix A Frequency response of Earphones



Appendix Figure 1 Frequency response of ER-1 Insert Earphones, retrived from URL

https://www.etymotic.com/auditory-research/insert-earphones-for-research/er1.html



Appendix Figure 2 Frequency response of Sennheiser HD 280 Pro, retrieved from URL https://reference-audio-analyzer.pro/en/report/hp/sennheiser-hd-280-pro.php#gsc.tab=0

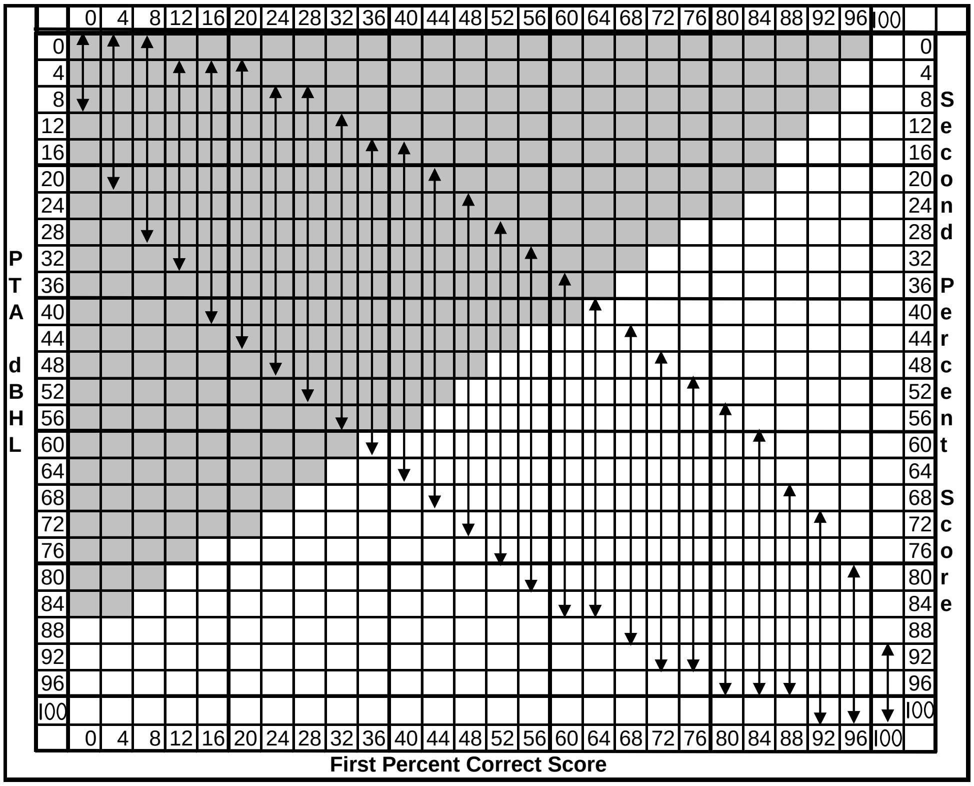
Appendix B Case History Form

- Participant Number: ________ Date: ________
- Sex: Male Female
- Age: ________ Date of birth: __________ [If younger than 18 or older than 40, exclude]
- Are you a native speaker of American English?YesNo
- Do you have hearing loss? Yes No
- Are you in good general health? Yes No
- If no, please explain your medical condition: _______
- Have you had any ear surgery? Yes No [If yes, exclude]
- Explain: ______________
- Have you had any recent ear infections, drainage, or pain in your ears?
- If yes, please indicate the date: ______[If within the last 3 months, exclude]
- Have you had any eye surgery? Yes No [if yes, exclude]
- Explain: ______________
- Have you ever had any condition that affects your brain such as: a stroke, seizure, hemorrhage, brain tumor, or other type of neurological condition? Yes No [if yes, exclude]
- If yes, please explain: ______
- Have you ever been diagnosed with any psychological disorder such as: anxiety disorder, schizophrenia, severe depression? Yes No[if yes, exclude]

- If yes, please explain: ___________
- Have you ever been diagnosed with any motor speech disorder such as: stuttering, apraxia of speech? Yes No[if yes, exclude]
- If yes, please explain: ___________

.

Appendix C Sprint Chart for 25-Word Lists (Thibodeau, 2000)

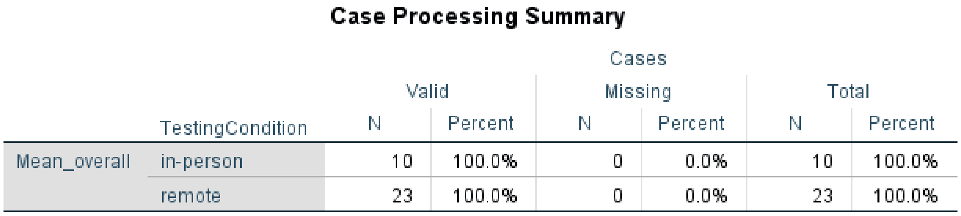


▨ 95% Confidence Limit for PBmax on NU6 25-word list.Plot score according to PTA on left ordinate and percent correct score on the abscissa. If it falls in the shaded area, it is considered disproportionately low. (Adapted from Dubno et al.,1995)

↔ 95% Critical differences for 25-word list. Plot first and second score according to the abscissa and right ordinate. If it falls within the arrow, the two scores are not significantly different (Adapted from Thornton & Raffin, 1978)

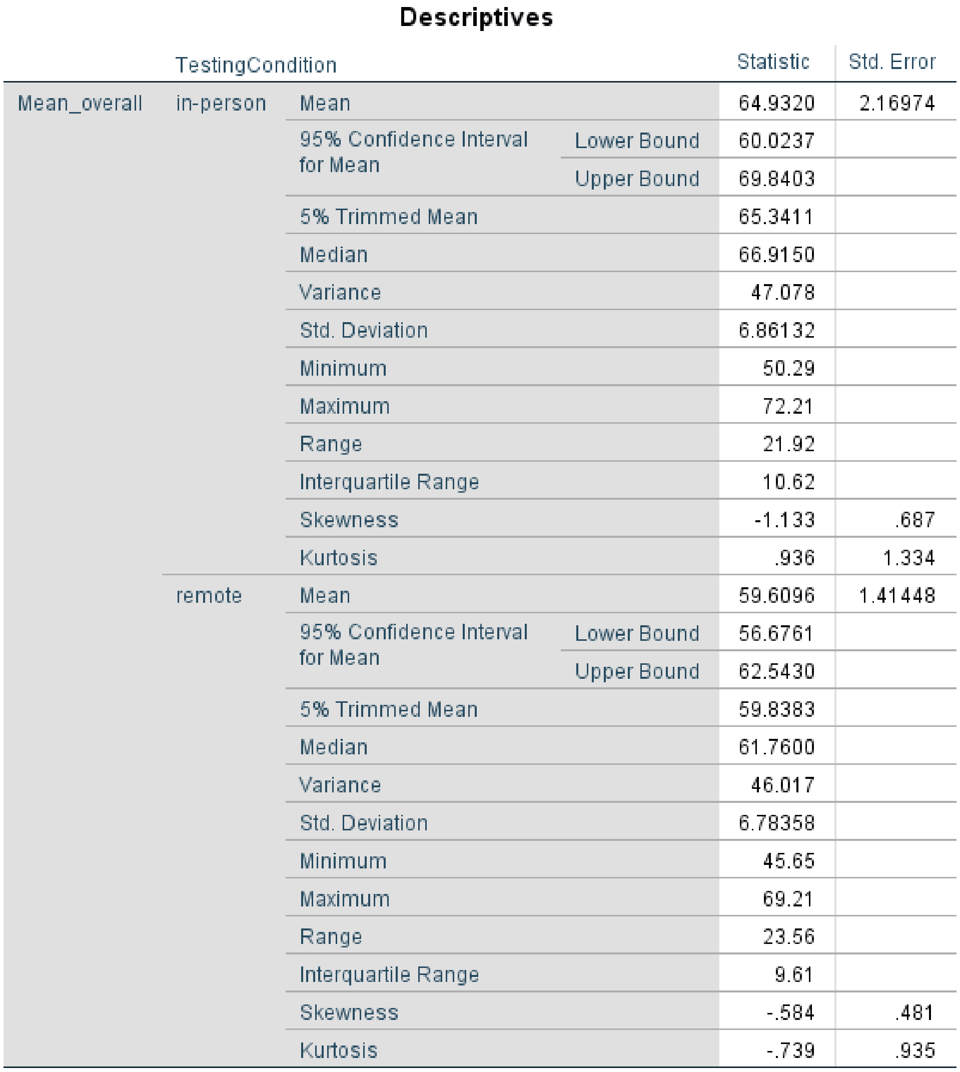
Appendix D Analysis Outputs from SPSS

Appendix D.1 Outputs of the Independent Samples T-Test



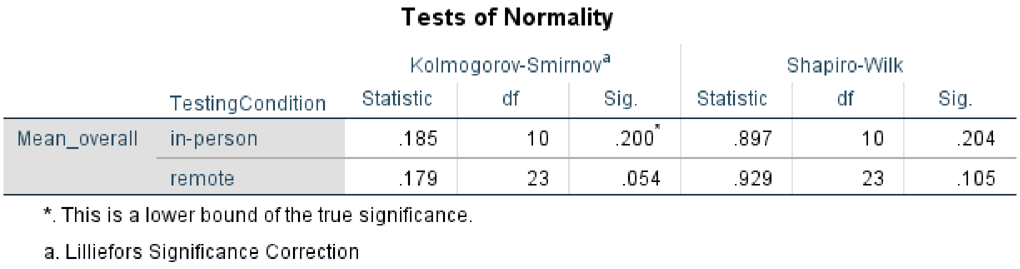
Case Processing Summary

		Cases					
		Valid		Missing		Total	
	TestingCondition	N	Percent	N	Percent	N	Percent
Mean_overall	in-person	10	100.0%	0	0.0%	10	100.0%
	remote	23	100.0%	0	0.0%	23	100.0%



Descriptives

	TestingCondition			Statistic	Std. Error
Mean_overall	in-person	Mean		64.9320	2.16974
		95% Confidence Interval for Mean	Lower Bound	60.0237	
			Upper Bound	69.8403	
		5% Trimmed Mean		65.3411	
		Median		66.9150	
		Variance		47.078	
		Std. Deviation		6.86132	
		Minimum		50.29	
		Maximum		72.21	
		Range		21.92	
		Interquartile Range		10.62	
		Skewness		-1.133	.687
		Kurtosis		.936	1.334
	remote	Mean		59.6096	1.41448
		95% Confidence Interval for Mean	Lower Bound	56.6761	
			Upper Bound	62.5430	
		5% Trimmed Mean		59.8383	
		Median		61.7600	
		Variance		46.017	
		Std. Deviation		6.78358	
		Minimum		45.65	
		Maximum		69.21	
		Range		23.56	
		Interquartile Range		9.61	
		Skewness		-.584	.481
		Kurtosis		-.739	.935

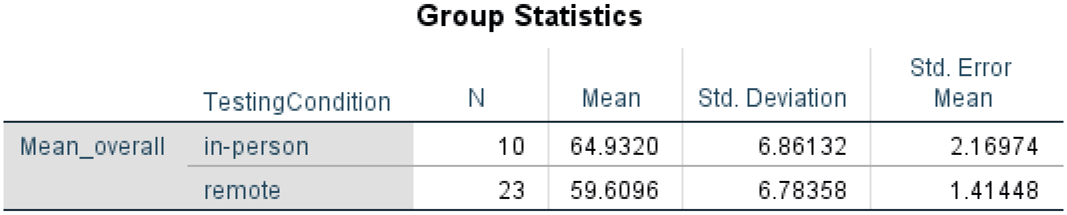


Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	TestingCondition	Statistic	df	Sig.	Statistic	df	Sig.
Mean_overall	in-person	.185	10	.200*	.897	10	.204
	remote	.179	23	.054	.929	23	.105

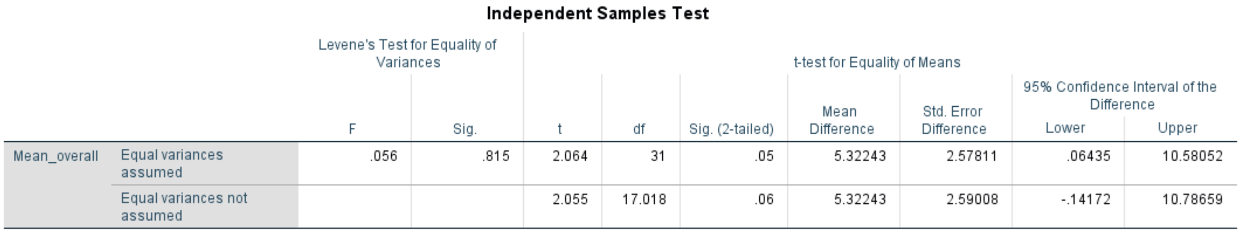
*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction



Group Statistics

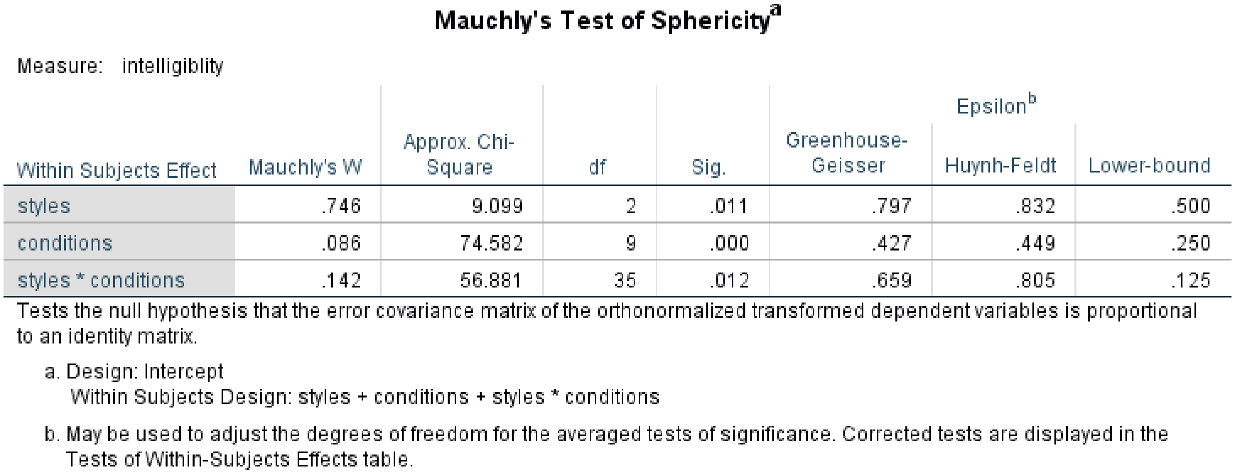
	TestingCondition	N	Mean	Std. Deviation	Std. Error Mean
Mean_overall	in-person	10	64.9320	6.86132	2.16974
	remote	23	59.6096	6.78358	1.41448



Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Mean_overall	Equal variances assumed	.056	.815	2.064	31	.05	5.32243	2.57811	.06435	10.58052
	Equal variances not assumed			2.055	17.018	.06	5.32243	2.59008	-.14172	10.78659

Appendix D.2 Outputs of Experiment 1



Mauchly's Test of Sphericity[a]

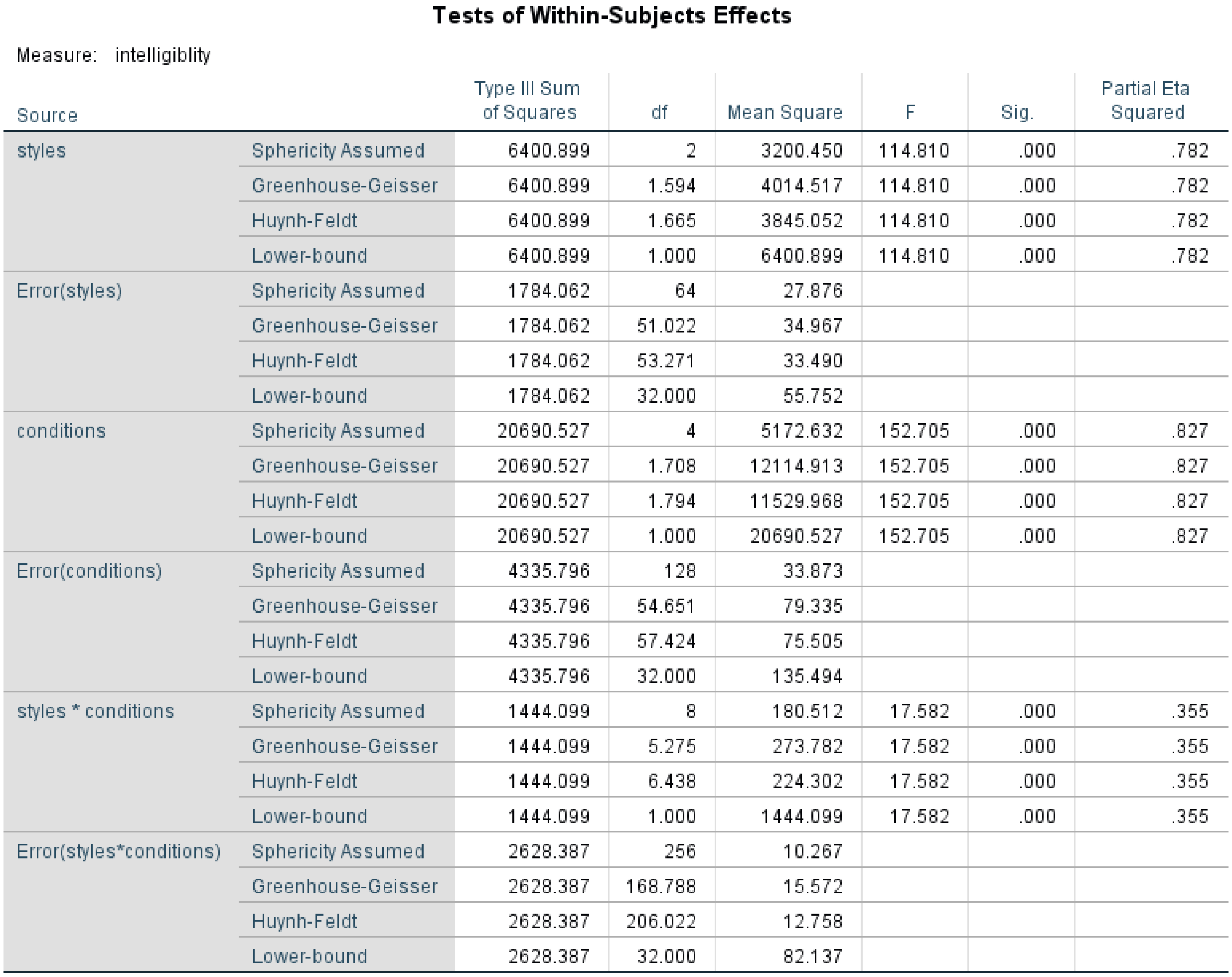
Measure: intelligiblity

					Epsilon[b]		
Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Greenhouse-Geisser	Huynh-Feldt	Lower-bound
styles	.746	9.099	2	.011	.797	.832	.500
conditions	.086	74.582	9	.000	.427	.449	.250
styles * conditions	.142	56.881	35	.012	.659	.805	.125

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept
Within Subjects Design: styles + conditions + styles * conditions

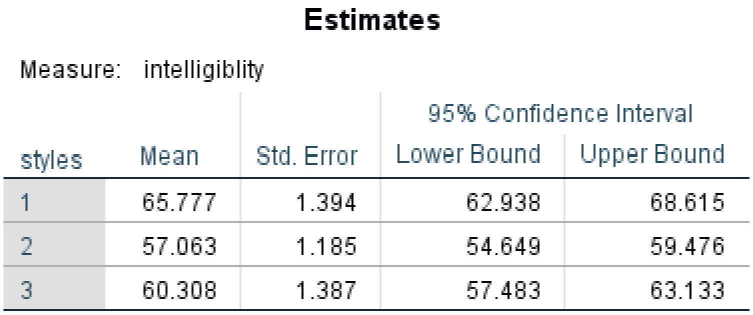
b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.



Tests of Within-Subjects Effects

Measure: intelligiblity

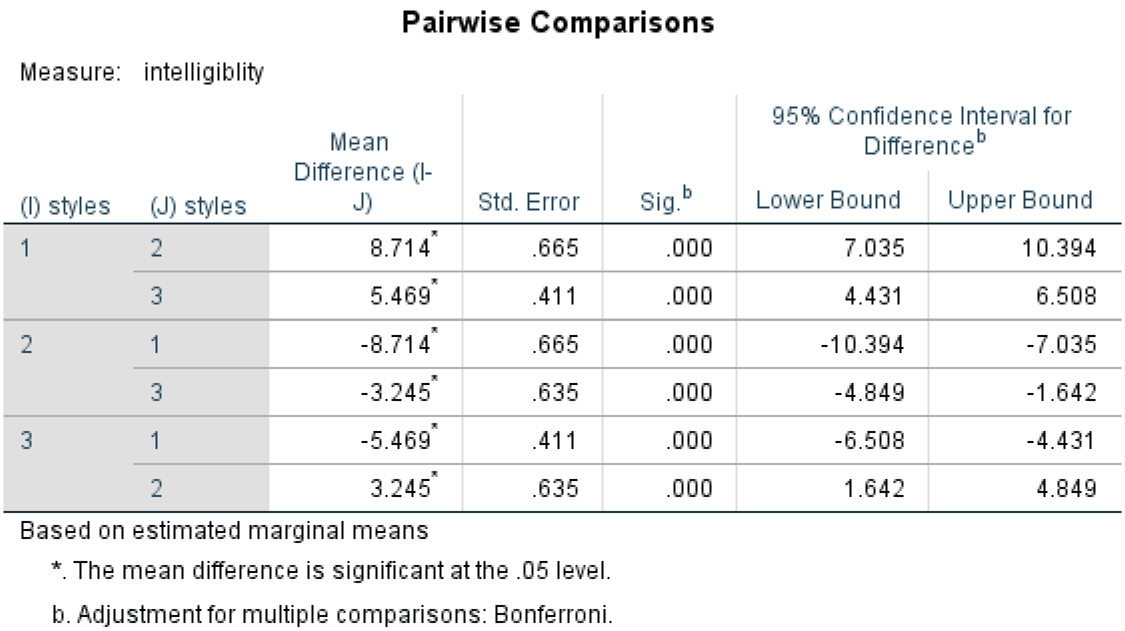
Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
styles	Sphericity Assumed	6400.899	2	3200.450	114.810	.000	.782
	Greenhouse-Geisser	6400.899	1.594	4014.517	114.810	.000	.782
	Huynh-Feldt	6400.899	1.665	3845.052	114.810	.000	.782
	Lower-bound	6400.899	1.000	6400.899	114.810	.000	.782
Error(styles)	Sphericity Assumed	1784.062	64	27.876			
	Greenhouse-Geisser	1784.062	51.022	34.967			
	Huynh-Feldt	1784.062	53.271	33.490			
	Lower-bound	1784.062	32.000	55.752			
conditions	Sphericity Assumed	20690.527	4	5172.632	152.705	.000	.827
	Greenhouse-Geisser	20690.527	1.708	12114.913	152.705	.000	.827
	Huynh-Feldt	20690.527	1.794	11529.968	152.705	.000	.827
	Lower-bound	20690.527	1.000	20690.527	152.705	.000	.827
Error(conditions)	Sphericity Assumed	4335.796	128	33.873			
	Greenhouse-Geisser	4335.796	54.651	79.335			
	Huynh-Feldt	4335.796	57.424	75.505			
	Lower-bound	4335.796	32.000	135.494			
styles * conditions	Sphericity Assumed	1444.099	8	180.512	17.582	.000	.355
	Greenhouse-Geisser	1444.099	5.275	273.782	17.582	.000	.355
	Huynh-Feldt	1444.099	6.438	224.302	17.582	.000	.355
	Lower-bound	1444.099	1.000	1444.099	17.582	.000	.355
Error(styles*conditions)	Sphericity Assumed	2628.387	256	10.267			
	Greenhouse-Geisser	2628.387	168.788	15.572			
	Huynh-Feldt	2628.387	206.022	12.758			
	Lower-bound	2628.387	32.000	82.137			



Estimates

Measure: intelligiblity

styles	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	65.777	1.394	62.938	68.615
2	57.063	1.185	54.649	59.476
3	60.308	1.387	57.483	63.133



Pairwise Comparisons

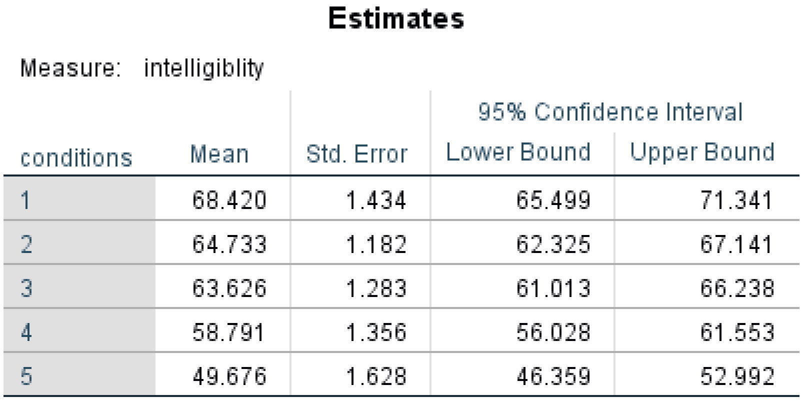
Measure: intelligiblity

(I) styles	(J) styles	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
1	2	8.714*	.665	.000	7.035	10.394
	3	5.469*	.411	.000	4.431	6.508
2	1	-8.714*	.665	.000	-10.394	-7.035
	3	-3.245*	.635	.000	-4.849	-1.642
3	1	-5.469*	.411	.000	-6.508	-4.431
	2	3.245*	.635	.000	1.642	4.849

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

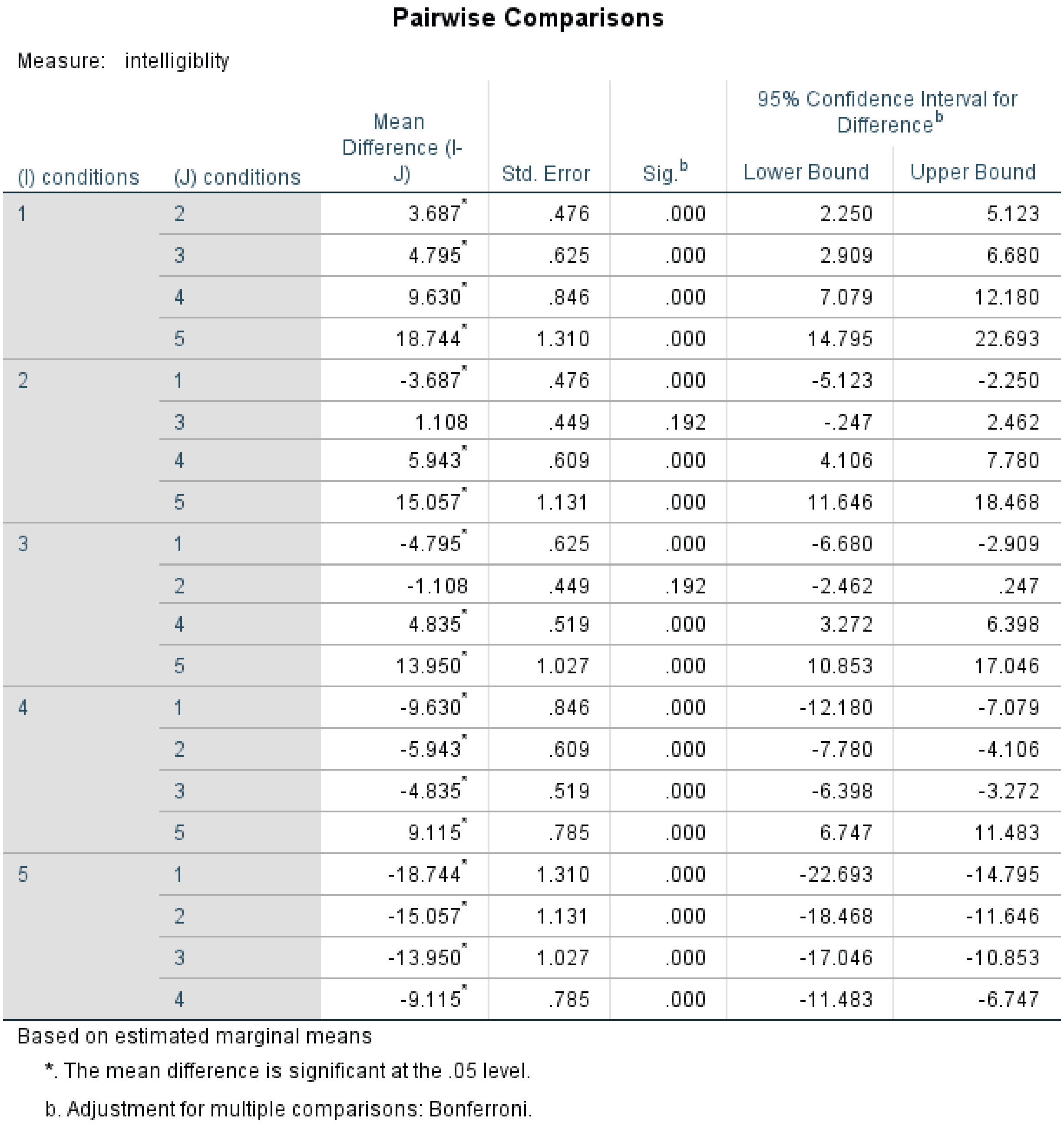
b. Adjustment for multiple comparisons: Bonferroni.



Estimates

Measure: intelligiblity

conditions	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	68.420	1.434	65.499	71.341
2	64.733	1.182	62.325	67.141
3	63.626	1.283	61.013	66.238
4	58.791	1.356	56.028	61.553
5	49.676	1.628	46.359	52.992



Pairwise Comparisons

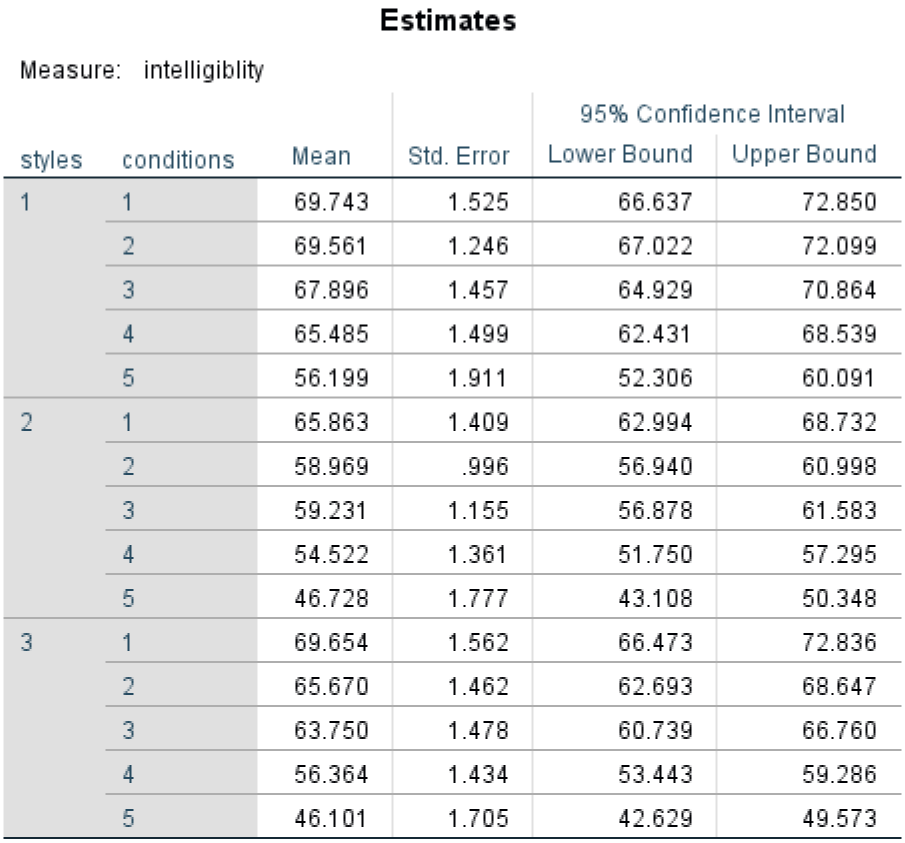
Measure: intelligiblity

		Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
(I) conditions	(J) conditions				Lower Bound	Upper Bound
1	2	3.687*	.476	.000	2.250	5.123
	3	4.795*	.625	.000	2.909	6.680
	4	9.630*	.846	.000	7.079	12.180
	5	18.744*	1.310	.000	14.795	22.693
2	1	-3.687*	.476	.000	-5.123	-2.250
	3	1.108	.449	.192	-.247	2.462
	4	5.943*	.609	.000	4.106	7.780
	5	15.057*	1.131	.000	11.646	18.468
3	1	-4.795*	.625	.000	-6.680	-2.909
	2	-1.108	.449	.192	-2.462	.247
	4	4.835*	.519	.000	3.272	6.398
	5	13.950*	1.027	.000	10.853	17.046
4	1	-9.630*	.846	.000	-12.180	-7.079
	2	-5.943*	.609	.000	-7.780	-4.106
	3	-4.835*	.519	.000	-6.398	-3.272
	5	9.115*	.785	.000	6.747	11.483
5	1	-18.744*	1.310	.000	-22.693	-14.795
	2	-15.057*	1.131	.000	-18.468	-11.646
	3	-13.950*	1.027	.000	-17.046	-10.853
	4	-9.115*	.785	.000	-11.483	-6.747

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

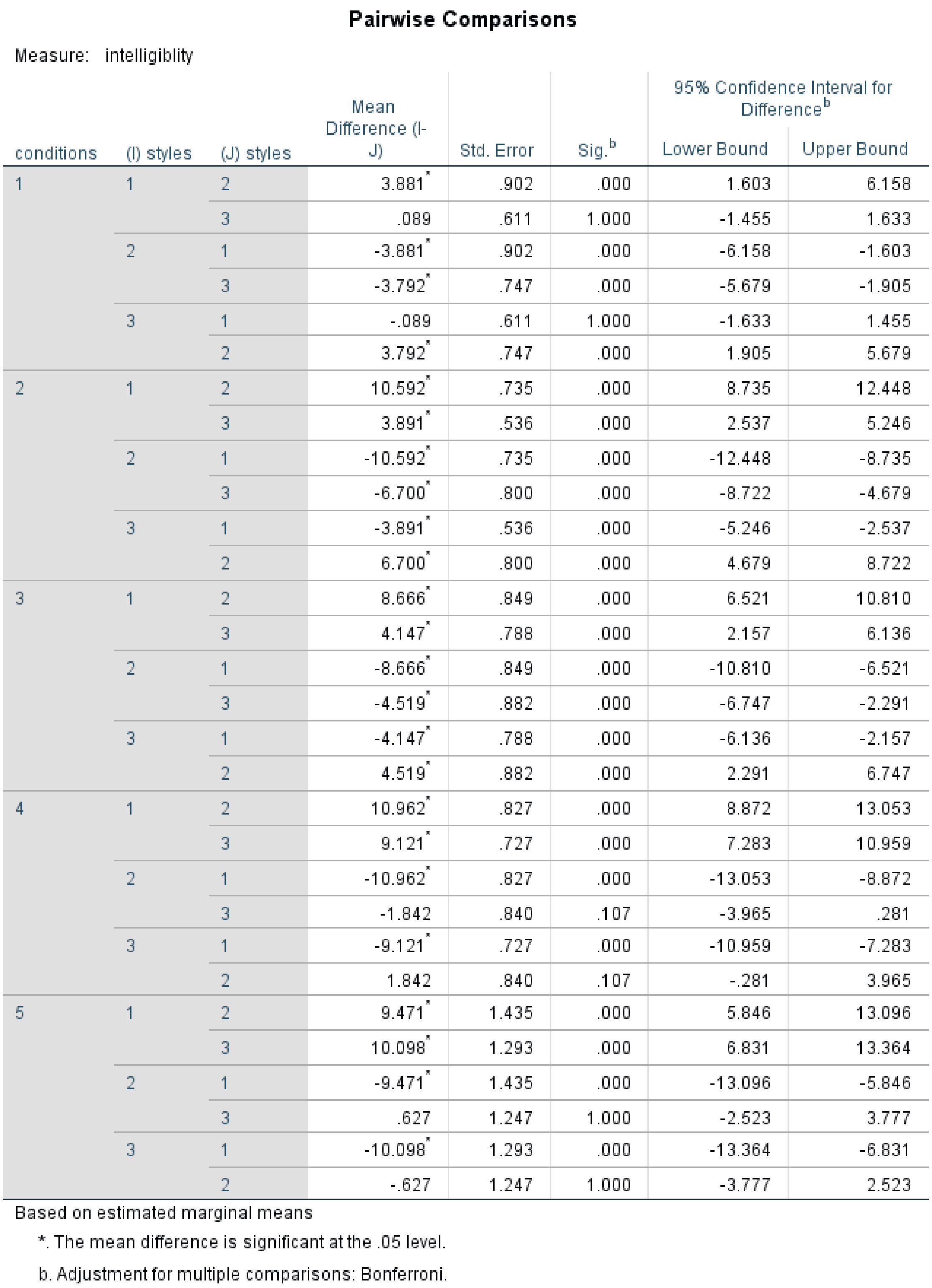
b. Adjustment for multiple comparisons: Bonferroni.



Estimates

Measure: intelligiblity

styles	conditions	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
1	1	69.743	1.525	66.637	72.850
	2	69.561	1.246	67.022	72.099
	3	67.896	1.457	64.929	70.864
	4	65.485	1.499	62.431	68.539
	5	56.199	1.911	52.306	60.091
2	1	65.863	1.409	62.994	68.732
	2	58.969	.996	56.940	60.998
	3	59.231	1.155	56.878	61.583
	4	54.522	1.361	51.750	57.295
	5	46.728	1.777	43.108	50.348
3	1	69.654	1.562	66.473	72.836
	2	65.670	1.462	62.693	68.647
	3	63.750	1.478	60.739	66.760
	4	56.364	1.434	53.443	59.286
	5	46.101	1.705	42.629	49.573



Pairwise Comparisons

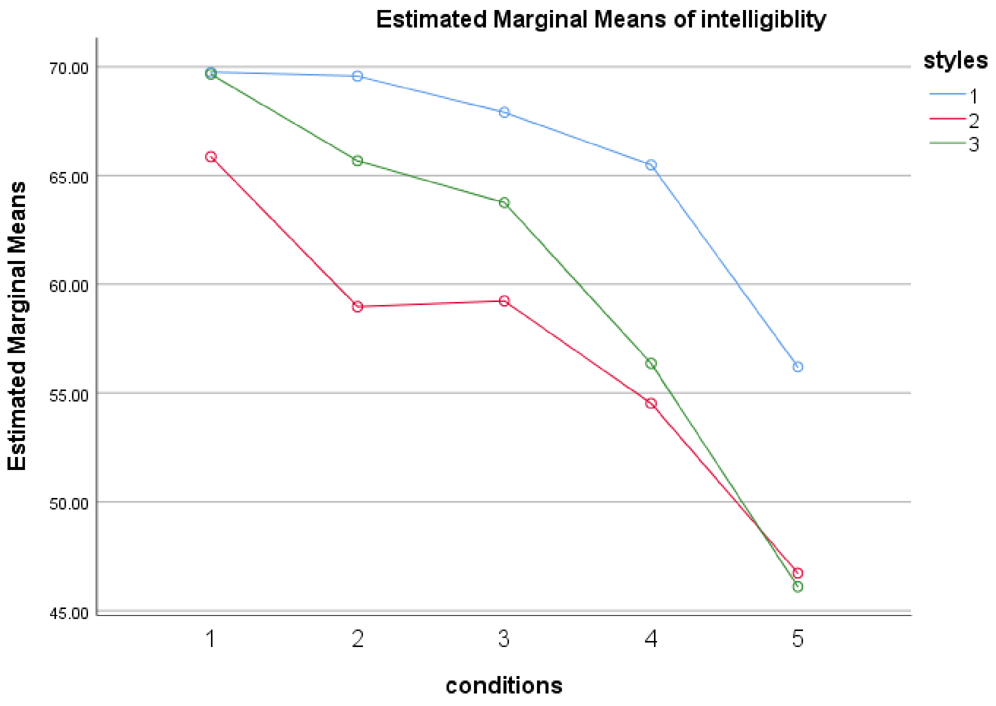
Measure: intelligiblity

conditions	(I) styles	(J) styles	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
						Lower Bound	Upper Bound
1	1	2	3.881*	.902	.000	1.603	6.158
		3	.089	.611	1.000	-1.455	1.633
	2	1	-3.881*	.902	.000	-6.158	-1.603
		3	-3.792*	.747	.000	-5.679	-1.905
	3	1	-.089	.611	1.000	-1.633	1.455
		2	3.792*	.747	.000	1.905	5.679
2	1	2	10.592*	.735	.000	8.735	12.448
		3	3.891*	.536	.000	2.537	5.246
	2	1	-10.592*	.735	.000	-12.448	-8.735
		3	-6.700*	.800	.000	-8.722	-4.679
	3	1	-3.891*	.536	.000	-5.246	-2.537
		2	6.700*	.800	.000	4.679	8.722
3	1	2	8.666*	.849	.000	6.521	10.810
		3	4.147*	.788	.000	2.157	6.136
	2	1	-8.666*	.849	.000	-10.810	-6.521
		3	-4.519*	.882	.000	-6.747	-2.291
	3	1	-4.147*	.788	.000	-6.136	-2.157
		2	4.519*	.882	.000	2.291	6.747
4	1	2	10.962*	.827	.000	8.872	13.053
		3	9.121*	.727	.000	7.283	10.959
	2	1	-10.962*	.827	.000	-13.053	-8.872
		3	-1.842	.840	.107	-3.965	.281
	3	1	-9.121*	.727	.000	-10.959	-7.283
		2	1.842	.840	.107	-.281	3.965
5	1	2	9.471*	1.435	.000	5.846	13.096
		3	10.098*	1.293	.000	6.831	13.364
	2	1	-9.471*	1.435	.000	-13.096	-5.846
		3	.627	1.247	1.000	-2.523	3.777
	3	1	-10.098*	1.293	.000	-13.364	-6.831
		2	-.627	1.247	1.000	-3.777	2.523

Based on estimated marginal means

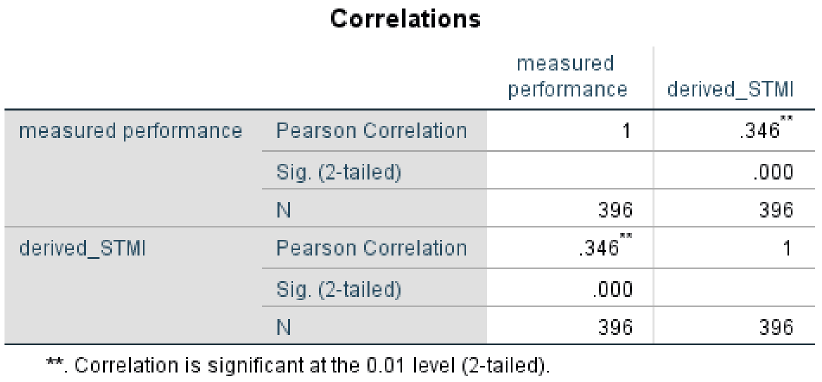
*. The mean difference is significant at the .05 level.

b. Adjustment for multiple comparisons: Bonferroni.



Appendix D.3 Outputs of Experiment 2

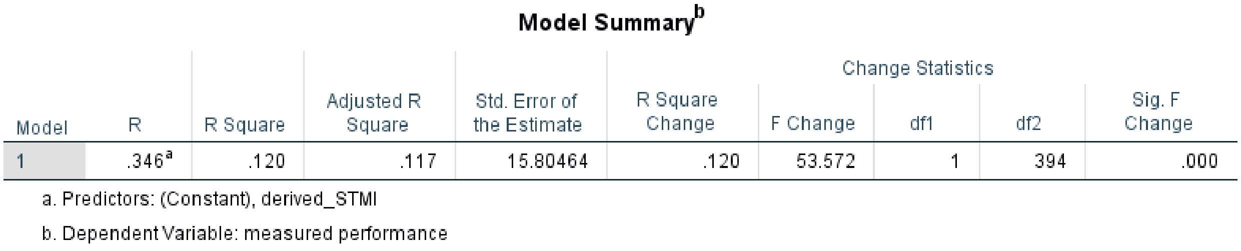
Appendix D.3.1 Overall Speech Intelligibility



Correlations

		measured performance	derived_STMI
measured performance	Pearson Correlation	1	.346**
	Sig. (2-tailed)		.000
	N	396	396
derived_STMI	Pearson Correlation	.346**	1
	Sig. (2-tailed)	.000	
	N	396	396

**. Correlation is significant at the 0.01 level (2-tailed).

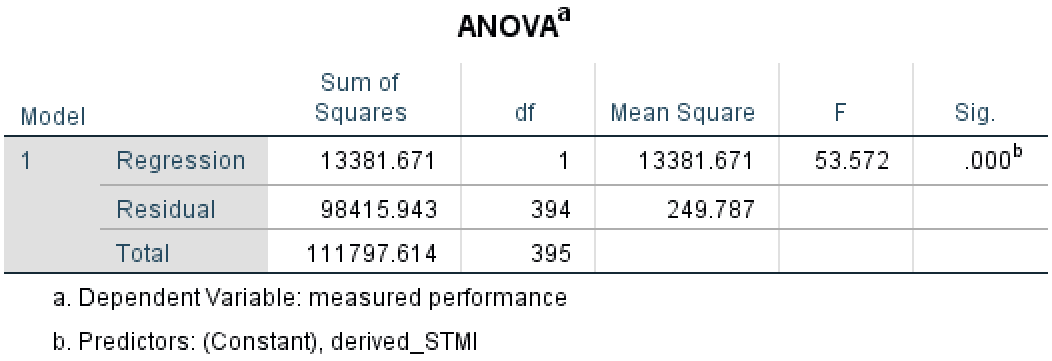


Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.346[a]	.120	.117	15.80464	.120	53.572	1	394	.000

a. Predictors: (Constant), derived_STMI

b. Dependent Variable: measured performance

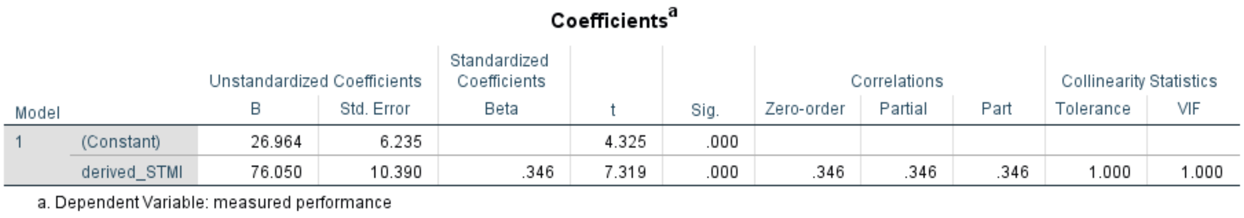


ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	13381.671	1	13381.671	53.572	.000[b]
	Residual	98415.943	394	249.787		
	Total	111797.614	395			

a. Dependent Variable: measured performance

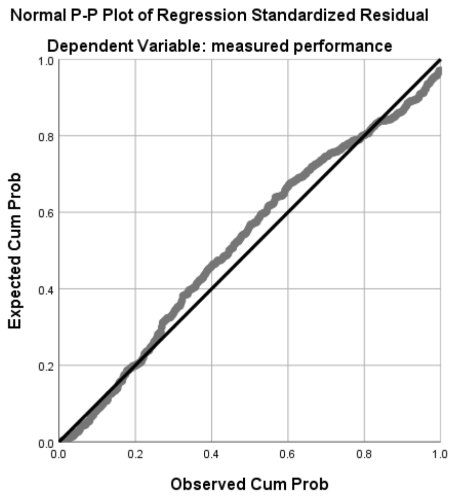
b. Predictors: (Constant), derived_STMI

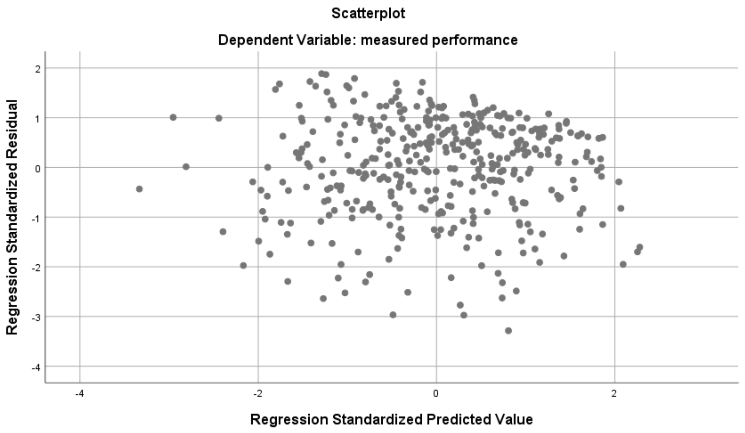


Coefficients[a]

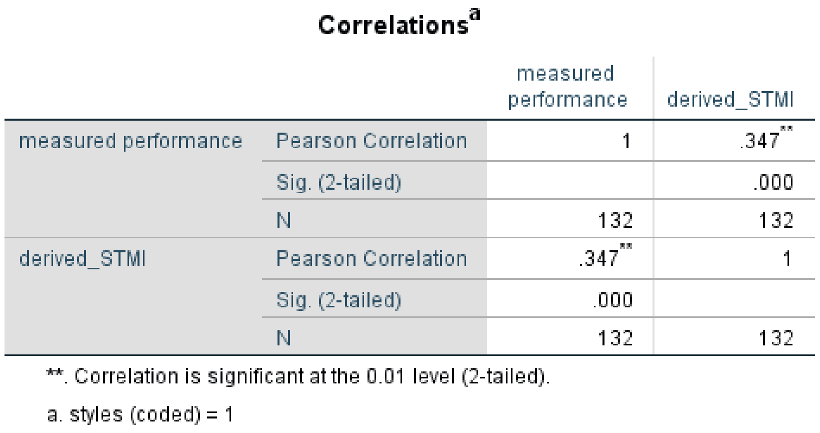
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	26.964	6.235		4.325	.000					
	derived_STMI	76.050	10.390	.346	7.319	.000	.346	.346	.346	1.000	1.000

a. Dependent Variable: measured performance





Appendix D.3.2 Clear Speech Intelligibility

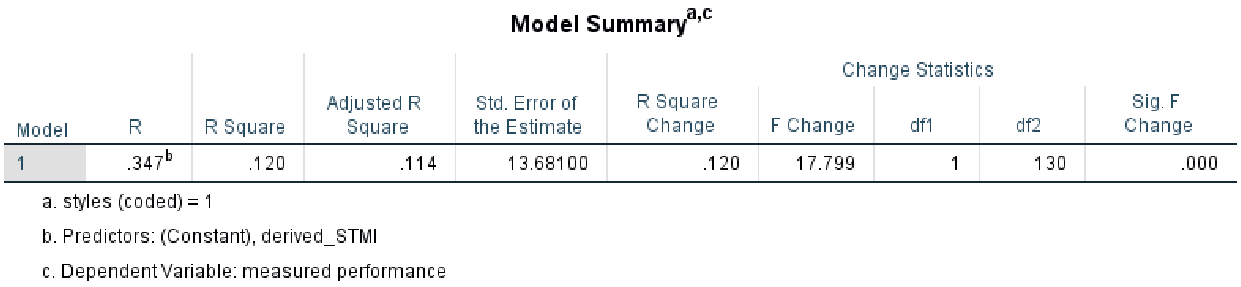


Correlations[a]

		measured performance	derived_STMI
measured performance	Pearson Correlation	1	.347**
	Sig. (2-tailed)		.000
	N	132	132
derived_STMI	Pearson Correlation	.347**	1
	Sig. (2-tailed)	.000	
	N	132	132

**. Correlation is significant at the 0.01 level (2-tailed).

a. styles (coded) = 1



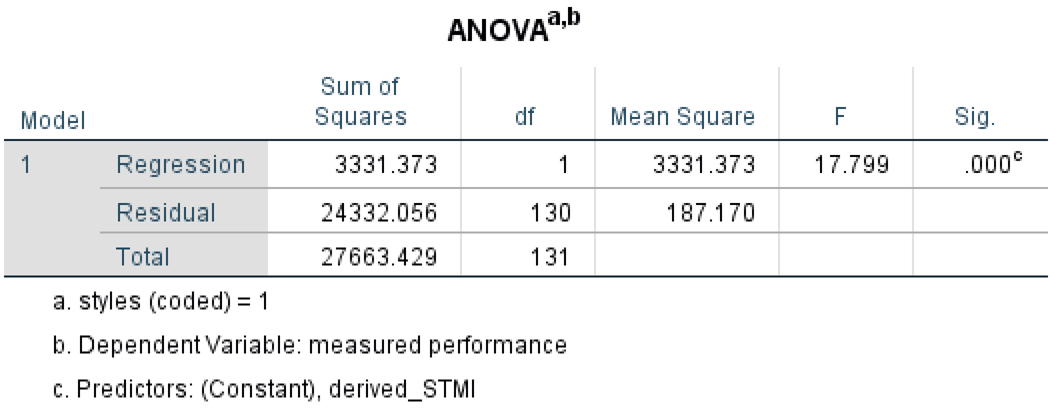
Model Summary[a,c]

					Change Statistics				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.347[b]	.120	.114	13.68100	.120	17.799	1	130	.000

a. styles (coded) = 1

b. Predictors: (Constant), derived_STMI

c. Dependent Variable: measured performance



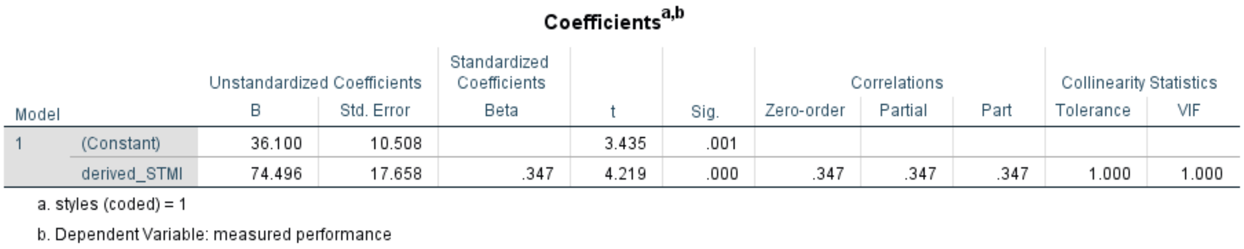
ANOVA[a,b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	3331.373	1	3331.373	17.799	.000[c]
	Residual	24332.056	130	187.170		
	Total	27663.429	131			

a. styles (coded) = 1

b. Dependent Variable: measured performance

c. Predictors: (Constant), derived_STMI

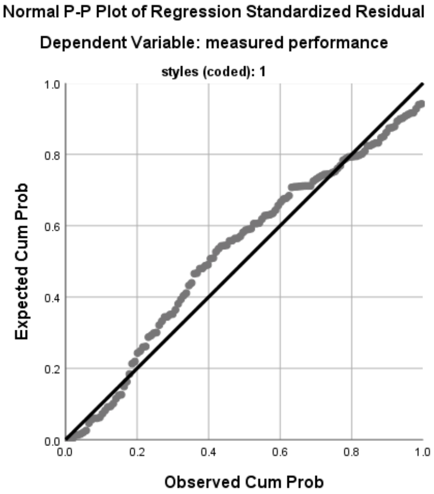


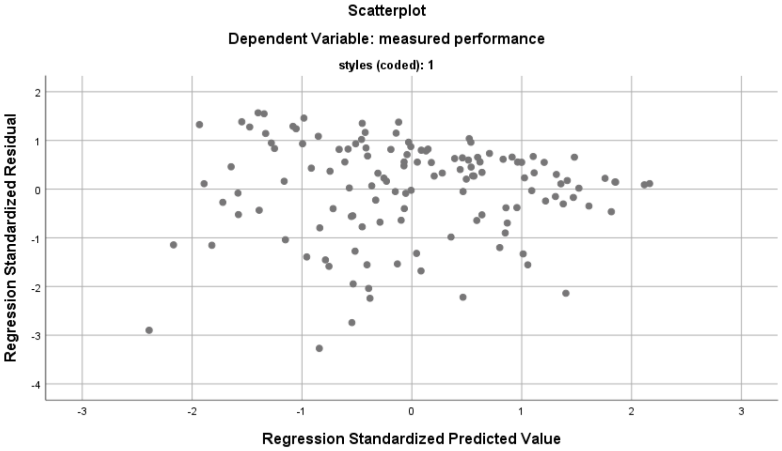
Coefficients[a,b]

		Unstandardized Coefficients		Standardized Coefficients			Correlations			Collinearity Statistics	
Model		B	Std. Error	Beta	t	Sig.	Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	36.100	10.508		3.435	.001					
	derived_STMI	74.496	17.658	.347	4.219	.000	.347	.347	.347	1.000	1.000

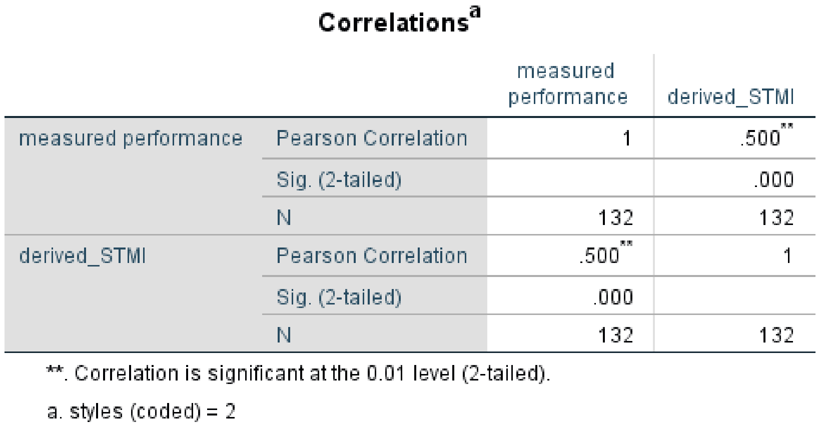
a. styles (coded) = 1

b. Dependent Variable: measured performance





Appendix D.3.3 Lab Conversational Speech Intelligibility

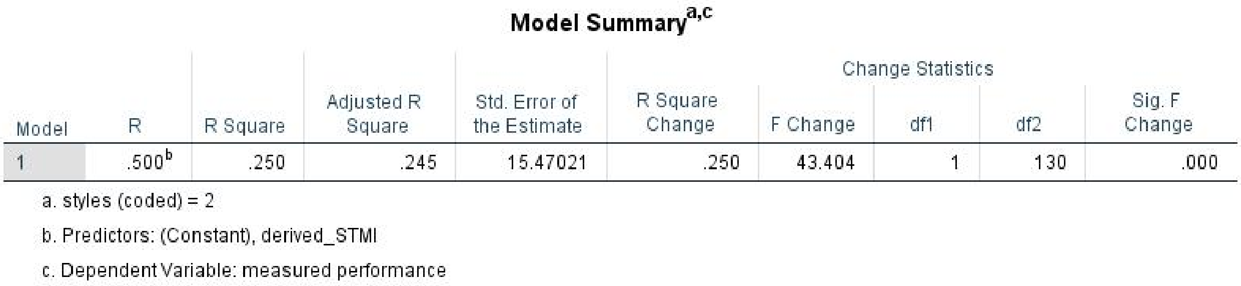


Correlations[a]

		measured performance	derived_STMI
measured performance	Pearson Correlation	1	.500**
	Sig. (2-tailed)		.000
	N	132	132
derived_STMI	Pearson Correlation	.500**	1
	Sig. (2-tailed)	.000	
	N	132	132

**. Correlation is significant at the 0.01 level (2-tailed).

a. styles (coded) = 2



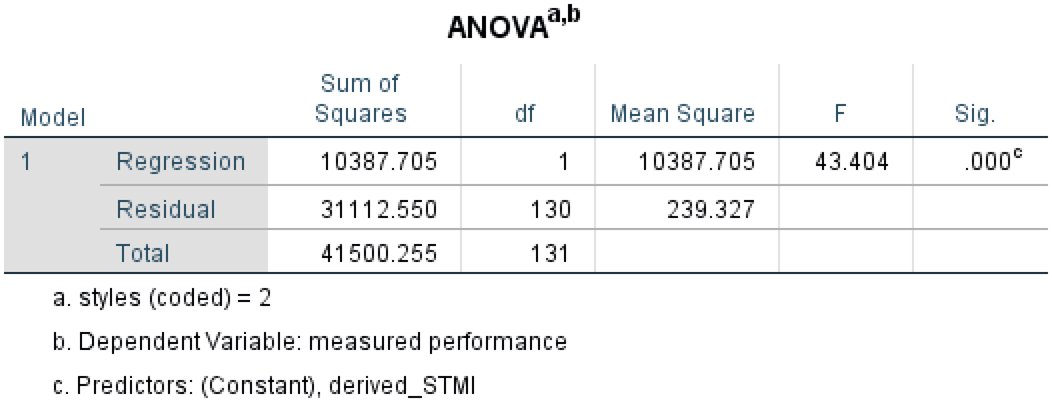
Model Summary[a,c]

					Change Statistics				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.500[b]	.250	.245	15.47021	.250	43.404	1	130	.000

a. styles (coded) = 2

b. Predictors: (Constant), derived_STMI

c. Dependent Variable: measured performance



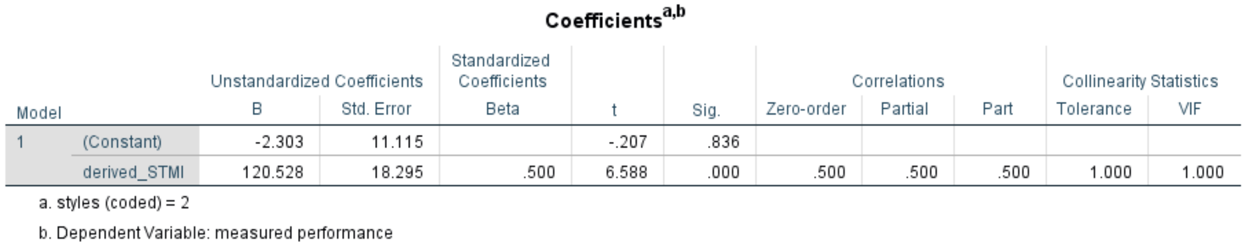
ANOVA[a,b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	10387.705	1	10387.705	43.404	.000[c]
	Residual	31112.550	130	239.327		
	Total	41500.255	131			

a. styles (coded) = 2

b. Dependent Variable: measured performance

c. Predictors: (Constant), derived_STMI

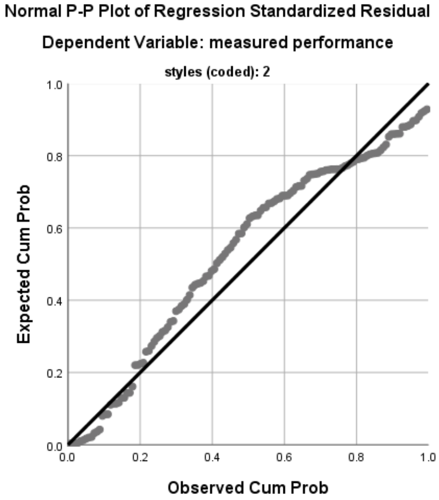


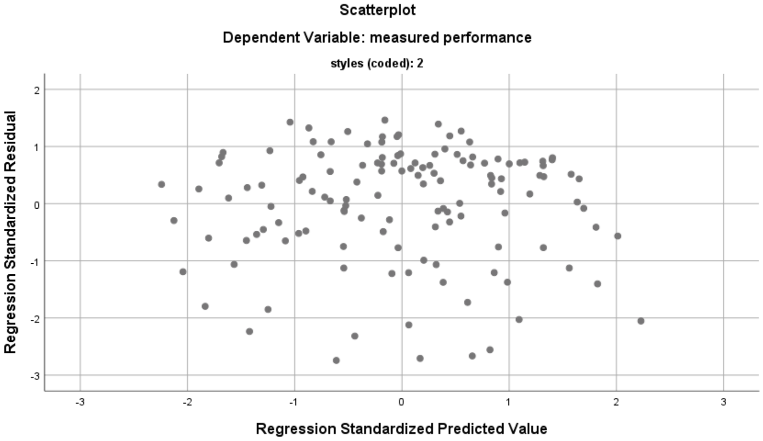
Coefficients[a,b]

Model		Unstandardized Coefficients		Standardized Coefficients			Correlations			Collinearity Statistics	
		B	Std. Error	Beta	t	Sig.	Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	-2.303	11.115		-.207	.836					
	derived_STMI	120.528	18.295	.500	6.588	.000	.500	.500	.500	1.000	1.000

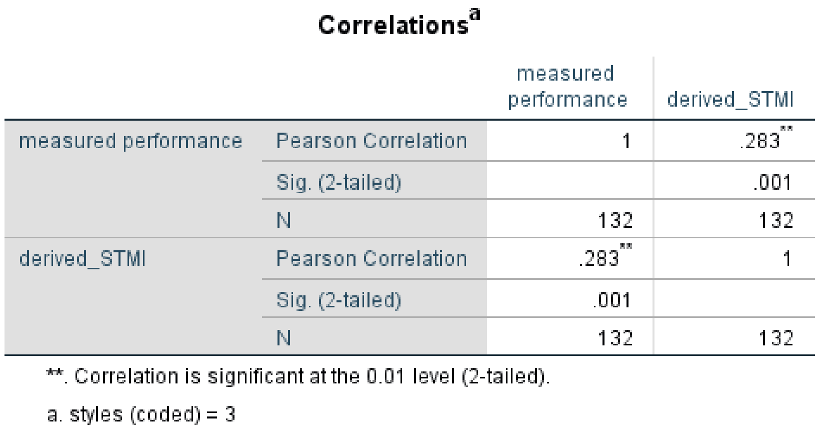
a. styles (coded) = 2

b. Dependent Variable: measured performance





Appendix D.3.4 Natural Conversational Speech Intelligibility

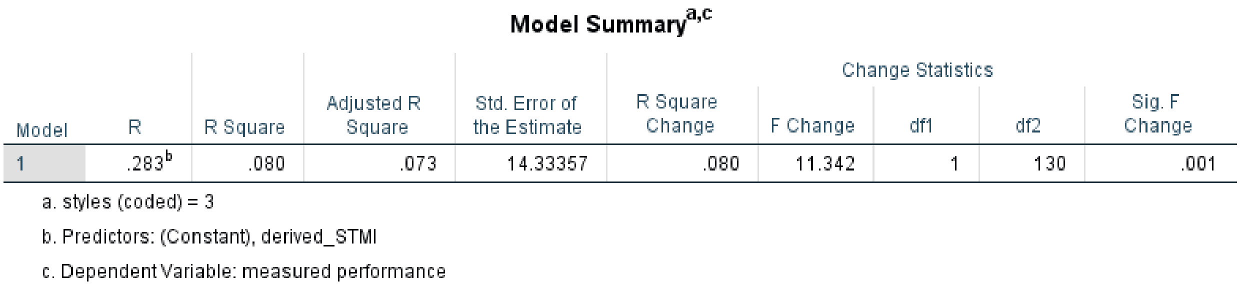


Correlations[a]

		measured performance	derived_STMI
measured performance	Pearson Correlation	1	.283**
	Sig. (2-tailed)		.001
	N	132	132
derived_STMI	Pearson Correlation	.283**	1
	Sig. (2-tailed)	.001	
	N	132	132

**. Correlation is significant at the 0.01 level (2-tailed).

a. styles (coded) = 3



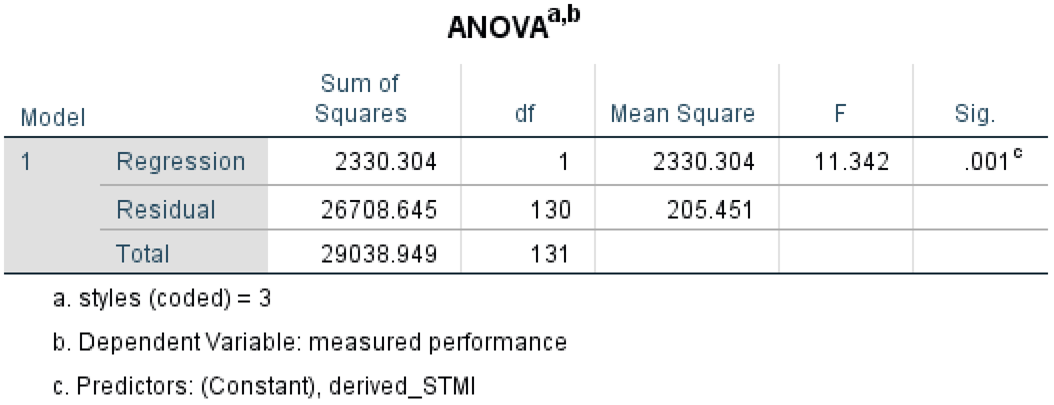
Model Summary[a,c]

					Change Statistics				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.283[b]	.080	.073	14.33357	.080	11.342	1	130	.001

a. styles (coded) = 3

b. Predictors: (Constant), derived_STMI

c. Dependent Variable: measured performance



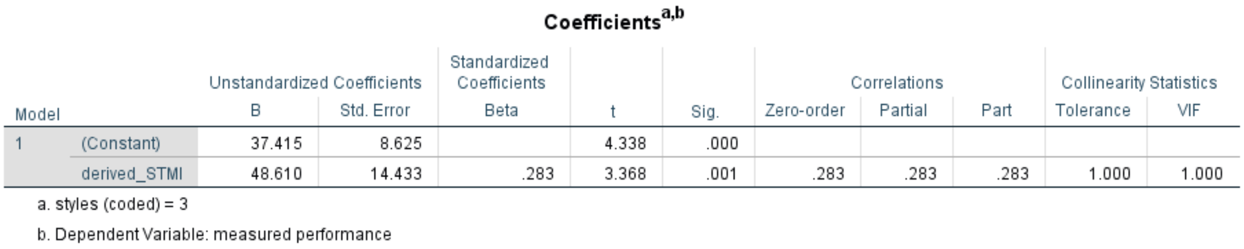
ANOVA[a,b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2330.304	1	2330.304	11.342	.001[c]
	Residual	26708.645	130	205.451		
	Total	29038.949	131			

a. styles (coded) = 3

b. Dependent Variable: measured performance

c. Predictors: (Constant), derived_STMI

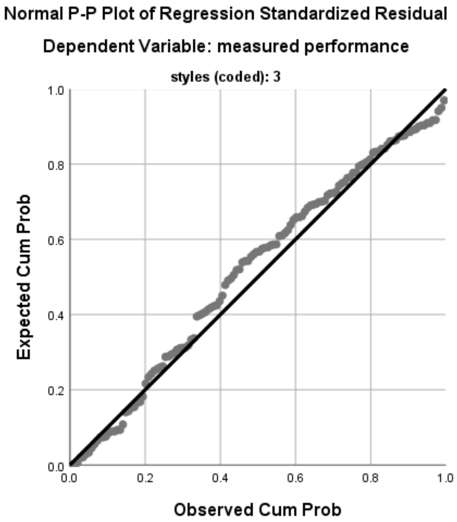


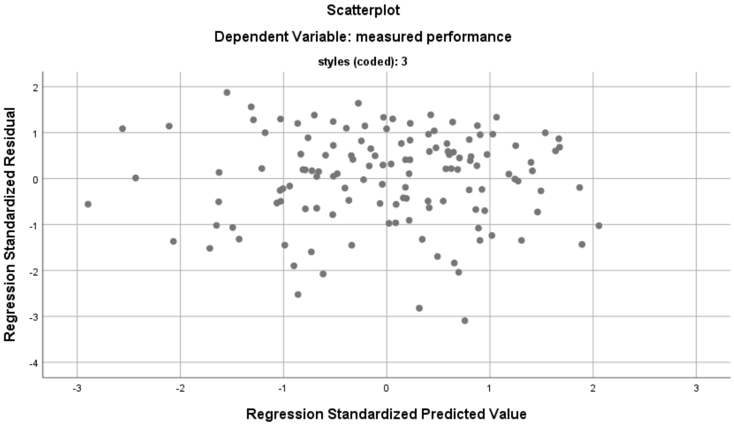
Coefficients[a,b]

		Unstandardized Coefficients		Standardized Coefficients			Correlations			Collinearity Statistics	
Model		B	Std. Error	Beta	t	Sig.	Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	37.415	8.625		4.338	.000					
	derived_STMI	48.610	14.433	.283	3.368	.001	.283	.283	.283	1.000	1.000

a. styles (coded) = 3

b. Dependent Variable: measured performance


Normal P-P Plot of Regression Standardized Residual
Dependent Variable: measured performance
styles (coded): 3
Expected Cum Prob
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
Observed Cum Prob


Scatterplot
Dependent Variable: measured performance
styles (coded): 3
Regression Standardized Residual
2
1
0
-1
-2
-3
-4
-3
-2
-1
0
1
2
3
Regression Standardized Predicted Value

Appendix E Sample of STMI Values

A sample of the derived STMI valued for one subject.

Filecode	Styles	STMI at +3	STMI at 0	STMI -3	STMI -6
001	Clear	0.49653	0.5731	0.45046	0.43403
002	Clear	0.4639	0.54432	0.41508	0.39756
003	Clear	0.687	0.73778	0.656	0.64543
004	Clear	0.62103	0.67365	0.5908	0.58107
005	Clear	0.35692	0.46698	0.27892	0.23723
006	Clear	0.59706	0.65643	0.56025	0.54646
007	Clear	0.65891	0.71031	0.62778	0.61764
008	Clear	0.56068	0.62766	0.51802	0.50021
009	Clear	0.58154	0.65302	0.53363	0.5126
010	Clear	0.66068	0.71641	0.62241	0.60329
011	Clear	0.62662	0.69143	0.58311	0.56399
012	Clear	0.56065	0.65005	0.48439	0.42938
013	Clear	0.52392	0.58925	0.48648	0.47478
014	Clear	0.4917	0.57369	0.43889	0.41569
015	Clear	0.61	0.67281	0.57034	0.55487
016	Clear	0.44265	0.52646	0.39428	0.38068
017	Clear	0.50468	0.57834	0.46038	0.44443
018	Clear	0.44191	0.53401	0.38211	0.35519
019	Clear	0.55197	0.6261	0.50118	0.47989
020	Clear	0.47903	0.57196	0.41082	0.37522
021	Clear	0.62281	0.67975	0.5865	0.57161
022	Clear	0.53833	0.62328	0.47321	0.43609
023	Clear	0.63181	0.69417	0.58828	0.56655
024	Clear	0.62948	0.70037	0.57402	0.54274
025	Clear	0.52476	0.59062	0.49136	0.48412
026	Clear	0.5001	0.58469	0.44069	0.41043
027	Clear	0.50288	0.58242	0.45616	0.43873
028	Clear	0.48703	0.56557	0.43429	0.40851
029	Clear	0.52866	0.59965	0.48805	0.47558
030	Clear	0.60503	0.66513	0.56899	0.5551
031	Clear	0.45537	0.56756	0.3626	0.30119
032	Clear	0.61541	0.68454	0.56362	0.53424
033	Clear	0.51655	0.61908	0.42515	0.35724
034	Clear	0.52659	0.61273	0.46214	0.42244
035	Clear	0.56252	0.63324	0.51993	0.50666

036	Clear	0.37102	0.4836	0.29239	0.25202
037	Clear	0.62277	0.68616	0.58237	0.56546
038	Clear	0.55107	0.61694	0.51846	0.51493
039	Clear	0.62481	0.68031	0.59464	0.58752
040	Clear	0.43953	0.54179	0.36477	0.32264
041	Clear	0.59691	0.66269	0.55629	0.5402
042	Clear	0.52214	0.61685	0.44734	0.39883
043	Clear	0.5963	0.67857	0.53201	0.49114
044	Clear	0.57159	0.65007	0.5158	0.48619
045	Clear	0.67897	0.73642	0.63467	0.60625
046	Clear	0.51661	0.60947	0.44824	0.41014
047	Clear	0.64751	0.71684	0.58641	0.53801
048	Clear	0.6831	0.73453	0.64924	0.63125
049	Clear	0.33694	0.45538	0.25068	0.20169
050	Clear	0.48374	0.58485	0.41238	0.36975
051	Clear	0.51284	0.59952	0.45558	0.43133
052	Clear	0.60124	0.66603	0.56303	0.54999
053	Clear	0.52868	0.59998	0.48489	0.46815
054	Clear	0.52845	0.60193	0.48314	0.4672
055	Clear	0.50121	0.57251	0.45975	0.4468
056	Clear	0.59996	0.66662	0.55272	0.52651
057	Clear	0.46707	0.55105	0.4159	0.39503
058	Clear	0.57564	0.63449	0.54084	0.52931
059	Clear	0.36186	0.48179	0.27389	0.22508
060	Clear	0.47021	0.56574	0.40188	0.36454
061	Clear	0.51689	0.59543	0.46261	0.43442
062	Clear	0.36478	0.48001	0.28307	0.24151
063	Clear	0.622	0.68838	0.57127	0.541
064	Clear	0.64881	0.71417	0.59415	0.55432
065	Clear	0.38841	0.49799	0.30756	0.26209
066	Clear	0.4525	0.55	0.38698	0.35696
067	Clear	0.46888	0.56776	0.40274	0.37296
068	Clear	0.53194	0.6007	0.49333	0.48054
069	Clear	0.56172	0.63565	0.51017	0.48502
070	Clear	0.62873	0.69078	0.58661	0.56465
071	Clear	0.55665	0.6391	0.49651	0.46325
072	Clear	0.58468	0.65984	0.53452	0.51334
073	Clear	0.60527	0.67733	0.55555	0.52939
074	Clear	0.55399	0.63445	0.4972	0.46804
075	Conv	0.59912	0.67364	0.54581	0.51663
076	Conv	0.49046	0.58009	0.42523	0.38761
077	Conv	0.48899	0.56147	0.44993	0.44248
078	Conv	0.59493	0.66213	0.54962	0.52785

079	Conv	0.53037	0.61087	0.47918	0.45688
080	Conv	0.66748	0.72459	0.62826	0.60696
081	Conv	0.55731	0.62638	0.51633	0.50192
082	Conv	0.63116	0.69395	0.58851	0.56737
083	Conv	0.38092	0.48543	0.30382	0.25971
084	Conv	0.68356	0.73597	0.64667	0.62693
085	Conv	0.71808	0.76967	0.67967	0.65694
086	Conv	0.4773	0.57289	0.4083	0.36909
087	Conv	0.45773	0.53851	0.41431	0.4016
088	Conv	0.39668	0.50364	0.32449	0.28952
089	Conv	0.56283	0.63267	0.52082	0.50551
090	Conv	0.58147	0.64739	0.53985	0.52284
091	Conv	0.43301	0.54006	0.35703	0.32033
092	Conv	0.58808	0.6615	0.53394	0.50211
093	Conv	0.44903	0.54908	0.37653	0.33553
094	Conv	0.64382	0.69904	0.6108	0.59877
095	Conv	0.61102	0.67561	0.57176	0.55612
096	Conv	0.70169	0.75345	0.66615	0.64806
097	Conv	0.64188	0.6997	0.60356	0.58559
098	Conv	0.36071	0.47872	0.2738	0.22358
099	Conv	0.69775	0.74666	0.66498	0.65
100	Conv	0.60706	0.67133	0.56181	0.53776
101	Conv	0.66833	0.7209	0.6339	0.61864
102	Conv	0.59626	0.66704	0.54633	0.519
103	Conv	0.70458	0.75541	0.66963	0.65172
104	Conv	0.65243	0.71256	0.61004	0.58752
105	Conv	0.45026	0.54644	0.38011	0.34008
106	Conv	0.66539	0.72369	0.62399	0.60157
107	Conv	0.68041	0.73659	0.64098	0.62108
108	Conv	0.5699	0.64277	0.52273	0.50157
109	Conv	0.42096	0.52456	0.34499	0.30206
110	Conv	0.59112	0.66064	0.54623	0.52508
111	Conv	0.66056	0.71372	0.62603	0.61012
112	Conv	0.38941	0.49154	0.31988	0.287
113	Conv	0.57785	0.65092	0.52839	0.504
114	Conv	0.50964	0.59315	0.44809	0.41184
115	Conv	0.57649	0.64547	0.53237	0.51321
116	Conv	0.66984	0.73321	0.62712	0.60522
117	Conv	0.53239	0.60776	0.48622	0.46712
118	Conv	0.61849	0.68816	0.56843	0.54199
119	Conv	0.30897	0.42615	0.22342	0.1763
120	Conv	0.55938	0.63876	0.4989	0.46173
121	Lab	0.67156	0.72383	0.63589	0.61807

122	Lab	0.54188	0.62515	0.47709	0.4374
123	Lab	0.66925	0.72513	0.63157	0.61557
124	Lab	0.62575	0.68427	0.58959	0.57592
125	Lab	0.64971	0.70676	0.61417	0.60211
126	Lab	0.64513	0.70009	0.6121	0.60041
127	Lab	0.64285	0.69804	0.60913	0.59746
128	Lab	0.64386	0.70653	0.59965	0.57507
129	Lab	0.54982	0.62788	0.49647	0.46984
130	Lab	0.58918	0.66412	0.53477	0.50645
131	Lab	0.60296	0.67697	0.5472	0.51278
132	Lab	0.62957	0.69115	0.58645	0.56297
133	Lab	0.56568	0.64108	0.51206	0.48359
134	Lab	0.37021	0.48412	0.29133	0.25219
135	Lab	0.52595	0.60017	0.47978	0.46309
136	Lab	0.59127	0.65175	0.5536	0.53682
137	Lab	0.48007	0.57735	0.41254	0.37937
138	Lab	0.51175	0.5952	0.45419	0.42419
139	Lab	0.3895	0.49755	0.31553	0.28073
140	Lab	0.48722	0.57416	0.43376	0.41673
141	Lab	0.5389	0.62221	0.47882	0.44591
142	Lab	0.65994	0.71964	0.6174	0.59446
143	Lab	0.40666	0.5148	0.32656	0.27912
144	Lab	0.52143	0.60185	0.46454	0.43511
145	Lab	0.47991	0.58236	0.40539	0.3676
146	Lab	0.48771	0.56472	0.44554	0.43418
147	Lab	0.56433	0.63441	0.51808	0.498
148	Lab	0.56328	0.64371	0.50292	0.46746
149	Lab	0.636	0.70079	0.58951	0.56314
150	Lab	0.56274	0.64828	0.49587	0.45213
151	Lab	0.60073	0.67105	0.55429	0.53191
152	Lab	0.49626	0.58867	0.43113	0.39749
153	Lab	0.49466	0.57925	0.43623	0.40767
154	Lab	0.55361	0.62664	0.50753	0.48742
155	Lab	0.46895	0.56394	0.39597	0.35065
156	Lab	0.60885	0.68314	0.55255	0.51751
157	Lab	0.43782	0.54249	0.35936	0.31264
158	Lab	0.45697	0.56061	0.3849	0.3515
159	Lab	0.57929	0.65036	0.53237	0.51076
160	Lab	0.48056	0.57139	0.41306	0.37569
161	Lab	0.53539	0.61712	0.47815	0.44487
162	Lab	0.66479	0.72828	0.6178	0.59013
163	Lab	0.62813	0.68761	0.58894	0.572
164	Lab	0.37671	0.49968	0.27957	0.21824

165	Lab	0.67399	0.73678	0.62604	0.59613
166	Lab	0.52785	0.61907	0.45848	0.41662
167	Lab	0.58259	0.64576	0.54335	0.52972
168	Lab	0.58906	0.66496	0.54123	0.52274
169	Lab	0.44481	0.54257	0.38089	0.35362
170	Lab	0.60763	0.66953	0.57056	0.55784
171	Lab	0.42282	0.52737	0.34624	0.30281
172	Lab	0.54577	0.62465	0.48913	0.45766
173	Lab	0.5617	0.64088	0.50534	0.47726
174	Lab	0.49489	0.5827	0.43107	0.3958
175	Lab	0.70051	0.75175	0.66665	0.65152
176	Lab	0.59027	0.66584	0.53467	0.5021
177	Lab	0.63263	0.70036	0.58621	0.56468
178	Lab	0.45483	0.54796	0.39283	0.36459
179	Lab	0.71832	0.76773	0.68374	0.66369
180	Lab	0.67596	0.73148	0.63803	0.61905
181	Lab	0.50664	0.58934	0.44828	0.41677
182	Lab	0.30682	0.43424	0.21351	0.16181
183	Lab	0.67572	0.73782	0.63144	0.60778

Bibliography

Abel, S. M., Alberti, P. W., Haythornthwaite, C., & Riko, K. (1982). Speech intelligibility in noise: effects of fluency and hearing protector type. *J Acoust Soc Am, 71*(3), 708-715. doi:10.1121/1.387547

Amano-Kusumoto, A., Hosom, J. P., Kain, A., & Aronoff, J. M. (2014). Determining the relevance of different aspects of formant contours to intelligibility. *Speech Commun, 59*, 1-9. doi:10.1016/j.specom.2013.12.001

ANSI. (1969). Methods for the Calculation of the Articulation Index. In. American National Standards Institute: New York.

ANSI. (1997). Methods for the calculation of the speech intelligibility index. In. American National Standards Institute: New York.

Beranek, L. (1947). The design of speech communication systems. *Proc. ICE, 35*, 880-890.

Bradlow, A. R., Kraus, N., & Hayes, E. (2003). Speaking clearly for children with learning disabilities: sentence perception in noise. *J Speech Lang Hear Res, 46*(1), 80-97. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12647890

http://jslhr.pubs.asha.org/data/Journals/JSLHR/929269/jslhr_46_1_80.pdf

Brungart, D. S., Barrett, M. E., Cohen, J. I., Fodor, C., Yancey, C. M., & Gordon-Salant, S. (2020). Objective Assessment of Speech Intelligibility in Crowded Public Spaces. *Ear and hearing, 41 Suppl 1*, 68S-78S. doi:10.1097/AUD.0000000000000943

Byrd, D., & Tan, C. C. (1996). Saying consonant clusters quickly. *Journal of Phonetics, 24*(2), 263-263. doi:10.1006/jpho.1996.0014

Chi, T., Gao, Y., Guyton, M. C., Ru, P., & Shamma, S. (1999). Spectro-temporal modulation transfer functions and speech intelligibility. *Journal of the Acoustical Society of America, 106*(5), 2719-2732. doi:10.1121/1.428100

Ching, T. Y., Johnson, E. E., Hou, S., Dillon, H., Zhang, V., Burns, L., . . . Flynn, C. (2013). A comparison of NAL and DSL prescriptive methods for paediatric hearing-aid fitting: predicted speech intelligibility and loudness. *Int J Audiol, 52 Suppl 2*, S29-38. doi:10.3109/14992027.2013.765041

Cox, R. M., Alexander, G. C., & Gilmore, C. (1987). Development of the Connected Speech Test (CST). *Ear and Hearing, 8*(5 SUPPLEMENT), 119s-119s. doi:10.1097/00003446-198710001-00010

Doyle, P. J., McNeil, M. R., Park, G., Goda, A., Rubenstein, E., Spencer, K., . . . Szwarc, L. (2000). Linguistic validation of four parallel forms of a story retelling procedure. *Aphasiology, 14*(5-6), 537-549. doi:10.1080/026870300401306

Durisala, N., Prakash, S. G., Nambi, A., & Batra, R. (2011). Intelligibility and acoustic characteristics of clear and conversational speech in telugu (a South Indian dravidian language). *Indian J Otolaryngol Head Neck Surg, 63*(2), 165-171. doi:10.1007/s12070-011-0241-7

Durrant, J. D., & Feth, L. L. (2012). *Hearing Science: A Foundational Approach*: Pearson.

Egan, J., & Weiner, F. (1949). On the intelligibility of bands of speech in noise. *Journal of the Acoustical Society of America, 18*, 435-441.

Elhilali, M., Chi, T., & Shamma, S. A. (2003). A spectro-temporal modulation index (STMI) for assessment of speech intelligibility. *Speech Communication, 41*(2), 331-348. doi:https://doi.org/10.1016/S0167-6393(02)00134-6

Ferguson, S. H. (2004). Talker differences in clear and conversational speech: Vowel intelligibility for normal-hearing listeners. *Journal of the Acoustical Society of America*, *116*(4 I), 2365-2373. doi:10.1121/1.1788730

Ferguson, S. H. (2012). Talker differences in clear and conversational speech: vowel intelligibility for older adults with hearing loss. *J Speech Lang Hear Res*, *55*(3), 779-790. doi:10.1044/1092-4388(2011/10-0342)

Ferguson, S. H., & Kewley-Port, D. (2007). Talker differences in clear and conversational speech: acoustic characteristics of vowels. *J Speech Lang Hear Res*, *50*(5), 1241-1255. doi:10.1044/1092-4388(2007/087)

Fletcher, H. (1952). The perception of sounds by deafened persons. *Journal of the Acoustical Society of America*, *24*, 490-497.

Fletcher, H. (1953). *Speech and hearing in communication*. New York, NY: Van Nostrand.

Fletcher, H., & Galt, R. (1950). The perception of speech and its relation to telephony. *Journal of the Acoustical Society of America*, *22*, 89-151.

French, N. R., & Steinberg, J. C. (1947). Factors Governing the Intelligibility of Speech Sounds. *Journal of the Acoustical Society of America*, *19*(1), 90-119. doi:10.1121/1.1916407

Hedrick, M. S., & Younger, M. S. (2007). Perceptual Weighting of Stop Consonant Cues by Normal and Impaired Listeners in Reverberation Versus Noise. *Journal of Speech, Language, and Hearing Research*, *50*(2), 254-269. doi:10.1044/1092-4388(2007/019)

Howell, P., & Kadi-Hanifi, K. (1991). Comparison of prosodic properties between read and spontaneous speech material. *Speech Communication*, *10*(2), 163-169. doi:http://dx.doi.org/10.1016/0167-6393(91)90039-V

Hulsch, D. (1975). Adult age differences in retrieval: trace dependent and cue dependent forgetting. *Developmental Psychology, 11*, 197-201.

Humes, L., Dirks, D., Bell, T., Ahlstrom, C., & Kincaid, G. (1986). Application of the Articulation Index and the Speech Transmission Index to the recognition of speech by normal-hearing and hearing-impaired listeners. *Journal of Speech and Hearing Research, 29*, 447-462.

Jensen, J., & Taal, C. H. (2014). Speech intelligibility prediction based on mutual information. *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP), 22*(2), 430-440.

Kain, A., Amano-Kusumoto, A., & Hosom, J.-P. (2008). Hybridizing conversational and clear speech to determine the degree of contribution of acoustic features to intelligibility. *Journal of the Acoustical Society of America, 124*(4), 2308-2319. doi:10.1121/1.2967844

Kalikow, D. N., Stevens, K. N., & Elliott, L. L. (1977). Development of a Test of Speech-Intelligibility in Noise Using Sentence Materials with Controlled Word Predictability. *Journal of the Acoustical Society of America, 61*(5), 1337-1351. doi:Doi 10.1121/1.381436

Kates, J. M., & Arehart, K. H. (2005). Coherence and the speech intelligibility index. *The Journal of the Acoustical Society of America, 117*(4), 2224-2237. doi:doi:http://dx.doi.org/10.1121/1.1862575

Killion, Mueller, G., Pavlovic, C., & Humes, L. (1993). A is for audibility. *Hearing Journal, 43*(4), 29-32.

Krause, J. C., & Braida, L. D. (1995). The effects of speaking rate on the intelligibility of speech for various speaking modes. *The Journal of the Acoustical Society of America, 98*(5), 2982-2982. doi:10.1121/1.413900

Krause, J. C., & Braida, L. D. (2002). Investigating alternative forms of clear speech: the effects of speaking rate and speaking mode on intelligibility. *The Journal of the Acoustical Society of America, 112*(5 Pt 1), 2165-2172. doi:10.1121/1.1509432

Krause, J. C., & Braida, L. D. (2004). Acoustic properties of naturally produced clear speech at normal speaking rates. *J Acoust Soc Am, 115*(1), 362-378. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/14759028

Krause, J. C., & Braida, L. D. (2009). Evaluating the Role of Spectral and Envelope Characteristics in the Intelligibility of Clear Speech. *Journal of the Acoustical Society of America, 125*(5), 3346-3357. doi:10.1121/1.3097491

Li, Y., Zhang, G., Kang, H.-y., Liu, S., Han, D., & Fu, Q.-J. (2011). Effects of speaking style on speech intelligibility for Mandarin-speaking cochlear implant users. *The Journal of the Acoustical Society of America, 129*(6), EL242-EL247. doi:10.1121/1.3582148

Liu, S., Del Rio, E., Bradlow, A. R., & Zeng, F.-G. (2004). Clear speech perception in acoustic and electric hearing. *Journal of the Acoustical Society of America, 116*(4 I), 2374-2383. doi:10.1121/1.1787528

Liu, S., & Zeng, F.-G. (2006). Temporal Properties in Clear Speech Perception. *The Journal of the Acoustical Society of America, 120*(1), 424-432. doi:10.1121/1.2208427

Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., & Brian, C. J. M. (2006). Speech Perception Problems of the Hearing Impaired Reflect Inability to Use Temporal Fine Structure. *Proceedings of the National Academy of Sciences of the United States of America, 103*(49), 18866-18869. doi:10.1073/pnas.0607364103

Ludvigsen, C. (1987). Predictions of speech intelligibility for normal-hearing and cochlearly hearing-impaired listeners. *Journal of the Acoustical Society of America, 82,* 1162-1170.

Maniwa, K., Jongman, A., & Wade, T. (2008). Perception of clear fricatives by normal-hearing and simulated hearing-impaired listeners. *Journal of the Acoustical Society of America, 123*(2), 1114-1125. doi:10.1121/1.2821966

Mattys, S. L., Brooks, J., & Cooke, M. (2009). Recognizing speech under a processing load: Dissociating energetic from informational factors. *Cognitive Psychology, 59*(3), 203-243. doi:10.1016/j.cogpsych.2009.04.001

McCreery, R. W., & Stelmachowicz, P. G. (2011). Audibility-based predictions of speech recognition for children and adults with normal hearing. *J Acoust Soc Am, 130*(6), 4070-4081. doi:10.1121/1.3658476

McNeil, M. R., Doyle, P. J., Fossett, T. R. D., Park, G. H., & Goda, A. J. (2001). Reliability and concurrent validity of the information unit scoring metric for the story retelling procedure. *Aphasiology, 15*(10-11), 991-1006. doi:10.1080/02687040143000348

McNeil, M. R., Sung, J. E., Yang, D., Pratt, S. R., Fossett, T. R. D., Doyle, P. J., & Pavelko, S. (2007). Comparing connected language elicitation procedures in persons with aphasia: Concurrent validation of the Story Retell Procedure. *Aphasiology, 21*(6-8), 775-790. doi:10.1080/02687030701189980

Miller, G. (1947). The masking of speech. *Psychological Bulliten, 44*, 105-129.

Nakamura, M., Iwano, K., & Furui, S. (2008). Differences between acoustic characteristics of spontaneous and read speech and their effects on speech recognition performance. *Computer Speech & Language, 22*(2), 171-184. doi:10.1016/j.csl.2007.07.003

Pavlovic, C. V. (1994). Band importance functions for audiological applications. *Ear and hearing, 15*(1), 100-104. doi:10.1097/00003446-199402000-00012

Payton, K. L., Uchanski, R. M., & Braida, L. D. (1994). Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing. *J Acoust Soc Am, 95*(3), 1581-1592. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8176061

http://scitation.aip.org/docserver/fulltext/asa/journal/jasa/95/3/1.408545.pdf?expires=1450733039&id=id&accname=2106341&checksum=20034FC2792AF3086CBFAD59E7495539

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1985). Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech. *J Speech Hear Res, 28*(1), 96-103. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3982003

https://jslhr.pubs.asha.org/article.aspx?articleid=1778112

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1986). Speaking clearly for the hard of hearing. II: Acoustic characteristics of clear and conversational speech. *J Speech Hear Res, 29*(4), 434-446. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3795886

Rhebergen, K. S., & Versfeld, N. J. (2004). An SII-based approach to predict the speech intelligibility in fluctuating noise for normal-hearing listeners. *The Journal of the Acoustical Society of America, 115*(5), 2394-2394. doi:10.1121/1.4780630

Ricketts, T. A., Henry, P. P., & Hornsby, B. W. (2005). Application of frequency importance functions to directivity for prediction of benefit in uniform fields. *Ear Hear, 26*(5), 473-486. doi:10.1097/01.aud.0000179691.21547.01

Rosen, S. (1992). Temporal Information in Speech: Acoustic, Auditory and Linguistic Aspects. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 336*(1278), 367-373. doi:10.1098/rstb.1992.0070

Schum, D. J. (1996). Intelligibility of clear and conversational speech of young and elderly talkers. *J Am Acad Audiol, 7*(3), 212-218. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8780994

Sherbecoe, R., & Studebaker, G. (2002). Audibility-index functions for the connected speech test. *Ear & Hearing, 23*(5), 385-398.

Steeneken, H. J. M., & Houtgast, T. (1999). Mutual dependence of the octave-band weights in predicting speech intelligibility. *Speech Communication, 28*(2), 109-123. doi:http://dx.doi.org/10.1016/S0167-6393(99)00007-2

Studebaker, G., & Sherbecoe, R. (1991). Frequency-importance and transfer functions recorded CID W-22 word lists. *Journal of Speech and Hearing Research, 34*(2), 427-438.

Studebaker, G., & Sherbecoe, R. (1993). Frequncy-importance and transfer functions for the Auditc of St. Louis recordings of the NU-6 word test. *Journal of Speech and Hearing Research, 36*(4), 799-807.

Studebaker, G., Sherbecoe, R., McDaniel, D., & Gwaltney, C. (1999). Monosyllabic word recognition at higher-than-normal speech and noise levels. *Journal of the Acoustical Society of America, 105*(4), 2431-2444.

Thibodeau, L. M. (2000). Speech audiometry. In R. J. Roeser, M. Valente, & H. Hosford-Dunn (Eds.), *Audiology: diagnosis*. New York: Thieme.

Tillman, T. W., & Carhart, R. (1966). An expanded test for speech discrimination utilizing CNC monosyllabic words. Northwestern University Auditory Test No. 6. SAM-TR-66-55. *Tech Rep SAM-TR*, 1-12.

Uchanski, R. M. (2005). *Clear Speech in D.B. Pisoni and R.E. Remez (Eds), The Handbook of Speech Perception*. Malden, MA: Blackwell Publisher.

Uchanski, R. M., Choi, S. S., Braida, L. D., Reed, C. M., & Durlach, N. I. (1996). Speaking clearly for the hard of hearing IV: Further studies of the role of speaking rate. *J Speech Hear Res, 39*(3), 494-509. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8783129

http://jslhr.pubs.asha.org/article.aspx?articleid=1781103

van Wingaarden, S., & Drullman, R. (2008). Binaural intelligibility predictions based on the speech transmission index. *Journal of the Acoustical Society of America, 123*(6), 4514-4523.

Venezia, J. H., Hickok, G., & Richards, V. M. (2016). Auditory "bubbles": Efficient classification of the spectrotemporal modulations essential for speech intelligibility. *J Acoust Soc Am, 140*(2), 1072. doi:10.1121/1.4960544

Vitela, A. D., Warner, N., & Lotto, A. J. (2013). Perceptual compensation for differences in speaking style. *Frontiers in Psychology, 4*, 399. doi:10.3389/fpsyg.2013.00399

Warner, N. L. (2005). Reduction of flaps: Speech style, phonological environment, and variability. *J Acoust Soc Am, 118*(3), 2035-2035. doi:10.1121/1.4785815

Wendt, D., Koelewijn, T., Książek, P., Kramer, S. E., & Lunner, T. (2018). Toward a more comprehensive understanding of the impact of masker type and signal-to-noise ratio on the pupillary response while performing a speech-in-noise test. *Hearing research*. doi:10.1016/j.heares.2018.05.006