SPEECH IN NOISE: EFFECTS OF NOISE ON SPEECH PERCEPTION AND SPOKEN WORD COMPREHENSION

by Jovan Eranović

A Thesis Submitted to the School of Graduate Studies
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy

McMaster University DOCTOR OF PHILOSOPHY (2022) Hamilton, Ontario (Cognitive Science of Language)

TITLE: SPEECH IN NOISE: EFFECTS OF NOISE ON SPEECH PERCEPTION AND SPOKEN WORD COMPREHENSION

AUTHOR: Jovan Eranović, MA (University of Sarajevo)

SUPERVISOR: Dr. Magda Stroińska

NUMBER OF PAGES: xix, 223

Lay Abstract

The study investigated the effects of noise, one of the major environmental stressors, on speech perception and spoken word comprehension. Throughout three different tasks (listening span task, in which participants were asked to recall a certain number of items from a list; listening comprehension task, in which listeners needed to demonstrate understanding of the incoming speech; and shadowing, in which listeners were required to listen and simultaneously repeat aloud the incoming speech), various types of background noise were presented in order to find out which ones would cause more disruptions to the two cognitive processes. The study found that general speech perception and specific word comprehension are not equally affected by the different noise maskers – provided that shadowing is considered primarily a task relying on speech perception, with the other two tasks considered to rely on working memory, word comprehension and semantic inference, or the way the listener combines and synthesizes information from different parts of a text (or speech) in order to establish its meaning. The results indicate that understandable background speech is most detrimental to speech perception, while any type of noise, if loud enough, as well as degraded target speech signal are most detrimental to spoken word comprehension. Finally, introducing a noise component to these tasks, adds another cognitively stimulating real-life dimension, which could potentially be beneficial to students of interpreting by getting them accustomed to working in a noisy environment, an inevitable part of this profession. Another field of application is the optimization of speech recognition software. In the last study,

the same types of noise as used in the first studies were tested on two automatic speech recognition programs. This technology was originally developed as an aid for the deaf and hard of hearing. However, its application has since been extended to a broad range of fields including education, healthcare and finance. The analysis of the transcripts created by the two programs found speech to text technology to be fairly resilient to a degraded speech signal, while mechanical background noise still presented a serious challenge to this technology.

Abstract

The study investigated the effects of noise, one of the major environmental stressors, on speech perception and spoken word comprehension. Three tasks were employed – listening span, listening comprehension, and shadowing – in order to find out to what extent different types of background noise affected speech perception and encoding into working verbal memory, as well as spoken word comprehension. Six types of maskers were used – (1) single babble masker in English, (2) single babble masker in Mandarin, (3) multi babble masker in Greek and (4) construction site noise, (5) narrow-band speech signal emulating phone effect and (6) reverberated speech signal. These could be categorized as energetic (2, 3, and 4), informational (1) and signal degradation (6 and 7) noise maskers. The study found that general speech perception and specific word comprehension are not equally affected by the different noise maskers – if shadowing is considered primarily a task relying on speech perception, with the other two tasks considered to rely on working memory, word comprehension and semantic inference. The results indicate that informational masking is most detrimental to speech perception, while energetic masking and sound degradation are most detrimental to spoken word comprehension. The results imply that masking categories must be used with caution, since not all maskers belonging to one category had the same effect on performance. Finally, introducing a noise component to any memory task, particularly to speech perception and spoken word recognition tasks, adds another cognitively stimulating real-life dimension to them. This could be beneficial to students training to become interpreters

helping them to get accustomed to working in a noisy environment, an inevitable part of this profession. A final study explored the effects of noise on automatic speech recognition. The same types of noise as in the human studies were tested on two automatic speech recognition programs: Otter and Ava. This technology was originally developed as an aid for the deaf and hard of hearing. However, their application has since been extended to a broad range of fields, including education, healthcare and finance. The analysis of the transcripts created by the two programs found speech to text technology to be fairly resilient to the degradation of the speech signal, while mechanical background noise still presented a serious challenge to this technology.

Acknowledgements

I would like to thank my supervisor, Dr. Magda Stroińska, for her support and encouragement during the four years I spent as a graduate student at McMaster University, and her patient and thorough guidance during the thesis writing. In addition, I would like to thank the committee members, Dr. Elisabet Service and Dr. Daniel Pape, for the constructive feedback and advice they provided during the writing process. I am also deeply grateful to Dr. Małgorzata Tryuk for providing critical comments on the final draft. Finally, I would like to thank a very dear colleague and friend of mine, Marijana Matkovski, whose expertise in statistical analysis proved vital for the successful completion of the project.

A very special thank you goes to Dr. Nikolai Penner for his extremely generous and invaluable help with participant recruitment, but also for his considerable support and technical assistance during my teaching fellowship.

List of Tables

List of Figures

CHAPTER 1: Introduction

1.1 The Scope of the Thesis

The study presented in this thesis investigated the effects of noise on speech perception and spoken word recognition. In particular, error analysis was conducted in order to find out to what extent and how different types of background noise maskers affected speech perception. The results of the study demonstrate that speech perception and spoken word recognition are equally affected by various noise maskers. They also reveal what type of maskers have the most detrimental effect on speech perception and spoken word recognition.

To expand on the human data, the same types of noise were tested on two automatic speech recognition programs. This technology had originally been developed as an aid for the deaf and hard of hearing, however, its application has since extended to a broad range of fields, including education and interpreting, the two most relevant for this project.

The idea for the choice of topic comes from the author's own experience of working as a simultaneous and consecutive interpreter, often in less than ideal conditions, exposed to various types of external noise. Conference interpreters usually work in the presence of other interpreters who may be speaking in other languages at the same time, and, more often than not in inadequately soundproofed booths. In addition, a working lunch or dinner is not an unusual setting for interpreters to work in. In these conditions they often cannot escape background speech and a variety of noises. Sometimes, interpreters are required to work in an outdoor setting, exposed to traffic noise, construction site noise, or a

cheering crowd at a sporting event. Those working as telephone interpreters are faced with another challenge – that of a reduced frequency bandwidth and reverberant auditory environment. Gaining a better understanding of the effects of these types of noise on speech perception and spoken word recognition can help schools of interpreting tailor their instructional strategies in order to better prepare their students for future practice.

1.2 The Structure of the Thesis

The structure of the thesis is as follows: in Chapter 2, we review existing literature that informed the current research. We first explain listening as a cognitive process, and make a clear-cut distinction between listening and hearing. Also explained are the obstacles that could impede our listening comprehension, as well as cognitive, metacognitive, and social strategies that the listener develops in order to cope with the auditory information. Importantly, listening takes place within a context, which can be any linguistic, physical and temporal environment in which interlocutors are situated. This context, as a result, informs our perceptual adaptation – or the adjustment of our phonological perception – but also syntactic and cultural perceptions. Three major theories that try to explain the link between perception and production are briefly discussed: the *motor theory* (Liberman & Mattingly, 1985; Moulin-Frier & Arbib, 2013), the *direct realist theory* (Goldstein & Fowler, 2003), and the *exemplar-based model of learning* (Pierrehumbert, 2003; Hay & Drager, 2006). This thesis relies on the exemplar-based model of learning, which holds that every time we hear speech, episodic

traces get activated in our memory. This view stems from Schema Theory, according to which every time we experience something new, that experience is being compared with similar experiences stored in our memory. Along the same lines, the theory of episodic traces assumes that long-term memory gets activated when we are presented with auditory input, which is reported to occur in both ordinary conversations and experimental settings, such as speech shadowing tasks. This is why the thesis concludes in favor of shadowing exercises to be used in interpreter training. While the main purpose for having this type of exercise is to help students get accustomed to simultaneously listening and speaking, such exercises should also help students of interpreting get accustomed to working in non-ideal auditory environment.

The second part of Chapter 2 defines the concept of noise, one of the most prominent environmental stressors, and then goes on to explain how different types of noise present in the environment – most notably *energetic* and *informational* maskers, but also degraded speech signal – interfere with speech processing and are detrimental to speech comprehension. Also discussed are adverse conditions classified by their outcomes, or the way they affect the listener, ultimately leading to obstructed transmission or otherwise distorted signal, and information loss, as well as some of the coping strategies used for overcoming such conditions.

The third part of Chapter 2 introduces the profession of simultaneous interpreting, connecting it to the research on cognition and memory. The section also discusses some the challenges an environmental noise present to interpreters.

The history of interpreting, and its different modalities will be briefly outlined, however, the very process of simultaneous interpreting will not be perceived as a simple one-way transfer of a message from source language to target language, on the contrary, it will be presented as a highly complex cognitive activity during which an interpreter is engaged in multiple processes – speech perception, listening comprehension, conversion of a message from source language into target language, speech production, and, finally, constant monitoring of both auditory input and own output.

The chapter also briefly discusses the concept of working memory – an important cognitive component responsible for temporarily storing and manipulating information – and the phonological loop, or the part of the system that temporarily stores verbal material during language processing. Also discused is a role of the phonological loop in our perception and production processes drawing on studies that analyzed its performance. These studies found that interpreters consistently demonstrated a working memory advantage over non-interpreters, while interpreter training was found to improve short-term memory and mental flexibility.

The final sections of Chapter 2 address bilingualism in the context of interpreting. Interpreters normally achieve superior fluency in at least two languages. This fluency goes beyond that of an average bilingual person, affecting linguistic processing, memory capacity and cognitive flexibility.

Chapter 3 introduces the research questions and the tasks studied in order to answer these questions.

Chapter 4 outlines the methods used in the research presented in this thesis, poses the research questions for which the author sought answers, and explains in detail the three tasks conducted – a listening span task, a listening comprehension task, and shadowing. In all three tasks commonly encountered types of noise maskers were added to prerecorded spoken stimuli (*construction* noise, *single babble energetic*, *single babble informational*, and *multi babble energetic*), or degraded sound (*phone* and *reverb* effects). All these maskers are explained in detail in the literature review section dealing with adverse listening conditions (Chapter 2, sections 2.2.2 and 2.2.3). Chapter 4 presents an in-depth data analysis of the results. The data analysis was performed using R Statistical Software (version 4.1.2), investigating and comparing error percentages and means obtained in the three tasks. In order to generalize the obtained results, inferential statistical analysis was conducted. In addition, the analysis is supported by numerous examples of the most typical errors found in the shadowing task – omissions, constructive errors, delivery errors and insertions.

Chapter 5 brings an overall discussion of the results obtained in the three tasks.

Chapter 6 is dedicated to a general discussion of the results presented in the previous chapter.

Chapter 7 briefly introduces automatic speech recognition technology. The results of testing two such programs using the same stimuli that were used in the shadowing task are presented. The chapter also includes statistical analysis of the

errors, as well as numerous examples of errors observed in various noise conditions.

Chapter 8 offers concluding remarks, and outlines some limitations of the study. Importantly, also discussed are implications for interpreter training. In addition to exploring speech perception in adverse conditions, the study argues that by being exposed to noise maskers in memory and shadowing exercises students of interpreting could potentially gain valuable experience of working in adverse conditions while still in the classroom, and that way prepare for one of the important professional challenges – that of listening to and processing speech in noise.

The Appendix contains transcripts of all the stimuli used in the tasks.

CHAPTER 2: Review of Literature

2.1 Listening

2.1.1 Listening Defined

This section discusses the phenomenon of listening, recognizing it as an activity whose aim is to perceive and process human speech. The only type of listening this thesis is concerned with is listening to human speech, including all its components. Listening in general, as in perceiving all possible sounds is of no concern for this project. Even though the terms *listening* and *hearing* are frequently used interchangeably, there is a significant difference between the two processes. They both start with sound perception; however, there is an element of intent that distinguishes listening from hearing. In other words, hearing is simply a psychological process during which listeners become aware of any noise in the environment, without any need for interpreting such noise, while listening, on the other hand is a multifaceted interpretative process during which listeners make sense out of the sounds they are hearing (Schnell, 1995). There have been numerous attempts to define listening, characterizing it either as a linguistic process (Brazil, 1995), or a perceptual process (Pickett & Morris, 2000) or a neurological process (Feldman, 2003). Even though each of the efforts offers a unique perspective while focusing mainly on one of the functional aspects of listening, there appears to be very little general agreement about what exactly listening entails, and how it operates (Dunkel, 1991). All the approaches to the definition of listening can be broadly assigned to one of the four umbrella categories – collaborative, constructive, receptive and transformative – yet, most

of them fail to give an all-encompassing definition of listening (Rost, 2011: 3). One of the reasons for this "underlying paradox" that research on listening typically runs into is caused by the nature of listening; for the very process that cannot be easily seen or accessed is extremely difficult to be investigated (Lynch, 1998). Listening being an invisible process, scholars sometimes resort to analogies, even metaphors, in order to define it. Listening can be described as largely "an active and complex process in which listeners must identify sounds and lexical items and make meaning of them through their grammatical structures, verbal and non-verbal cues and cultural context" (O'Bryan & Hegelheimer, 2009: 11). Perceiving listening as a *multidisciplinary endeavor*, Worthington and Bodie point out that "our understanding of key aspects of listening processes is woefully lacking," suggesting that instead of trying to come up with an ideal all-encompassing definition of listening, researchers need to focus on "determining the key features of specific listening processes and/or behaviors of interest to their particular research process" (Worthington & Bodie, 2018: 11). Essentially a language skill, listening is a precursor to speaking. The process starts with the listener perceiving auditory input, and continues through the identification of speech sounds and syllables combining it into words, then identification of prosodic features, and finally parsing all these elements into meaningful messages. Of course, while this description presents the process as strictly sequential, one should bear in mind that listening can also be conceived as "primarily a cognitive activity, involving the activation and modification of concepts in the listener's mind" (Rost, 2011: 57). All the different types of skills

and knowledge the listener employs are "capable of interacting and influencing each other" (Buck, 2001: 3). Thus, listening is foremost an interactive process during which "the various types of knowledge involved in understanding language are not applied in any fixed order" (Buck, 2001: 3). It takes a skilled and competent listener to successfully integrate all the components of social interaction and arrive at the intended meaning.

Listening must not be treated as an isolated activity with no direct reference to other language skills or personal experiences for "the conceptual knowledge that the listener brings to text comprehension needs to be co-ordinated in ways that allow him or her to activate it efficiently and continuously arrive at an acceptable cognitive understanding of the input" (Rost, 2011: 57). Importantly, in addition to physiological and cognitive factors, listening is also dependent on the listener's social knowledge or background. The absence of an exact and precise definition of listening "limits communication research in listening and lessens the chance of finding effective methods of training individuals to be effective listeners" (Dunkel, 1991: 433).

It was James (1984) who identified several components or stages of listening: the *sonic realization*, which includes physical hearing of language and distinguishing language and non-language sounds; the *segmental/suprasegmental form*, which relies on pitch, amplitude, duration (rhythm) or phoneme reduction for identification of different phonological and prosodic structures where the information about how something is said contributes to the overall meaning just as much as what is said; the *lexical phrasing*, where individual words and phrases

get encoded in order for the overall message to be understood; and, finally, the *purpose of the message* and the *actualization of the message*, where the interlocutors realize the ideas and intentions of the spoken messages (James, 1984).

2.1.2 Obstacles to Listening and Listening Strategies

Researchers have identified seven possible obstacles that could impede our listening comprehension: the speed of delivery, which the listener cannot control; the impossibility to hear the original speech repeated; limited vocabulary; the inability to recognize discourse markers (ranging from linguistic ones used in formal settings to nonverbal ones used in informal settings); foreign and regional accents that could hinder listening comprehension of native and non-native speakers alike; learning habits developed in a language classroom where the stress is put on understanding every single word while not paying enough attention to the overall context; and, finally, the lack of contextual information due to cultural or other differences between the interlocutors (Underwood, 1989). All these factors contribute to the mismatch between input and knowledge, creating gaps in comprehension. In order to successfully overcome these obstacles and comprehend the intended message listeners make use of various strategies or mental processes (Mendelsohn, 1995; Thompson & Robin, 1996; Vandergrift & Goh, 2012). Categorized into three groups – *cognitive*, *metacognitive* and *social* – these strategies include different mental processes they activate in order for the

listener to comprehend and retain novel, sometimes ambiguous, information (O'Malley, Chamot & Küpper, 1989).

Listeners resort to *cognitive* strategies, which assist the acquisition of both knowledge and skills, in order to handle the auditory linguistic information, process their learning tasks, and store and recall new information (Derry & Murphy, 1986). Scholars typically divide these cognitive strategies into two groups: top-down and bottom-up strategies, precisely corresponding to the two processes we engage in when we listen – top-down and bottom-up processes (Engel, 2010). Importantly, while distinction is made between bottom-up and top-down information, different computations in the brain rely to a different extent on incoming (bottom-up) and previously existent (top-down) information, which is why the two processes should never be seen as dichotomous. Both of these processes are interactive, requiring prior linguistic and pragmatic knowledge in order for the message to be understood. True listening occurs as a result of the two processes interacting successfully. Top-down listening primarily involves the listeners' contextual or background knowledge based on their prior experiences, including the knowledge of the subject discussed, but also any cultural knowledge that may be of relevance. On the other hand, with bottom-up processing, listeners are first registering and identifying the incoming auditory information, building the meaning from the smallest phonetic and phonological units and progressing to lexical meanings and syntactic relationships. This process is enriched by top-down information that becomes available to interact with it at successive neural locations on the way to cortical auditory areas. As noted earlier, listening does not

rely exclusively on only one of these processes; all the processes involved, as well as all levels of knowledge, rather interact while searching for the intended meaning (Rumelhart & McClelland, 1981: 37). The success with which listeners use the combination of bottom-up and top-down information depends on their language proficiency and familiarity with the topic discussed, both features being the result of experience, but also overall motivation. Interestingly, listening is also perceived as a form of intrapersonal communication, or "a mild form of daydreaming whereby we reflect on the meaning of what is said to us," (Schnell, 1995: 3). This process of reflection "draws heavily from our own past experiences, which can obviously be different than past experiences of others," and is also part of empathy building (Schnell, 1995: 3).

With *metacognitive* strategies, listeners monitor and evaluate the incoming speech by focusing on the auditory input they are receiving, but also making decisions about what segments they need to pay particular attention to (O'Bryan & Hegelheimer, 2009). These strategies include awareness of the task at hand, as well as the linguistic information the incoming auditory input contains, where the listener monitors and evaluates the incoming speech with the aim to successfully arrive at the correct meaning (O'Malley, Chamot & Küpper, 1989).

Finally, *social* or interpersonal strategies can be defined as those employed by listeners in collaboration with their interlocutors in order to facilitate better understanding of the message received. They include cooperative learning and asking questions for clarification (O'Malley, Chamot & Küpper, 1989). Social strategies are also important in the context of language learning, since the

learners' own feeling about their learning experience is found to significantly affect their learning (Rabinowitz & Chi, 1987). Studies show that the choice of strategy is determined by the listener's level of proficiency. Overall, skilled listeners are found to use these strategies more frequently than their less skilled counterparts (O'Malley, Chamot & Küpper, 1989). In addition, it has been demonstrated that listening strategies can be perfected and listening comprehension improved through targeted and extensive training (Macaro, 2006).

2.1.3 Listening as a Cognitive Process

There are two types of cognitive processing that we humans typically employ – controlled processing and automatic processing – the former requiring attention and effort, and the latter occurring automatically without any need for conscious attention or active control (Shiffrin & Schneider, 1977). More recent accounts see the two processes as two ends of an effortfulness–automaticity continuum (Logan, 1992). Given the speed and complexity with which typical speech unfolds, automatic processing is of utmost importance for the effective and efficient listener (Buck, 2001).

Early research on listening strategies suggests that in typical listening situations when processing no ambiguities, English native speakers primarily rely on semantic cues, or the broader contextual information provided, while non-native speakers focus their attention on prosodic and syntactic elements (Conrad, 1985; Harley, 2000). In that regard, it is important to note that foreign-language learners have been reported to either make inferences relying on acoustic and

contextual cues, or use elaboration in order to activate background knowledge of the topic (Berne, 2004). Subsequently, once their background knowledge has been activated, they first summarize the received message, after which they begin to self-monitor and self-evaluate their own comprehension and strategies employed (Berne, 2004). Unsurprisingly, the main cognitive component of listening is memory, and it is only by integrating memory models with the concept of listening that researchers were able to establish connections between different types of listening and specific individual predispositions (Bodie & Worthington, 2018).

Just like any other cognitive activity, listening is an interactive process, and treating it as if it were a process that consisted of a sequence of operations is simply wrong for the processing of the diverse types of knowledge may take place either in a fixed or in any other convenient order (Buck, 2001). Cognitive models of listening postulate that a successful listening experience depends on the listener's integrating all the available linguistic and extralinguistic information and matching it against the preexisting schemata (Brown, 1995).

Various facets of listening can be measured, such as recognition, recall, or retention. Regardless of the listening component a researcher is interested in measuring, an appropriate operationalization of the construct first needs to be established, for "very few listening phenomena are ever directly measured, but we can make claims about listening constructs based on their operationalizations" (Worthington & Bodie, 2018: 22).

2.1.4 Listening in Context

Listening typically takes place in or within a context – which is defined in linguistic literature as the language surrounding an expression – save for experimental scenarios when the listener is presented with either isolated target units ranging from speech sounds and syllables to words, phrases and full sentences. Being somewhat vague and an infinitely flexible concept to define, context needs to be understood as an exceptionally intricate framework encompassing physical, social and cognitive surroundings of language. Adopting Duranti and Goodwin's (1992) taxonomy, four contextual dimensions can be distinguished: *setting*, or the immediate physical environment in which communication takes place; *behavioral environment*, or the way the interlocutors use their facial and body gestures as well as any other nonverbal cues, such as posture or orientation; *language* as context, or the way in which speech itself provides context for the speech that follows; and, finally, *extrasituational* context, or the listener's general and specialist knowledge, as well as personal experience that together facilitate the grasp of meaning. Listener's understanding of the message received depends on continuous interaction between the input information and these four dimensions. Importantly, any contextual analysis should be understood from the perspective of those whose actions or behavior are being analyzed – in this particular case, the listener – where, again, the role that these four dimensions play is a crucial one.

2.1.4.1 Perceptual Adaptation

Similarly to any noise the listener may experience, unfamiliar accents are known to be quite challenging to understand, hampering the listening process. When exposed to heavily-accented speech, we tend to adapt our perception accordingly. The processing delays reported in studies on accented speech suggest that different mechanisms are relied upon when accented and unaccented speech are processed (Bürki-Cohen et al., 2001). Studies show that the standard American accent was perceived differently when accompanied with a picture of an Asian person, as opposed to when accompanied with a picture of a Caucasian person (Rubin, 1992). The mismatch between what listeners expect to hear and the actual acoustic signal they receive can result in heavier cognitive loads, slower processing, and impaired reception of messages. If the mismatch is too big, as in the case of heavily-accented speech, additional cognitive resources need be employed, negatively affecting both the listening process and comprehension (Hamada & Suzuki, 2020). Despite the fact that foreign accents and non-standard dialects can at first impair language processing, the available research demonstrates that more experience results in prompt adaptation (Cristia et al., 2012).

The process during which listeners adapt their perception of foreign accents, unfamiliar dialects, and various phonological features they encounter is called *perceptual adaptation* (Hamada & Suzuki, 2021). The process involves listeners adjusting their “preexisting phonemic categories to accommodate speakers’ pronunciation” (Kraljic & Samuel, 2007: 1). This is essentially a three-

stage process during which the listener initially perceives the unfamiliar input, which then gets mapped onto the stored lexical knowledge, leading to the generalization of the new mappings to other lexical items (Banks et al., 2015).

Different communication theories have offered explanations on how speakers adapt their speech to that of their interlocutors in order to facilitate more effective communication. During the process of adaptation, speakers either converge toward or diverge away from the speech of their interlocutors (Galloway & Rose, 2015). Studies show when both native and non-native speakers adjust their speech, mutual intelligibility can be achieved – if the conversation, for example, is carried out in English, interlocutors typically use non-standard English, that is, they simplify their speech for easier understanding (Carey, 2010; Jenkins, 2009). Despite the decades of research invested into the issue, the mechanism of perceptual adaptation is still not a fully understood one. Some scholars propose that the process is to some extent lexically-driven with speakers relying on acoustic cues and their lexical knowledge (Kraljic & Samuel, 2007). Even though this avenue of research has not produced any conclusive results regarding lexical motivations, it may be useful in explaining perceptual adaptation (Drozdova, van Hout, & Scharenborg, 2016). Others maintain that the degree of success of perceptual adaptation depends on extralinguistic factors such as the number of exposures to the accent, positive or negative feelings about the accent, as well as the variety of contexts in which these exposures took place (Bradlow & Bent, 2008).

One of the limitations of this body of research is that it has almost exclusively focused on interaction between native speakers, with only a few studies exploring the process of adaptation between native and non-native speakers. It should be noted that native and non-native speakers of any language listen differently (Cutler, 2012). When engaged in listening, native speakers primarily rely on accuracy in pronunciation, while non-native speakers' judgment depends on multiple elements, such as fluency, pronunciation, grammar, lexis (Saito et al., 2019).

One theory that partially explains the process of adaption is the Perceptual Assimilation Model (Best, 1995). The theory holds that when listening to unfamiliar non-native phonemes, naïve listeners typically assimilate such phonemes to the phonemes of their native language that appear most similar in terms of their articulatory properties (Hamada & Suzuki, 2020). During this process, such non-native phoneme is assimilated to a phoneme in the listener's native tongue, and perceived as a poor, moderate, or excellent exemplar of that category (Faris et al., 2018). However non-native phonemes can also remain uncategorized, and, thus, fall "in an untuned region in between categories," or not at all be perceived as speech and thus remain non-assimilable, falling "outside the listener's L1 phonological space" (Faris et al., 2018: 1). As flexible and dynamic a process perceptual adaptation is known to be, researchers suggest that "if given sufficiently extensive and intensive exposure to foreign-accented speech, native talkers may begin to shift their pronunciations in the direction of the ambient foreign-accented speech" (Bradlow & Bent, 2008: 727).

Similarly to this view, the Speech Learning Model (Flege, 1995) also attempts to explain how listeners perceive and eventually acquire incoming phonetic information. The model postulates that completely new sounds would not be causing any difficulties for if the phonetic dissimilarity between the closest non-native and native sounds is big enough, the learners will likely be able to acquire non-native sounds easily since "more dissimilar sound will be perceived as more obviously 'different' from L1 categories and thus the learner may eventually, given enough input and experience, be able to establish a separate category from the existing L1 categories" (Carlet & Cebrian, 2014: 86). However, the sounds that are perceived as similar to those in the listener's native language would be the most problematic ones for learners since both "L1 and L2 phonemes will be assimilated into one phoneme category" (Feng, 2020: 60). Despite this category *merger*, the model also holds that creation of new L2 categories can lead to category *dissimilation*, which is usually the case with younger learners (Flege, 1995).

Regardless of the model adopted, it is perception of non-native contrasts that always takes place first, only then to be followed by successful production of the same contrasts.

2.1.4.2 Spoken-word Recognition

The sole function of word recognition is to arrive at the word meaning. Some scholars maintain that during this process "listeners must map a dynamic, variable, spectro-temporally complex continuous acoustic signal onto discrete

linguistic representations in the brain, assemble these so as to recognize individual words, access the meanings of these words, and combine them to compute the overall meaning" (Johnsrude & Buchsbaum, 2017: 5). Namely, listeners are able to successfully engage in this process by selecting "a word from tens of thousands of words they know on the basis of the incoming perceptual information" (Auer, 2009: 419). Importantly, listeners do not perceive the speech signal as a continuous signal space; instead, they perceive speech in terms of unites, i.e. "distinct categories, along one or more linguistic dimensions or levels of analysis," which include "articulatory gestures or features, or phonemes, syllables, morphemes, or words" (Johnsrude & Buchsbaum, 2017: 5). There are several aspects of spoken-word recognition that affect the listener's working memory load.

The first of these aspects is the fact that a phonemic inventory of any language is extremely small in comparison with the extensive vocabulary it supports. Typically, an average number of phonemes a language may have is 31, with most languages having around 25, while vocabularies of these languages number tens, even hundreds of thousands of words (Lecumberri et al., 2010). One of the obvious implications of this relationship is that vocabulary items simply cannot manifest significant dissimilarity. A lot of words in some ways resemble each other, and shorter words are sometimes embedded in longer ones. In practice, this means that whenever the listener encounters a spoken word, a whole range of possible mapping candidates gets selected. *Mail* could become *mailman* or *mailbox* or *mailing*, or it can become part of a bigger compound such as

mailing list. Thus, the spoken-word recognition task is "one of sorting out what is intended to be there from a large set of alternatives that are only partially or accidentally there" (Lecumberri et al., 2010: 866). This phenomenon is also known as the *multiple activation process*, during which all the words that "partly overlap with the input, irrespective of their onsets, are activated simultaneously" (Scharenborg & van Os, 2019: 54). Activation of multiple lexical representations that overlap with the input is reported to "rapidly decline as soon as there is mismatch with the input" (Friedrich et al., 2013: 1). However, in real-life scenarios, incoming speech is often mixed with noise or degraded due to the speaker's accent, which makes the selection process more difficult.

The second aspect affecting processing load is the very process of speech production. The articulatory gestures used for production of speech sounds may exhibit variations in different contexts – therefore, one articulatory movement for a given sound made in one phonetic context may differ from the movement used for the production of the same sound in a different phonetic context. What such phonetic coarticulation adds to speech is variability, while its continuous articulation makes phonemes, but also words, difficult to be individually differentiated (Lecumberri et al., 2010). Thus, while *mailman* and *mailgram* are fairly easy to differentiate when pronounced in isolation, in rapid speech, the listener may have hard time differentiating between the two. On the other hand, coarticulation also adds redundancy to the speech signal, making it easier to anticipate what will be coming next. The number of words that get activated, as well as the nature of these words, has been proved to affect accuracy and speed at

which word recognition takes place; as a result, extra cognitive effort is required for processing words belonging to a dense or high-frequency neighborhoods (Scharenborg & van Os, 2019). In its essence, word recognition is a process of selection during which "rival lexical alternatives compete," and "evidence in favor of any one of them simultaneously counts against all rival candidates" (Lecumberri et al., 2010: 866). Finally, during this competition "active candidates that fail to match the acoustic input and/or the semantic context are inhibited, leaving the optimal word candidate given the acoustic input and semantic context to be recognized" (Scharenborg & van Os, 2019: 54). Of course, in case of homophones, the lack of contextual cues in the acoustic input would inhibit the disambiguation (Stroińska & Drzazga, 2018).

The third aspect affecting processing load is the fact that speech can take place virtually anywhere – so that the distance between interlocutors, or the conversational setting itself can introduce variability between conditions and additional levels of processing complexity. The listener is constantly being exposed to speech input that varies in its rate and quality, yet the task of spoken-word recognition most of the time unfolds without great apparent effort, at least for a proficient language user. In the final stage of the word recognition, which is also referred to as the *integration process*, the appropriate semantic information "related to the selected words is integrated into the ongoing sentence" (Scharenborg & van Os, 2019: 54).

In addition, there are also motor constraints that are usually not taken into account when speech perception is analyzed, but they do play a vital role in the

listener's selection of possible mapping candidates and their overall comprehension of speech. There is a number of impossible combinations of phonetic features in articulation, such as producing a fricative and a glottal stop simultaneously. Such constraints "put limits on the way that acoustic information covaries in the speech signal, and listeners have been shown to be highly sensitive to such co-varying cues, even with novel non-speech sounds" (Scott, 2017: 31).

2.2 Speech in Noise

2.2.1 The Effects of Noise on Speech Recognition

Noise is one of the major environmental stressors. Just like any other sound, noise is "the result of vibration within any physical medium" (Szalma & Hancock, 2011: 682). Of concern in this study is noise produced and transmitted in the medium of air. Similar to other forms of energy, noise varies in amplitude, frequency and duration (Szalma & Hancock, 2011). Defining noise is a challenging task. Noise can be perceived to be "all external acoustic energy," in which sense all speech should also be considered to be "meaningful or variable noise" (Koelega & Brinkman, 1986: 466). Conversely, noise can be regarded as something that "conveys no information to the brain," and therefore treated just like "music or the squeal of a tire as varied auditory stimulation or environmental stressors" (Koelega & Brinkman, 1986: 466). Noise can also be described as "nuisance" and "unwanted sound" (Stansfeld & Matheson, 2003: 243). For the purpose of this study, any sound signal arriving at the listener concurrently with the speech target and interfering with it will be considered as noise. In addition,

all the artifacts resulting from degradation of speech signal will also be considered a noise.

The unit of measure for sound intensity is a decibel (dB). This, of course, applies to any sound, speech included, but also any type of noise. The loudness level of typical speech measures usually between 45 and 65 dB, and it can go as low as 10 dB in case of a whisper. Importantly, the decibel scale is not linear, thus, with each 3 dB increase, the sound intensity and loudness doubles. Loud traffic or a construction site can register up to 90 dB, while loud concerts typically exceed the 100 dB mark.[1] Any prolonged exposure to sounds of intensity of 80 dB or more can be dangerous for one's hearing, while exposure to sounds exceeding 120 dB in intensity can immediately damage one's hearing. Some studies indicate that exposure time is of greater importance than the actual exposure level (Prell et al., 2011). More often than not, signals that reach our ears are coupled with background noise, and the relationship between the two is expressed by the signal-to-noise ratio (SNR). Expressed in decibels, the ratio of 0 dB means that the levels of signal and noise are equal. In experimental settings, the level of noise of manipulated variables typically varies between 0 dB to -8 dB, in other words, the volume of noise is either equal to the target signal or lower.

Speech is ordinarily conveyed by transient acoustic signals of variable, yet complex structure, which present the listener with the challenge of mapping these incoming signals onto mentally stored representations (Guediche et al., 2014). In a typical listening environment, the listener is faced with the task of "processing

[1] For more information on how to calculate sound intensity and volume see Speaks (2018).

of complex auditory sequences such as speech," the success of which depends on "the accurate analysis of multiple elements within time-varying patterns" (Leek & Watson, 1984: 1038). This task is often augmented by various adverse conditions arising from noisy environments. There are several factors that determine speech intelligibility – the distance between the speaker and listener, reverberation time and the level of environment noise, the loudness of the speaker's vocal output (Assmann& Summerfield, 1992). Any of these factors can be a source of adverse conditions for the listener, in which case the overall intelligibility depends not only on what researchers refer to as *spectral overlap* between speech and noise, but also on the duration of the speech peaks in the spectrogram (Mattys et al., 2012). Importantly, unlike the case for visual or tactile stimuli, research demonstrates that it is speech sounds that primarily access our cognitive processing structures (Armstrong & Sopory, 1997). Thus, in any information-processing task performed under adverse conditions, it is our phonological working memory, and the ability to accurately process incoming auditory signals, that gets affected the most. In such cases, two or more auditory or other stimuli compete for our limited processing resources.

Listeners typically find the task of following a conversation in a noisy environment challenging, for noise routinely affects several levels of speech processing – "acoustic-phonetic cues may be insufficient (or absent because of coarticulation and deletion) for word identification when listening to speech in noise or when listening to faster, more casual speech" (Smiljanic & Sladen, 2013: 1086). Moreover, as the complexity of auditorily perceptible information

increases, the listener's ability to follow the target speech decreases. Adverse conditions that listeners encounter in everyday speech typically result from added sound energy from other sound sources, sound reverberating off reflecting surfaces, and channel distortions (Lecumberri et al., 2010). The types of noise listeners are exposed to fall into two broad categories: other meaningful conversations they can attend to, or meaningless noise, for instance, coming from a nearby construction site or overhead fan exhaust. *Auditory masking* is the term used for the deficit in sound perception, and it is in general defined as "the process by which the threshold of hearing one sound is raised by the presence of another" (Wang & Xu, 2021: 110).

In a typical conversational environment, our ability to clearly and accurately perceive speech relies on the ability of our auditory systems to distinguish incoming speech from the background noise (Smiljanić & Sladen, 2013). Any noise present in the environment interferes with signal processing and can be detrimental to speech comprehension, not only because noise acts as an *energetic* masker covering portions of the frequency spectrum, but also because noise can act as an *informational* masker confusing the listener in the choice of sound source to attend to. The type of background noise is directly responsible for the degree of processing challenge the listener is presented with.

2.2.2 Energetic and Informational Masking

The first type of interference, which distracts the listener by integrating the speech and masker at the level of peripheral auditory processing, is labeled

energetic masking (Lecumberri et al., 2010). Energetic masking is defined as a situation in which "competing signals overlap in time and frequency in such a way that portions of one or more of the signals are rendered inaudible" (Brungart et al., 2001: 2528). Energetic masking typically affects speech perception by making potential segmental cues unavailable, but also by blocking access to prosodic cues (Lecumberri et al., 2010). This type of masking, therefore, results in some of the signal components being lost, so the interpretation of incoming speech only relies on partial information (Cooke et al., 2008). If intense enough, energetic masking can sometimes lead to ambiguity or a total loss of the information contained in the target-signal (Lecumberri & Cooke, 2006). Any sound produced by a competing source, spectrally resembling the speech target, and interfering with the processing of the target can act as an energetic masker (Ezzatian et al., 2012).

The second type of masking where competing speech obstructs the processing of the target speech is labeled *informational masking* (Schneider et al., 2007). An informational masker can be any information that affects overall intelligibility making both the target and masker similar and confusable along several acoustic dimensions (Helfer & Freyman, 2005). Unlike energetic masking, informational masking occurs in listening situations in which both the target and masker are clearly audible; however, the listener is not able to segregate the components of the target signal from the components of the distracters which sound very similar to the target signal (Brungart et al., 2001). In other words, informational masking takes place when there is no acoustic overlap between the

signals, resulting from high-level factors such as attention and perceptual grouping (Snyder et al., 2012). The masking signals being variable and perceptually resembling target sounds, this type of masking is characterized by poor auditory detection of target signals which are embedded in the masker (Wightman et al., 2006). Informational masking is characterized by reduced intelligibility for the available attentional resources are busy processing the masker (Lecumberri & Cooke, 2006).

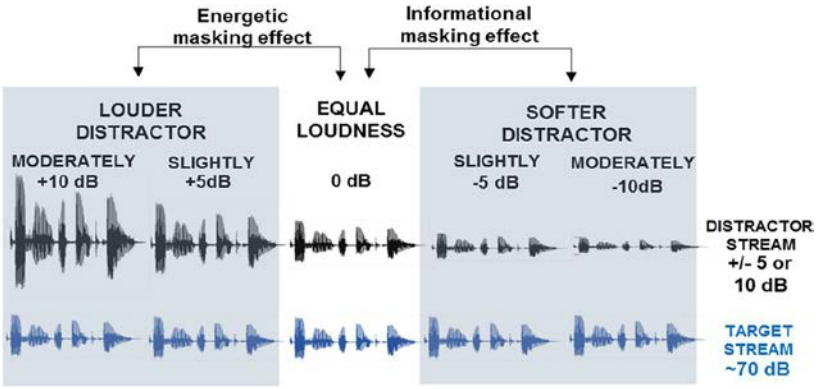

Figure 2.1. Difference between energetic and informational masking (adapted from Szalárdy and colleagues, 2019: 4).

Figure 2.1., adapted from Szalárdy and colleagues (2019: 4), aptly illustrates the difference between the two types of masking: the bottom blue stream depicts target speech, while the upper black stream is the masker. In the gray box on the left, the black stream works as an energetic masker, for its peaks coincide with the peaks of the target sound, but also have much larger amplitudes, which sonically obscures the target signal. On the other hand, the black signal in the right gray box is presented as an informational masker as it has similar energy as the target signal, and also its peaks are unequally distributed.

Following the speech of a single speaker in a noise-free room does not pose much of a problem for the typical listener. However, following a multi-person conversation in a crowded bar, does decrease one's ability to segregate and process the target speech. The key factor that hinders our ability to closely follow speech in auditorily complex environments is the signal-to-noise ratio (SNR). Another important factor that adds to the difficulty is the listener's inability to properly locate, identify and separate the competing sound signals. Any breakdown in the analysis of an auditorily presented signal can lead to phonological or semantic interference. Importantly, "when the target is speech and the masker is noise, the target will elicit activity in the phonetic, semantic, and linguistic systems whereas the masker is unlikely to do so" (Schneider et al., 2007: 560). On the other hand, when both the target and the masker are speech, in most listening scenarios both will cause activity in the systems responsible for language processing (Schneider et al., 2007). For instance, studies have found that following a single person talking when there are other talkers around is challenging. The challenges in such situations are the result of intrusions of the competing speech because listeners are either unable to separate the different incoming streams, or their attention keeps switching between the target speech and the intruding speech. Therefore, masking of target speech occurs precisely because of the inability of listeners to parse the auditory scene into different components in the way in which they can focus all their attention to one source while ignoring the others (Schneider et al., 2007).

Studies investigating the effects of energetic masking have often been concerned with spoken-word recognition among native versus non-native listeners. Since energetic masking primarily interferes with the acoustic information in the speech signal, it provides a good starting point for exploring the phenomenon of signal masking. Informational masking, on the other hand, is the masking effect that persists after the consequences of energetic masking have been taken into account (Scharenborg & van Os, 2019). Once energetic maskers have been investigated, their effects can be assessed against the remaining effects, those typically attributed to informational maskers. Therefore, understanding the effects of energetic masking on the listener's perception and comprehension of the incoming speech makes the understanding of informational maskers easier.

Two types of energetic maskers have been regularly used across the studies. The first is speech-shaped noise with an average long-term spectrum similar to that of an adult male speaker (Scharenborg & van Os, 2019). This type of masker is designed to simulate the noise commonly found in real-life situations, such as at a traffic intersection or a loud party. Also used frequently is multi-speaker babble masking, which presents one of the most taxing adverse conditions for the listener due to its time-unfolding nature and resemblance to the target speech. When challenged with this type of masker, the listener's ability to recognize the target speech will depend on the gaps in the babble. With an increased number of speakers in the babble, gaps become fewer, making identification of the target speech more difficult. The multi-speaker variable can range from two to a dozen.

Testing bilingual subjects with a word-pair paradigm, Golestani and colleagues (2009) wanted to examine whether the semantic context contributes in any way to the intelligibility of words heard through different levels of noise. They were also interested in finding out whether native speakers would outperform non-native speakers under the same experimental conditions, that is, under the same levels of noise. In each trial of the lexical decision task the subjects were presented with both auditory and visual stimuli. The auditory stimuli consisted of word pairs – the first word was a semantically related prime presented in noise, and second word was the target presented without noise. Upon hearing the target word, a subject would be visually presented with two words, having to decide by pressing a button which of the two visually presented choices corresponded to the prime. The study found that the perception of previously presented primes was facilitated by semantically related targets only for native speakers (Golestani et al., 2009). Their results suggest that the semantic context plays a key role for the native language advantage under the adverse conditions. However, comparing the results between the two groups, the authors report the same performance, the only difference being the "signal quality level needed to show context benefits," concluding that "native and non-native listeners have a different 'threshold' for being able to make use of higher-level, linguistic context" (Golestani et al., 2009: 391). The authors concluded that speech processing in non-fluent bilinguals is significantly more automatic in their native language, so that in case of any adverse conditions, such as poor articulation or signal degradation, parsing in the non-native language becomes increasingly difficult for

both the higher-level linguistic resources and the lower-level speech input processing levels reaching a bottleneck (Golestani et al., 2009). By contrast, their native tongues only limit the lower-level speech input processing (Golestani et al., 2009).

By testing native and non-native (Dutch) speakers of English in three word-recognition tasks, Brouwer and colleagues (2012) wanted to examine the effects of different maskers on word recognition on the two groups. In addition, the authors also wanted to investigate if linguistic similarities contained in both target and masker signals would contribute to speech-on-speech masking (Brouwer et al., 2012). The stimuli they used comprised of English sentences mixed with either English or Dutch two-speaker background babble. An interesting finding was reported for the bilingual group whose first language was Dutch, showing that when engaged in processing their second language their processing resources were mainly preoccupied with relevant English speech, making fewer processing resources available for irrelevant Dutch speech (Brouwer et al., 2012). These findings were in line with another study which also noted that “bilinguals invest more of their resources in processing a target and/or processing a target leads to a stronger inhibition of competitors” (Colzato et al., 2008: 310). The study found that speech-in-speech recognition accuracy was in fact determined by linguistic similarity (Brouwer et al., 2012). Namely, the results illustrated that word recognition was more accurate when the background babble was in a different language from the target language. Hence, as the target speech became more similar to the masker speech, be it the same or a different language

or in terms of semantic content, the recognition accuracy decreased. Finally, the study illustrated an interesting relationship between linguistic masking and the listener's knowledge of both the target language and the background masker language, arguing that "while signal-bound, energetic masking differences may dominate stream segregation for speech-in-speech recognition, target-masker linguistic similarity likely makes an independent contribution raising the possibility of speech-in-speech enhancement strategies that focus on these factors" (Brouwer et al., 2012: 1462). In other words, masking at the energy level occurs independently of masking at the linguistic level.

Jin and Liu (2012) wanted to investigate sentence recognition in long-term speech-shaped noise and multi-talker babble. Their starting premise was when the spectrum and amplitude of noise are fluctuating, resembling a multi-speaker babble, the listener should be able to make use such temporal gaps in the masker (Jin & Liu, 2012). However, the authors wanted to find out whether the listener's native language would influence sentence processing in adverse conditions. To that end, in addition to a native speaker group, they recruited two groups of non-native speakers, one Korean and the other Mandarin, to explore the role of language background on English sentence processing in noise. The study found that the masking release, or the listeners' ability to recognize the target when the masker is not at its peak, was greater for the native speaker group than their non-native counterparts, indicating that the amount of masking release indeed depended on the listeners' native tongue (Jin & Liu, 2012). Interestingly, despite the quite similar levels of English proficiency reported for the two groups, the

Korean group showed greater masking release than the Mandarin group, suggesting that the release from masking was associated with something other than English proficiency (Jin & Liu, 2012). The results were in line with previous research reporting that as Mandarin is a tonal language, listeners would rely on different acoustic and phonetic cues for speech perception than speakers of non-tonal languages. Since frequency and amplitude were modulating in multi-talker babble, this might have affected the tonal perception of the listeners, leading to a lower masking release (Jin & Liu, 2012).

Van Engen (2010) tested speakers of English and Mandarin in the babble condition. The author first normalized the participants' tolerance for energetic masking so that the effect of babble speech on the two groups could be investigated (Van Engen, 2010). The results obtained confirmed the initial predictions that non-native listeners would require a better signal to noise ratio – which turned out to be at around 8dB – in order to correctly recognize English sentences in strictly stationary, speech-shaped noise (Van Engen, 2010). The study confirmed previous findings in which native speakers' performance was markedly superior to that of their non-native counterparts. The study also reported that both native speakers of English and L2 English speakers found the English background babble more disruptive than the Mandarin babble, suggesting that "acoustic and/or linguistic similarity between the speech signal and the noise may be the most critical factor in driving noise language effects" (Van Engen, 2010: 951). Finally, the author observed that in addition to noise acting as an energetic masker by mechanically interfering with audio signal, another factor that had to

be borne in mind is concurrent informational masking caused by the native-language noise (Van Engen, 2010).

Interested in exploring the effects of both energetic and informational masking in a word-recognition task, Cooke and colleagues (2008) used single-talker competing speech with similar masking and target sentences for both the target speech and the background speech. In line with their predictions, the results indicated that background noise had greater impact on non-native listeners' identification of keywords in simple sentences (Cooke et al., 2008). Of the two noise conditions, the authors found that for both native and non-native listeners informational masking was more detrimental to spoken-word recognition. The authors also noted that in low-noise listening situations, in which detailed acoustic information of individual speakers was available, previous experience, such as exposure to different accents, led to the native advantage (Cooke et al., 2008). The explanation for this was that drawing upon a richer knowledge native listeners were able to interpret the signal more accurately. Nevertheless, when high levels of noise were present in the signal, both native and non-native listeners found the same talkers easy or difficult to recognize, thus, it was not background knowledge, but rather acoustic factors that they ultimately relied on (Cooke et al., 2008).

Overall, all these studies reported no difference between native and non-native listeners for word-recognition tasks in quiet condition. Differences in performance between the two groups commonly referred to as *native advantage* become apparent in the presence of background noise. One should bear in mind

that any comparison in performance between native and non-native listeners can be further complicated by the fact that non-native listeners make up a very heterogeneous group (Lecumberri et al., 2010). However, what also needs to be borne in mind is that in its essence the structural design of the spoken-word recognition system is not language-dependent, thus, when the listener is in the search for the ideal mapping of the incoming acoustic signal onto a word, activation of multiple word candidates, as well as competition between the candidates will always be taking place, regardless of the vocabulary (Lecumberri et al., 2010). The differences between any two groups of speakers can be measured using one of the following three methods.

The most commonly used method to measure word recognition in noise is to vary signal-to-noise ratio across trials, calculating the number of correctly identified words. As the noise level becomes higher, word recognition deteriorates. Another approach recommends establishing signal-to-noise ratios in which both groups have almost identical scores. The size of the reported adjustment for the non-native listeners varies from +2dB (Brouwer & Bradlow, 2016) and +4dB (Cooke et al., 2008) to +6dB (Lecumberri et al., 2010) and +8dB (Van Engen, 2010). The third approach recommends varying signal-to-noise ratio for each participant until an accuracy of 50% is reached. Studies report that non-native listeners reach the accuracy of 50% when noise levels are much lower than those used with native listeners, which is typically -2dB to -4dB (Kaandorp et al., 2015). Generally, based on the available research, it is believed that differences in performance in adverse conditions that favor native listeners reflect their less-

than-optimally developed phonetic categories, which are influenced by factors such as native-language interference and differences in quality and quantity of foreign-language exposure (Lecumberri & Cooke, 2006).

2.2.3 Reverberation

Sound reflecting off nearby surfaces and reaching the listener following indirect paths is said to be *reverberant*. Reverberation is referred to as "the persistence of a sound in an enclosed environment" (Rogers et al., 2006: 466). It is measured in time that is needed for a specific frequency sound pressure wave do decay by 60 dB after the original signal ends (Rogers et al., 2006). Even though reverberation is usually considered to be a background noise, unlike energetic or informational masking, reverberation is not a type of noise that comes from a different source – rather, its masking energy originates in the target speech itself. In case of reverberation, the masker contains additional energy which, unlike with the other types of maskers, is "correlated with the sound source which produced it, leading to different masking patterns" (Lecumberri et al., 2010: 873). Sounds that continue to produce prolonged reflections are known to degrade the speech transmission and hinder speech perception (Reinhart & Souza, 2018). Reverberated speech is speech mixed with time-delayed reflections that represent scaled versions of the original signal, and result in a smeared signal (Assmann & Summerfield, 1992). In reverberated speech offsets are typically muddled, phonemes prolonged, and bursts smoothed (Scharenborg & van Os, 2019). Reverberation is reported to smear the spectro-temporal information of the

speech across both phoneme and word boundaries (Reinhart & Souza, 2018). In addition, reverberation has been found to more reduce intelligibility of late segments than of early segments, but also to better preserve vowel identity than consonant identity (Mattys et al., 2012). Affecting both prosodic and segmental information, reverberation amplifies horizontal structures appearing on the X axis of the spectrogram, such as static formants, at the same time blurinng vertical ones that appear on the Y axis, such as transient and bursts (Lecumberri et al., 2010). Therefore, the spoken message that eventually reaches the listener is a combination of direct and reflected signals, resulting smeared speech signal and blurred spectral detail (Assmann & Summerfield, 1992).

Testing early Spanish-English bilinguals and English monolinguals, Rogers and colleagues (2006) investigated the effects of noise and reverberation on word recognition. The authors considered early bilinguals those children who were exposed to Spanish since infancy and to English before age 6. Their study found no differences in performance in a no-noise condition between the two groups, however, in both groups decrease in SNR resulted in poorer performance (Rogers et al., 2006). The study also found that overall performance was much inferior in a simulated noisy and reverberant condition than in noise alone (Rogers et al., 2006). Importantly, the bilingual group demonstrated significantly lower performance across the noisy and reverberant conditions, suggesting that despite the fact that early bilinguals were quite capable of understanding low-volume speech while in typical listening situations, they were much less resistant to signal degradation than their monolingual counterparts (Rogers et al., 2006). A potential

explanation for these differences in performance could be interpreted through greater demands for attentional and processing resources demand for the bilinguals for one of the following reasons: the bilinguals had to deactivate one of their languages prior to making selections; and the bilinguals were presented with a much larger selection of alternatives when selecting a target phoneme (Rogers et al., 2006). The study concluded that even though bilingual speakers were usually perceived as having an advantage over their monolingual counterparts for cognitive skills such as creativity and problem solving, recognizing words in a noisy or reverberant environment presented a greater challenge for them (Rogers et al., 2006).

2.2.3.1 The Precedence Effect

A sound produced in a reverberant environment travels in multiple directions and reflects off nearby surfaces. This amalgamation of sounds presents the listener's auditory system with a challenging task of perceiving and localizing the competition between the original signal and its refections (Litovsky et al., 1999). Besides its capacity to detect and receive incoming sound information, an essential feature of our auditory system is the ability to analyze incoming acoustic signals and distinguish between different auditory elements in them (Blauert & Braasch, 2005). When the listener is exposed to a complex acoustic input, the auditory system is responsible for differentiating between the direct signal and its reflections (Blauert & Braasch, 2005). This multifaceted phenomenon known as

the precedence effect enables the listener to properly identify and localize the competing sound sources in cognitively taxing reverberant environments.

Researchers studying the precedence effect typically use two-source paradigm to understand the process. This allows them measure the leading direct signal and the lagging reflection. If a listener happens to be in a space with sound-reflecting barriers or walls, the sound emitted by a source would inevitably arrive to the listener via different paths: via a direct path, which is typically the shortest distance between the source and the listener; or via a number of reflective paths which represent the sound coming from the source, hitting any reflective surface on its way and bouncing off it, changing its direction accordingly. Thus, when the listener is surrounded with multiple sound sources emitting signals simultaneously, the sound information arriving at the listener is the mixture of direct waves coming from the sources and their numerous reflections. Each individual sound reflection arrives at the listener at a different time, and with a different delay. For example, if the listener is placed in the middle of an anechoic room with a single sound source on the right hand side and only one reflecting barrier on the left hand side, there would be two signals to attend to: the first a direct waveform coming from the sound source (e.g. a loudspeaker), and the second a waveform reflecting from the barrier.

Alternatively, the listener can be seated in the center of an anechoic chamber surrounded by two sound sources: one to the right and the other to the left. In this setup, a reflection can be imitated by having the second loudspeaker producing time-delayed version of the original signal. This way, two competing

signals arrive at the listener from different directions and with different time delays. When faced with such competing sounds, the listener's auditory system needs to decide whether the two waveforms are coming from the same, single sound source, or two different sources. In case of a shorter time delay, which means that the spectro-temporal characteristics of the delayed signal are close to those of the original sound signal, the listener's perceptual system tends to perceive the source of the leading waveform as the sound source (Schneider et al., 2007). The direct sound and the reflection appear as a single blended image situated between the two sounds (Blauert & Braasch, 2005). In a scenario like this, our auditory system is able to identify the direct sound, which corresponds to the location of the sound source, while successfully disregarding its reflections (Blauert & Braasch, 2005). When this knowledge is applied to speech heard in a noisy environment, one of the following will occur: when the listener's target is speech and the masker is noise, unlike the masker, the target will elicit phonetic and semantic activity (Schneider et al., 2007). In situations when both the target and the masker are speech, they are both likely to initiate activity in language-processing systems (Schneider et al., 2007). As the delay between two sources increases, the precedence effect gets more pronounced. With slight delays, our auditory system is still able to disregard localization cues contained in reflections, and recognize the direct leading sound as the dominant one (Blauert & Braasch, 2005). However, as the delay between the sounds increases, the auditory image widens reaching the threshold level of echo suppression. Once the threshold has been reached, the sound reflection becomes perceived as a new auditory image

(Blauert & Braasch, 2005). Finally, the information contained in all the reflected sounds is also significant since it enables the listener to perceive the actual environment as reverberant.

2.2.3.2 Auditory Scene Analysis

A model of psychoacoustics that has traditionally been involved in exploring our auditory perception and organization of incoming auditory stimuli in the form of discrete sounds or sequences of sounds from different sources around us is known as *auditory scene analysis* (Bregman, 1990). The listener regularly receives complex auditory signals from an unidentified number of sound sources. The main duty of our auditory system is to process this complex mixture of sounds so that it can successfully identify the sources of the incoming signals (Szabó et al., 2016). Researchers started exploring auditory scene analysis hoping to understand how the listener segregates the available acoustic stimuli and groups them into meaningful word *streams*[2]. For example, the listener in a crowded conference room with numerous people talking simultaneously will need to segregate the irrelevant background speech form utterances produced by the interlocutor sitting just opposite, and will also need to arrange the various sound components of this attended speech into a meaningful and coherent flow of words. Along the way, study paradigms have been developed equally fitting for study of perceptual mechanisms, attention and the effect of previous knowledge on perception (Snyder et. al., 2012).

[2] In this sense, a *stream* is defined as "the percept of a group of successive and/or simultaneous sound elements as a coherent whole, appearing to emanate from a single source" (Moore et al., 2012: 919).

Two or more tones of an adequate frequency separation continuously alternating at a faster rate, create a streaming effect, which is strictly a perceptual phenomenon determined by the rate of stimulation and the frequency relationship between tonal sequences (Sussman et al., 1999). As a result of this perceptual effect, the listener is able to segregate between the streams and hear the groups of high and low tones that are split into separate sound streams, one composed of the high tones and the other of the low tones (Sussman et al., 1999). At this point, an important distinction needs to be made between two facets of *streaming*. The first of these is the *auditory stream segregation* during which different parts of the sensory data are linked together, and eventually determine what auditory events are included or excluded from our perceptual descriptions (Bregman, 1990). The second process, referred to as *streaming effect*, occurs when sounds of distant frequencies are heard as separate audio streams, while sounds of adjacent frequencies are perceived as unified auditory objects (Chakalov et al., 2013). Thus, streaming effect is a phenomenon when during a sequence of rapid high and low tones, the two groups are perceived as forming separate streams (Bregman, 1990). It is the streaming effect that is responsible for our ability to accurately perceive different sonic events, and its success depends largely on the perceived difference between successive sounds, but also the rate at which these sounds are presented (Moore & Gockel, 2012).

Subjects taking part in experiments testing auditory scene analysis are expected to report their subjective experience on how they perceived two or more incoming sounds. Typically, if dissimilarities between adjacent sounds in a quick

sequence are too small, the sequence is then perceived as a single stream (Moore et al., 2012). However, when dissimilarities are too large, the sequence is then perceived as two or more streams (Moore et al., 2012). In addition, also available and fairly useful are incidental measures based on the subjects' performance, measuring the same effects and lending themselves to direct comparison. Importantly, the majority of auditory scene analysis studies rely on data obtained through rather simple experiments, with sounds that appear to be not too hard to either generate or manipulate (Snyder et. al., 2012). A typical experimental setting involves stimuli comprised of sine wave tones bearing almost no resemblance to the complex sounds the listener encounters in everyday life (Teki et al., 2013). Similarly, usually absent from these experiments are complex real-world acoustic scenarios whose interpretation normally depends on the activation of long-term memory or processes related to expert knowledge (Snyder et. al., 2012).

Two types of mechanisms have been suggested to participate in auditory scene analysis: a *primary* mechanism, which automatically processes incoming streams of sounds; and a *schema-based* mechanism that is believed to be dependent on our attention and prior knowledge (Bregman, 1990). A prototypical example of the primary mechanism is the frequency-separation effect taking place during segregation of sequential auditory samples (Bregman & Campbell, 1971; Van Noorden, 1975). This primary mechanism is seen as pre-attentive sensory segregating mechanism for which the auditory system relies on basic cues present in the stimulus (Moore et al., 2012). These cues indicate whether one sound is similar to the next along the spectrum, or whether there are significant changes

across different sounds. In an experimental condition, the subject is played two alternating pure tones (A and B) of different frequencies in the following pattern ABA-ABA, in which the hyphen represents silence. The two tones are initially perceived as a single galloping pattern; however, after a number of repetitions of the whole sequence, the tones begin to be perceived as if splitting in two completely different streams (Bregman, 1990). Also, as the gap between the two tones is expanded, or the rate of there is presentation is increased, subjects are likely to begin to hear two separate streams as well (Snyder et. al., 2012). Scholars have been using this experimental paradigm to learn how we auditorily perceive the world around us. The evidence that streaming takes place at central sites argues in favor of the possibility that auditory perception is a consequence of multiple auditory-level processing, involving schema-based mechanisms (Snyder et. al., 2012). These schema-based mechanisms are seen as top-down selection mechanisms in which our stored knowledge about sounds is activated during processing and organizing of the incoming auditory information (Moore et al., 2012). This means that when we hear a sound or a stream of sounds, we either use the available cues, such as frequency spectrum and timbre, building up our perception of the stream from the bottom up, or we rely on the schematically represented knowledge we already have of auditory object and events, employing top-down selection.

Finally, another equally important facet of auditory scene analysis is the segregation of concurrently occurring sounds in scenarios in which two speakers speak at the exact same time. In cases like this, in which the listener must

perceptually integrate various spectro-temporal components of the speaker' voice, and successfully segregate them from those of other speakers, there are several available cues that can help the listener accomplish this task (Snyder et. al., 2012). Harmonically related sounds, or the family of sound frequencies related to the individual fundamental frequency of the speaker and makes up differences between different voices, as well as sounds that appear to come from the same location at the same time are most likely produced by the same source (Ciocca, 2008). Laboratory experiments, in which only one spectral component of a pure tone is detuned (i.e., a harmonic), show that such a tone becomes easily distinguishable from the rest of the presented tones (Moore, 1986). Attention has been found to play a vital role in auditory scene analysis. Streaming has been reported to be less successful in situations in which there was a concurrent competing task – with the actual detrimental effects of the competing task depending on the task demands (Cusack et al., 2004).

2.3 Outcomes of Adverse Conditions

In addition to the previously discussed adverse conditions based on their origin (energetic and informational maskers, and reverberation) adverse conditions can also be classified based on their outcomes, that is, the way they affect the listener. Regardless of their characteristics, adverse conditions in general lead to an extent of mismatch between the segments perceived by the listener and their canonical forms (Mattys et al., 2012). They ultimately affect and shape "the quality and reliability of the acoustic speech signal and negatively

impact word recognition, reducing intelligibility" (Guediche et al., 2014: 2). Interestingly, while a number of such conditions seem to appear unrelated, they do elicit a shared perceptual effect, indicating that the same compensatory mechanisms responsible for them (Mattys et al., 2012).

Typical consequences of adverse conditions are a failure in mapping the acoustic and phonetic properties to their segmental representations, and a failure to map segmental representations to appropriate lexical representations. These could result in information loss, due to obstructed transmission, too big a distance between the signal source and the listener, or otherwise distorted or narrowed spectral domain, as in the case of cochlear implants or telephone transmission.[3] Adverse conditions can be a result of the departure from the listener's expectations in both acoustic and phonetic domains, as is the case of accented speech or mispronounced words. Acoustically degraded, but also accented, speech normally deviates from what the listener is expected to hear, typically resulting in lexical uncertainty on the part of the listener or, in severe cases, incorrect lexical selection, or no selection at all. The effects of accented speech manifest themselves along the lines of both segmental and suprasegmental features of the incoming speech. However, in addition to having an effect on speech intelligibility, accented speech is known to affect the accuracy and efficiency of linguistic processing, as well as auditory memory (Van Engen & Peelle, 2014).

Any mismatch between incoming auditory input and stored representations will lead to a loss of acoustic information. Depending on the

[3] Telephone speech transmission involves a narrow-band frequency range from approximately 300Hz to 3400Hz, which degrades both the intelligibility and overall quality of the transmitted speech signal. For more detail, see Pulakka et al., 2012.

nature and quantity of the lost material, our comprehension system selects an appropriate method of coping with these challenges. The information contained in speech is typically highly redundant, thus, missing parts can be easily reconstructed (Bronkhorst, 2015). Therofore, relying on coarticulation, echoic memory, and the comprehension system, the speaker can successfully bridge brief gaps in the input signal. The lost information can be recovered if it contains permissible segment deletions in conditions that are richly contextualized, as is the case in conversational speech" (Mattys et al., 2012). Redundancy is typically defined as "any set of factors that reduces the number of alternatives from which a stimulus might be chosen" (Zola, 1981: 4). This measure of certainty actually applies to two concepts – structural redundancy and contextual redundancy. For the purpose of this study, concerned with word recognition and processing, only structural redundancy will be considered. In case of deviations along the acoustic and phonetic structures, the overall impact of adverse conditions will depend on word frequency, for rich lexical activation will most likely to lead to larger competition and less robust lexical selection (Mattys et al., 2012). This implies that comprehension of unambiguous, predictable and redundant material will not be severely affected by degradation of the input signal.

2.3.1 Perceptual Interference

The speech signal competing with other non-target speech results in perceptual interference. The resulting interference can be low-level, when the competing speech masks only part of the target speech, but it can also be higher-

level, when the competing speech prevents an accurate interpretation of the target speech by diverting the listener's attention from it. The masking effect is, in this case, determined by either syntactic or semantic interference of the masker (Assmann & Summerfield, 1992).

Our comprehension system is believed to handle energetic masking as gaps in information, however, the competing signal presents a challenge for both signal/noise separation and selective attention (Mattys et al., 2012). Even in situations when parts of the target words are inaudible as a consequence of energetic masking, our auditory system is able to perceptually reconstruct the missing information. This process, known as *phonemic restoration*, involves processing at the spectro-temporal level, as well as reflecting lexical and semantic expectations about the speech content (Carlile & Corkhill, 2015). The process of phonemic restoration is primarily unconscious and automatic. It relies on speech redundancy for minimizing the interference effects of any irrelevant signal, while simultaneously evoking overlearned or otherwise familiar patterns from long-term memory (Assmann & Summerfield, 1992). In other words, using both prior and subsequent contexts, phonemic restoration enables the listener to restore the missing sounds (Warren & Obusek, 1971).

Similarly to energetic masking, informational masking engages our selective attention, however, this type of masking does depend on both lexical and segmental familiarity of the masker (Mattys et al., 2012). The maskers reported to cause the greatest degree of interference are identifiable babble speech and native speech (Assmann & Summerfield, 1992). They both hamper our speech

perception causing a spectro-temporal overlap and leading to auditory masking, because their own syntactic and semantic features compete with the target speech processing (Assmann & Summerfield, 1992). Importantly, it has been reported that this type of masking can be circumvented through training (Leek & Watson, 1984). Analyzing the results of their experiments, Leek and Watson (1984) found that after thousands of exposures, their participants who prior to the experiments had not been familiar with the sound sequences were ultimately able to perceive individual components in the complex pattern. The authors concluded that "the informational masking associated with stimulus uncertainty can be reduced by providing extensive experience with a stimulus" (Leek & Watson, 1984: 1043). This point is of particular relevance for the present thesis that will argue in favor of shadowing exercises in interpreter training programs, through which novice interpreters can improve their perception, selective attention, and overall listening skills on which adverse conditions place considerable demands.

In addition to the above-mentioned outcomes, adverse conditions bring about secondary consequences, such as reduced attentional capacity and reduced memory capacity. Any linguistic or nonlinguistic task that is simultaneously performed with speech certainly puts more demand on our attentional resources, resulting in reduced attentional capacity. Under dual-task conditions, supplementary visual cues have been reported to place an extra strain on processing resources, clearly leading to poorer performance on the secondary task (Gosselin & Gagné, 2011). Similarly, TV as a source of background noise has been found to degrade performance on various types of complex cognitive tasks,

including reading tasks, nonverbal problem-solving tasks and cognitive flexibility tasks (Armstrong & Sopory, 1997). Of course, different adverse conditions will tax our attentional capacity in different ways, especially if the distractor has its own semantic content.

Adverse conditions tax our memory, too, when it has to be engaged in multiple tasks. Research shows that when listeners are presented with enough variation in acoustic-phonetic structure and speaker characteristics, this will put extra demands on their attention, resulting in slow recognition (Nusbaum & Morin, 1992). Both short-term and echoic memory is known to be affected by adverse conditions (Mattys et al., 2012). In the case when the listener is trying to attend to more than one speaker, the process will exploit additional memory resources for within-speaker variation from utterance to utterance seems to initiate an estimation process relying on the information contained in the utterance to provide the context for interpreting that utterance (Nusbaum & Morin, 1992). A recent study in eye-tracking found that the spoken word recognition system interacts with background speech since patterns of lexical competition change as a function of the background speech content (Brouwer & Bradlow, 2016). This indicates that during phonological competition in a word-recognition task, the specific lexical content of background speech that must be ignored, affects lexical competition patterns throughout the time period of spoken word recognition (Brouwer & Bradlow, 2016). Recognizing, segmenting and processing degraded speech also exhaust our memory resources. Reduced working memory capacity becomes particularly apparent when the listener is trying to process longer chunks

of speech simultaneously performing semantic integration (Caplan & Waters, 1999).

2.3.2 Overcoming Adverse Conditions

Unmasking, or overcoming the effects of adverse conditions, can be either low-level or high-level, depending on the auditory processing involved. Low-level unmasking involves the reduction of energetic factors, while high-level unmasking activates selective attention and linguistic competition.

One way of release from masking is spatial separation of the competing signals – when the target speech and noise masker are spatially separated, the target speech becomes more intelligible. Spatial separation is expected to reduce informational masking since it is easier for the listener to ignore the information contained in the masker and focus on the target only (Schneider et al., 2007). In addition, spatial separation is known to improve detection and recognition of target speech (Schneider et al., 2007). Likewise, spatial separation is known to facilitate the perceptual segregation of voices competing for processing attention (Assmann & Summerfield, 1992). Studies investigating spatial separation report a release from both energetic and informational masking (Kidd et al., 1998). In addition to physically separating the sources of sound signal and enabling the listener to hear previously obscured parts of the target speech more clearly and accurately, spatial separation also reduces attentional stress (Sussman, 2017) and cognitive load (Andéol et al., 2017).

Another way to compensate from effects of masking is through familiarity with the content. Listeners' task of processing a message in a noisy environment will be much easier if they are familiar with the topic of conversation, in which case, parts of the conversation may help to recover missed words and phrases (Schneider et al., 2007). Familiarity with the conversational topic should also draw the listener's attention to the relevant speaker in case of being surrounded by multiple talkers. Interestingly, research shows that as the number of background talkers increases, our attention is less engaged in selecting between them suggesting that informational masking is most likely to take place when the listener's focus is on the speech of one speaker, while in the presence of nearby conversations (Freyman et al., 2004). Investigating the effects of different informational maskers Freyman and colleagues conclude that earlier familiarity with the content of the target talker's speech releases informational masking to a degree, the evidence of which could be found in superior recognition of key words that were omitted from a target-sentences preview (Freyman et al., 2004).

2.3.3 Contextual Information

Spoken-word recognition failure in adverse conditions, as a result of the listener's reduced attentional and memory resources, may be avoided if the listener is able to make use of higher-level linguistic and contextual information (i.e. lexical, semantic, and syntactic cues) in order to compensate for the losses at the perceptual level (van der Feest et al., 2019). Nativeness is also a major factor that determines the listener's ability to draw on contextual cues since it fosters

accessing of “contextual information in degraded listening environments” (Smiljanic & Sladen, 2013: 1093). Contexts of high-predictability have been found to increase the possibility of successful target word identification by reducing the number of possible word candidates, which, in turn, facilitate word recognition (van der Feest et al., 2019). Semantic information has also been reported to enhance comprehension using contextual information (Smiljanić & Sladen, 2013). It has been revealed that grammatical forms and semantic meaning facilitate the target-word recognition by either increasing their probability or reducing the number of possible lexical candidates, which consequently reduces the burden placed on the processing resources of the listener (McCoy et al., 2005). Interested in finding out whether wider discourse and sentential context information would improve word recognition, Bouwer and colleagues tested recognition of reduced target words in a spontaneous conversational setting (Bouwer et al., 2013). They found that when exposed to discourse information listeners did improve their target-word recognition (Bouwer et al., 2013). In addition, the authors concluded that the role the wider contextual information plays in word-recognition is also the result of bottom-up speaker adaptation processes (Bouwer et al., 2013).

2.3.4 Visual Cues

In situations where auditory and visual modalities provide complementary information, visual cues may provide a release from both energetic and informational masking. Providing supplementary information to the ongoing

speech in the forms of segmental and suprasegmental features, visual cues are reported to reduce the auditory processing demands, and increase the certainty of a recognition response (Grant & Sitz, 2000). Importantly, in order for the speaker to be able to extract useful information from both channels, both visual and auditory information have to be perceived as originating from the same source (Helfer & Freyman, 2005). Integrating visual and auditory input has also been reported to improve auditory perception of accented speech (Banks et al., 2015). What also testifies to the prominence of visual cues in speech perception is the fact that when making prosodic judgments listeners typically focus on the speaker's eyes and the area around the eyes, while the area around the mouth is reported to be vital for lexical decisions (Swerts & Krahmer, 2008). A study exploring perception of accented speech across different modalities found that when faced with an accented speech, the listener may be able to adapt to it by exploiting the visual information contained within the message (Banks et al., 2015). This is because during the perception process, the information from both auditory and visual channels is at one point integrated to form a coherent percept, facilitating the listener's identification of unclear phonemes and prosodic cues (Swerts & Krahmer, 2008). Finally, visual cues aid the listener in localizing the target signal and separating it from the rest of the incoming auditory input, which basically constitutes noise, thereby directing the listener's auditory attention and "reducing temporal and spectral uncertainty" (Grant & Sitz, 2000: 1206). Therefore, in a typical real-life scenario, speech perception in adverse conditions

can be substantially improved by supplementary visual cues provided by the speaker (Sumby & Pollack, 1954).

Lastly, training listeners to discriminate between various sound features, that is to distinguish between fine-detailed spectral or spatial differences in ordinary non-speech sounds proved to be beneficial for improving speech recognition (Gao et al., 2020). In addition to the immediate consequences of adverse conditions, there are later cognitive processes that do not facilitate comprehension of the current auditory input, but rather help the perceptual system adjust, in order to have better understanding of subsequent speech (Mattys et al., 2012). Such processes, involving fairly long-lasting changes to one's perceptual system, are referred to as *perceptual learning*; and they are known to improve one's ability to respond to environmental stimuli (Goldstone, 1998). Perceptual learning entails that the listener had been exposed to degraded or in some way noncanonical speech different from the speech typically experienced, and the exposure having led to a change in subsequent language processing. It has been well-documented that one's native language experience affects representation of individual speech sounds; however, brief auditory trainings, typically the ones taking place in a clinical environment or laboratory, are also known for their ability to modify speech perception (Kraus, 1999). Common paradigms for investigating perceptual learning involve evaluating the effects of prior exposure to degraded speech on later perception where listeners are first presented with an unfamiliar speech stimulus, and where such exposure results in an improved ability to successfully identify or discriminate that particular type of speech

stimuli (Samuel & Kraljic, 2009). Fine-tuning performance of their perception system, listeners progressively learn to adapt to unfamiliar accents and improve their ability to process degraded speech. Based on the results of the studies reviewed in this section, it is safe to say that perceptual learning commonly occurs in adverse conditions. Of course, perceptual learning will not be as effective, or will not take place at all, in adverse conditions in which the target speech is degraded by an unpredictable masker which is difficult to be learned from a limited number of trials, or in scenarios in which either completely unintelligible speech is present, or no external feedback given (Mattys et al., 2012).

The following subchapter will introduce the profession of interpreting, connecting the research on cognition and memory, with the theory of simultaneous interpreting. Research has shown that working memory does improve with practice. This is why it is important for interpreters to be aware of how working memory – a primary mechanism for overcoming any distractors during an act of interpreting, background noise included – operates, but also what exercise to use to improve it.

2.4 Interpreting

2.4.1 Brief History of Simultaneous Interpreting

When exactly interpreting developed as a form of mediated communication is difficult to pinpoint, for throughout history interpreting has been known as a fairly common activity that “did not merit special mention” (Pöchhacker, 2004: 159). Interpreting, sometimes referred to as the world’s

second oldest profession, is a human activity that involves "a direct, immediate and highly personal act of mediation between individuals, who are nonetheless subjected to the political imperatives of the situation they occupy in the spaces between cultures" (Delisle & Woodsworth, 2012: 279). Cognitively complex behavior, interpreting can be perceived as a form of elaborate "human information processing involving the perception, storage, retrieval, transformation, and transmission of verbal information" (Gerver, 1975: 119).

Among the early interpreters were priests, monks, missionaries, merchants and adventurers from whose accounts we learn about their encounters with worlds unknown to them at the time – Roman Monks exploring the British Isles and Scandinavia (Delisle & Woodsworth, 2012), Portuguese conquistadors and mercenaries in India in the early 16th century (Jackson-Eade, 2018), Chinese monks (Lin, 2002) and Jesuit travelers (Restif-Filliozat, 2019) exploring India, Spanish monks acting as interpreters in the Spanish Inquisition Tribunals in the period between the 15th and the 19th centuries (García-Morales, 2016), British colonials in their conquest of the New World (Luca, 1999) and French Catholic missionaries spreading their faith in Canada (McLean, 1890). These are only a few examples of documented contacts that would have been unimaginable without interpreting, or misinterpreting for that matter. The methods and approaches these missionaries sometimes used are beyond the scope of this thesis.

Historically, there has always been an overlap between interpreting and diplomacy. The further back in time we look, the more difficult it becomes to separate the two professions. Diplomatic positions were typically held by the

upper classes, those with better education, and until the Renaissance, Latin was the *lingua franca* among the upper classes and diplomats. During the period of Renaissance, other languages found their way into diplomatic communication, most notably French. Going even further back in history, no contact, or conflict for that matter, could have been possible without some sort of language intermediaries. Unfortunately, unlike translation, where surviving artifacts can help us put the historical puzzle together, significantly fewer records exist that are related to interpreting activities. There are two reasons for this – the very nature of speech as a transient or short-lived phenomenon, and the low status that interpreters characteristically enjoyed throughout history (Baigorri-Jalón & Takeda, 2016).

Most scholars date back the origins of simultaneous interpreting to the end of World War I and the 1920's. Inadequate as they are, the existing oral and written accounts help us paint a rather fragmentary picture of the history of interpreting. Many questions remain unanswered, particularly those related to the distant past. Cary (1956) was among the first scholars to offer a historical perspective on what he referred to as *official interpreting*, which mostly dealt with diplomatic, military and judicial settings. Using archival documents, Baigorri-Jalón (2000), a former UN interpreter himself, paints a detailed history of simultaneous interpreting starting with the 1919 Paris Peace Conference and leading all the way to the Nuremberg Trials. Chernov (2016) informs us about this novel mode of translation being successfully employed at the 11th International Labor Conference in Geneva and the 4th Congress of the Comintern in Moscow,

both around the same time in 1928. There is no record on what type of interpreter training was offered before the Moscow conference, but what we do know are the details of the training the interpreters received in Geneva. The training sessions would involve a person "reading aloud transcripts of speeches delivered at previous sessions of the conference, while another person interpreted them, and a third person listened in order to assess the interpretation" (Seeber & Arbona, 2020: 370). Upon the completion of the training, the candidates were tested, and the best ones selected based on their scores (Baigorri-Jalón, 1999). The training provided varied in length – between two weeks and two months – and the trainees mostly worked with past conference documents and improvised speeches (Mackintosh, 1999). There are mixed accounts on exactly what kind of training the interpreters working for the Nuremberg trials received. One of the points they seem to be consistent on is that most of the interpreters who worked there had no previous training in simultaneous interpreting. While some had worked as translators before the Trials, others found themselves in a sink or swim situation from day one (Keiser, 2004). The very first audio system for simultaneous interpreting was developed by IBM in the 1920's, and it successfully debuted at the League of Nations in 1931. Having undergone several modifications, the system was also used during the Nuremberg Trials. Despite the scarcity of the accounts detailing all the technical aspects of the interpreting process at the Trials, we know that some of the challenges the interpreters faced were huge amounts of feedback and crosstalk, as well as speakers who spoke too loudly or who were too close to the microphone (Guise, 2020). Also reported were heavy amounts of

reverberation in headphones, making both interpreting and listening to interpretation ever more challenging (Visman, 2011). After the Nuremberg Trials, simultaneous interpreting became a recognized profession and a staple at every major international event, including meetings, conferences, seminars, workshops etc. Demand for simultaneous interpreters drastically increased, which resulted in first interpreting schools being opened throughout Europe, but also in the US. During the Trials, there was only one school of interpreting in the world, and it was based in Geneva.

The Nuremberg veteran, Léon Dostert was instrumental in bringing this mode of interpreting to the U.S. After his success at the Nuremberg Trials, Dostert introduced simultaneous interpreting at the United Nations, where he also oversaw the installation of the interpreting equipment in council halls and conference rooms. He also partnered with IBM with an aim to improve the equipment used at the Trials, and soon founded Division of Interpreting and Translation at Georgetown University.

The second part of the twentieth century saw the rise of interpreter training programs, associations, and publications, all contributing to the advancement of interpreting as a profession, but also the establishment of ethical principles and standards, and professional codes and guidelines. With a sudden need for more trained interpreters, new schools were soon being opened. Interestingly, even in this formal setting, simultaneous interpreting was not taught at first. It was its alumni association that first introduced informal training in simultaneous interpreting in Geneva, after a request by their graduate student working at the

Nuremberg trials at the time (Keiser, 2004). Unsupervised sessions took place in the school basement, using improvised equipment. The lack of qualified and experienced interpreters prompted the then-European Economic Community to establish a six-month training course for their simultaneous interpreters in 1964 (Heynold, 1994). Over the next several years, this program evolved into the European Masters in Conference Interpreting program, which is currently made up of sixteen member universities offering a coordinated core curriculum, including the theory and practice of consecutive and simultaneous interpreting.[4]

Microphone interpretation, as it was referred to back in the day, was first introduced in Canada in 1949 at the University of Montreal. Two years later, the course was incorporated into the graduate program in translation and interpreting (Dalisle, 2020). In January 1959, simultaneous interpreting was introduced in the Canadian House of Commons. Every speech, address, debate, question and answer has since been interpreted between English and French.

Glancing at the curricula used in these programs, we learn that they cover various topics, registers and styles. They also employ various exercises and practice materials so that students can master all the necessary skills a qualified interpreter should possess.

2.4.2 Modes of Interpreting

An important distinction needs to be made between simultaneous and consecutive interpreting since the two modes put quite different demands on

[4] For details on the curriculum and member universities see https://www.emcinterpreting.org

memory and cognitive abilities. While the core tenets of both modes have remained unchanged throughout decades, teaching methods and strategies have significantly advanced; and they included insights and results from other disciplines, while beginning to adapt to the new working environment (Riccardi, 2005).

When engaged in simultaneous interpreting, the interpreter's attention is divided between several different processes including listening to the incoming speech, processing and transforming the incoming speech into target language, verbally producing and delivering the message in target language while listening to a new sequence in source language, monitoring all these concurrent processes, and correcting possible mistakes (Pöchhacker, 2004). In this regard, simultaneous interpreting should be perceived as a typical scenario of divided attention for it involves multiple cognitive tasks that are carried out more or less at the same time (Lambert, 2004).

In contrast, when engaged in consecutive interpreting, interpreters are handling speech chunks varying in length from several seconds to several minutes (Phelan, 2001). It is the job of the interpreter to memorize such a chunk for a brief period of time, accurately transforming it into target language, and finally delivering it in target language as well. One of the standard aids comes in the form of note-taking, however, throughout the whole process, the primary focus remains on listening and memorizing.

Scholars agree that simultaneous interpreting is one of the most cognitively demanding tasks. Simultaneous interpreting goes way beyond the

simple manipulation of verbally received information for it involves constant monitoring and updating of both input and output, where the incoming information has to be clearly understood and translated into a different language, and at the same time old information updated with new, relevant one (Morales et al., 2014). In terms of intellectual demands, simultaneous interpreting can be seen as a skill, procedural knowledge, and competence, thanks to the multiple cognitive processes it activates (Riccardi, 2005). Literature indicates that professional interpreters usually demonstrate advantages over those who had not acquired this skill in both short- and long-term memory tasks (Padilla et al., 1995; Bajo et al., 2000; Babcock et al., 2017). In addition, both professional interpreters and students of interpreting have proved to be almost immune to *articulatory suppression* – the decreased span that is a result of irrelevant phonological information rehearsal during memorization tasks[5] (Padilla et al., 1995; Yudes et al., 2012). The fact that these advantages were observed in comparison to bilingual and multilingual participants suggests that even though a necessary prerequisite for the very process of interpreting, bilingualism alone is not all it takes to be a successful interpreter. Instead, specific and intensive training programs are generally required through which a student can acquire necessary skills. Gerver and colleagues (1989) offer a practical overview of necessary skills an interpreter must possess in order to be able to meet the extraordinary challenges this profession will throw their way: a thorough knowledge of the

[5] Articulatory suppression is a term used to describe a blocking of a short-term memory subvocal rehearsal (explained in 2.5.1) typically during a memory task during which a participant is required to utter speech sounds simultaneously.

languages they work with, as well as the respective cultures; ability to quickly understand the meaning of what is being said; the ability to clearly and precisely formulate and articulate the intended message; broad general knowledge and interest to acquire new knowledge. Drawing on empirical evidence, it is important to assert that specific training in the cognitive processes of updating and shifting has elicited cognitive changes, especially when engaged in complex tasks that concurrently manipulate several information streams (Babcock et al., 2017). In particular, along with providing necessary knowledge and competencies, a systematic and comprehensive training in simultaneous interpreting that imposes excessive demands on memory as well as other executive control processes may significantly enhance the interpreter's overall cognitive performance.

2.5 Interpreting and Memory

2.5.1 Working Memory and the Phonological Loop

So far, the notion that human memory has two systems for information storage – short-term memory and long-term memory – has been generally accepted among scholars (Conway et al., 2005; Baddeley, 2007). The first system holds what we are focused on in the moment – arguably it can store up to about seven items (monosyllabic digits) at a given time, a much debated proposition to this day – while the second holds everything that we know and can recall (Liu et al., 2016). Both systems have been subject to extensive academic and clinical research, with interpreting studies drawing on both in developing its scholarship.

Short-term memory or working memory refers to the memory system "used for holding and manipulating information while various mental tasks are carried out" (Hiramatsu, 2000: 317). It is useful think of working memory as a multicomponent system consisting of a number of different systems that interact with each other, while varying in their degree of modularity (Baddeley, 2017). This system is responsible for constant "maintenance of information in the face of the ongoing processing and/or distraction," where the active maintenance in itself is "the result of converging processes – most notably, domain-specific storage and rehearsal processes and domain-general executive attention" (Conway et al., 2005: 770). By studying working memory, scientists have been trying to figure out the ways in which we temporarily manipulate and store information. Also of interest is how we process information while the mind is busy with various other tasks. The study of short-term memory first started as the study of the idea of a simple temporary memory storage, however, it soon became clear that a system acting as an interface between memory and perception, but also attention and action, undoubtedly requires much more capacity than simple temporary storage (Baddeley, 2007).

Most scholars today subscribe to Baddeley's model of working memory, which was originally proposed in 1974 by Alan Baddeley and Graham Hitch, but the model has since been modified and expanded to explain the new data Baddeley was getting in his research (Baddeley, 2007). The latest iteration of this model holds that our working memory consists of four components: the *central executive*, the *phonological loop*, the *visuospatial sketchpad* and the *episodic*

buffer (Baddeley, 2007). According to this view, our working memory functions as "a temporary storage system under attentional control that underpins our capacity for complex thought" (Baddeley, 2007: 1).

It is the phonological loop that plays a crucial part in listening to and processing auditory information, speech included. The phonological loop, sometimes also referred to as *phonetic loop* or *articulatory loop* is responsible for processing auditory and verbal information, subvocal rehearsal, and immediate and temporary storage of verbal information (Byrne, 2017). It is practically a slave system that has two components that is responsible for temporarily storing verbal material (Vicari et al., 2004). The phonological loop consists of the phonological store or "the passive maintenance of verbal information in a phonological code," and a subvocal rehearsal process, responsible for rehearsing phonological information and preventing "the material stored in the phonological store from decaying by refreshing the temporary traces" (Vicari et al., 2004: 81).

The evidence for the phonological store came from the *phonological similarity effect* (also referred to as *acoustic similarity*), an impairment of short-term memory for visual letter sequences, when the letters have similar sounding names (Conrad, 1964; Wickelgren, 1965; Baddeley & Dale, 1966). Based on this effect, Baddeley proposed that short-term memory depended on "an acoustic memory trace, with visually presented items being converted into an acoustic code by subvocalization" (Baddeley, 2007: 8).

In addition, the phonological loop is also thought to be the very system in which verbal material that is visually presented gets recoded into a phonological

format (Vicari et al., 2004). Auditorily presented material is believed to be kept in a phonological store of limited capacity, from which traces disappear fairly rapidly, but can also be revived by subvocal articulation (Baddeley, 2007). The information gets stored in the phonological store for approximately two seconds and then it fades away; however, in order for it to be retained in working memory for a longer time, the information must be constantly repeated or rehearsed (Hiramatsu, 2000). Similarly, subvocal articulation can also perform verbal recoding of visually presented stimuli, enabling such stimuli to be stored in the phonological loop (Baddeley, 2007). This constant repetition is performed by the articulatory component of the phonological loop, in which information revived through the store, and its life extended (Hiramatsu, 2000). The process of rehearsal is thought to be reflected in the *word length effect*, the finding that as the words became more complex and longer, observed was a decline in immediate memory for strings of words (Baddeley, 2007). The concepts of acoustic similarity and rehearsal incorporated, our memory span seems to be determined by the speed at which traces fade, but also the rate at which items are rehearsed (Baddeley, 2007). Shorter words lend themselves to rapid rehearsal which leads to a greater span. The function of the phonological loop can be tested by using the *articulatory suppression effect* (Padilla et al., 1995; Yudes et. al., 2012; Liu et. al., 2016). Vocal speech is known to hinder the normal process of subvocal rehearsal that is taking place in the phonological loop, by allowing irrelevant information to dominate the control process (Daró, 2002). The articulatory suppression effect, thus, decreases the short-term memory span when subjects simultaneously

rehearse phonological information in the short-term memory and articulate aloud irrelevant material (Liu et. al., 2016). The overt articulation blocks the phonological loop by preventing the encoding of phonological information (Yudes et. al., 2012). The articulation will only disrupt the phonological rehearsal if "it is performed in the same modality as the linguistic information being rehearsed" (Liu et. al., 2016: 363). Because of its dual nature, articulatory suppression is capable of involving demanding attentional processes, but also affecting overall performance (Saeki & Saito, 2004). Thus, if a subject utters an irrelevant sound during rehearsal, the rehearsal will typically be interrupted and performance impaired. Based on the data obtained through clinical work with patients – neurophysiological testing and neuroimaging – it has been concluded that auditorily presented items automatically access the phonological store, whereas visually presented items had to be subvocalized in order to gain access to the store (Baddeley, 2007). Despite the fact that articulatory suppression does not impose a heavy cognitive load, memory traces in this condition cannot be activated, which, as a result, impairs the recall (Baddeley, 1986). Interestingly, the articulatory suppression effect has been found not to occur for simultaneous interpreters (Padilla et al., 1995; Yudes et. al., 2012). Through their formal training and professional experience, interpreters become accustomed to constantly shifting their attention between the input and output. The empirical evidence suggests that interpreters are able to disregard the sound of their own voice, which essentially enables them to focus on the input and ignore the output (Yudes et. al., 2012). Exploring the effect of articulatory suppression on

simultaneous interpreters, Padilla and colleagues (1995) found that the interpreters' recall remained unaffected in this condition. In addition, Yudes and colleagues (2012) reported that the only context in which interpreters showed the effects of articulatory suppression was under extreme task demands when tested on pseudo-words.

Another well-accepted account of working memory was offered by Nelson Cowan (1988, 2014) who perceived working memory as a cognitive process capable of storing old and new information in a form that makes it easy to be accessed and manipulated with other tasks performed concurrently. According to this model, working memory is "the small amount of information that can be held in mind and used in the execution of cognitive tasks, in contrast with long-term memory, the vast amount of information saved in one's life" (Cowan, 2014: 197).

Contrary to Baddeley's model, Cowan does not think of working memory and long-term memory as separate modules, but rather as a single, highly complex mechanism embedded into and interrelated with other cognitive mechanisms. In addition, this view emphasizes the interconnectedness of various memory components. According to Cowan, working memory cannot exist on its own; it is rather a collection of processes that includes both attention and long-term memory. As such, it is intrinsically associated with "intelligence, information processing, executive function, comprehension, problem-solving, and learning" (Cowan, 2014: 197). Cowan's model suggests that a distinction be made between two components of working memory: the *focus of attention*, and the *activated part* of long-term memory (Cowan, 2014). Exposure to any stimulus normally

activates items in the long-term memory, so that each new stimulus initially gets briefly stored in a sensory store only to be sent to an activated part of long-term memory or the focus of attention. However, not all of the stimuli are focused. As the exposure continues and the number of stimuli increases, the focus of attention shifts from the first activated items to other items. The first items, however, will remain active for a limited time, usually between 20 and 30 seconds (Hiltunen, 2016). The focus of attention prioritizes the processing of one piece of information over another. Any stimulus that is not different from what has previously been experienced typically goes to the activated long-term memory, whereas novel stimuli remain within the focus of attention. Finally, the activated part of long-term memory retains the information necessary for the completion of the activated task. This part is referred to as short-term memory. The process in charge of facilitating all the cognitive processes needed for task execution that itself is capable of sending and modifying instructions is called the *central executive*.

2.5.2 Long-Term Memory

Long-term memory refers to all the knowledge and information stored in our brains, and it comprises *episodic* memory and *semantic* memory. Episodic memory is a collection of previously experienced events or episodes that exists within our own spatio-temporal framework and it allows conscious recollection of these events. Practically any perceptual event can be stored in episodic memory based on its perceptible attributes or properties; therefore, such event is always

stored based on its self-reference to the pre-existing contents in the episodic memory store (Tulving, 1972). Semantic memory, on the other hand, is not concerned with this spatio-temporal framework as it contains ideas, facts and concepts about the world we live in. This type of memory, or "mental thesaurus," is essential for language use since it holds systematically organized and accessible knowledge one has about "words and other verbal symbols, their meaning and referents, about relations among them, and about rules, formulas, and algorithms for the manipulation of these symbols, concepts, and relations" (Tulving, 1972: 386). Semantic memory in simultaneous interpreters is believed to be regularly increasing thanks to their constant acquiring of novel knowledge through both exposure and practice (Daró, 2002).

2.5.3 Memory Capacity and Fading

The different approaches to memory notwithstanding, memory in general can be perceived as having two distinct stages – short-term memory, and long-term memory – with sensory memory sometimes added as the initial stage, referring to a very short-term memory responsible for processing information received by sensory organs, and lasting typically less than one second.

One's memory capability and capacity increase over the life span; thus, it is the increase in one's knowledge and experience that is responsible for the development of memory. Memory also increases as a result of learning, during which new concepts are being formed and linked with the existing ones. Finally, memory can be expanded through training by either simply improving

functionality of its basic processes, in the same way a muscle can be developed through practice, or by discovering new task-completing strategies that are more efficient that those used up to that point (Cowan, 2014). The link between aging and brain function has been well-documented. A faster rate of brain atrophy, resulting in reductions in the volume of gray matter and, in general, cortical thickness have been reported to be larger in monolinguals than bilinguals (Del Maschio et al., 2019; Tao et al., 2021). The information that is not recalled from memory for a certain period of time may simply fade away. With short-term memory this process can occur within the matter of seconds since the last recall. One of the possible explanations for this is the limited availability of space that short-term memory needs for new incoming information. This view was first proposed by Brown (1958) and it still remains a widely debated topic among scholars. The view holds that forgetting or fading occurs as a result of memory traces losing their activation as time passes (Brown, 1958). This view also seems to be underpinning the concept of human memory consisting of two units – short-term working memory that is "limited to a small number of items or to a short period of time," and long-term memory for which such limitations do not apply (Ricker et al., 2017: 1970).

2.5.4 Memory Advantage for Interpreters

Studies typically report a working memory advantage for interpreters over non-interpreters. Specifically, interpreter training has been found to "develop more efficient comprehension strategies involving processing of larger units and

deeper semantic analyses" (Yudes et al., 2013: 1052). In addition, such training expands verbal short-term memory (Babcock et al., 2017), mental flexibility and task-switching control (Babcock & Vallesi, 2017), resulting in better interpreting performance (Chmiel, 2018). Similarly, professional interpreter training has been reported to have "a substantial effect on lexical retrieval in L1 at the semantic (or conceptual) level" (Stavrakaki et al., 2012: 631).

Using the digit span task, a reading span working memory task, and a free recall task, Padilla and colleagues (2005) tested a group of interpreters versus non-interpreters. By measuring the working memory span, as one's ability to engage in additional cognitive tasks that impose their own processing and storing demands on the working memory while a subject is engaged in a primary task, the authors reported that the first group outperformed the second one in all three memory tasks, suggesting the advantage was not inherent, but rather gained through training and experience (Padilla et al., 2005).

Somewhat different results were reported by Christoffels and colleagues (2006). Comparing Dutch-English professional interpreters against Dutch-English bilingual students and Dutch-English teachers on a word span task, a speaking span working memory task, and a reading span working memory task, the authors found that the interpreters and teachers performed similarly, but also that the two groups had a memory advantage over the students, which suggests that "word retrieval is not uniquely related to simultaneous interpreting," and that it cannot be augmented "any further by professional interpreting than by another profession that demands high proficiency in L2" (Christoffels et al, 2006). However, the

authors seemed to have overlooked the age factor of the two groups of professionals and the experience that comes with it.

Interestingly, testing professional interpreters and students of interpreting against two control groups of monolinguals and multilinguals, Köpke and Nespoulous (2006) reported contradicting results to those from previous studies by their colleagues. They did not observe an interpreter memory advantage in the digit span task and word span tasks, finding it only in the listening span task and in free recall. The authors concluded that "simultaneous interpreters do not rely on phonological rehearsal since their memory performance is disrupted to a lesser degree by articulatory suppression," suggesting additional research be done to investigate "other factors that influence performance on these tasks; specifically, the role of age and of different kinds of experience" (Köpke & Nespoulous, 2006: 14-15).

Comparing memory abilities and executive control between a group of simultaneous interpreters and two control groups consisting of language experts – one of the foreign language teachers with same level of education and professional experience as the interpreter group, and one of so-called non-experts with no professional proficiency of their second language, Hiltunen and colleagues found that simultaneous interpreters achieved superior results in free recall and dichotic listening tasks, attributing their success to their being "accustomed to dividing their attention between two channels: listening to the incoming source text and to their own voice speaking the target text, and even comparing the two" (Hiltunen et al., 2016: 307). The dichotic listening task

included babble noise added to the stimuli, and the study reported that the interpreter group was successful at resisting all external distractions, suggesting that in addition to having better memory skills the group also demonstrated superior performance of executive control in contrast to non-linguistic experts (Hiltunen et al., 2016).

Stavrakaki and colleagues (2012) examined a potential link between verbal fluency and working memory. Specifically, the authors tried to establish if proficiency in a second language was in any way related to improved working memory and phonological processing (Stavrakaki et al., 2012). In addition, the study sought to explain whether working memory and phonological processing were facilitated by training and professional experience in simultaneous interpretation (Stavrakaki et al., 2012). To that end, the authors recruited one group of professional interpreters, one of language teachers, and a control group made up of monolingual speakers that matched the other two groups on age and level of education, and tested them on a listening recall task, which tested their working memory, and a semantic task, and a phonological task, both of which were described as verbal fluency tasks. The study found that the interpreter group, significantly outperformed the other two groups on almost all tasks, suggesting that both extensive training and professional experience in simultaneous interpreting are linked with superior phonological loop processing and semantic processing, and to a degree even phonological processing (Stavrakaki et al., 2012). Interestingly, the study also found that foreign language proficiency in

itself was associated with superior working memory, and either semantic or phonological processing (Stavrakaki et al., 2012).

In addition to demonstrating a clear memory advantage of simultaneous interpreters over non-interpreters, the results of these studies are also in line with evidence from experimental psychology suggesting that given sufficient and intense practice human beings can be trained to concurrently perform several independent tasks (Lambert et al., 1995).

2.5.5 Additional Aspects of Interpreting

The very process of simultaneous interpreting, one can argue, is not a one-way process, like it has been typically thought of – a simple transfer of a message from a source language to a target language. Most of the cognitive challenges that an interpreter has to navigate in order to get the understanding of the input message have already been addressed. Importantly, there are several extralinguistic factors that ultimately affect the shape and tone of the message produced in the target language.

The most crucial one would be the context; for the context is the constant in the interpreter's mind, based on which all the incoming linguistic information is being processed (Schweda-Nicholson, 1987). In this sense, the context is a cognitive category, a set of ideas the interpreter conveys and expresses based on personal knowledge and understanding of the world, but also based on assumptions about the listener's expectations. From the perspective of Relevance Theory, it is precisely these assumptions that inform the dynamics and flexibility

of context, since context is taken to be "a subset of the individual's old assumptions, with which the new assumptions combine to yield a variety of contextual effects" (Sperber & Wilson, 1995: 132). One of the ideas this theory endorses is that incoming information is constantly being verified against selected background knowledge and assumptions, with context formation being open to other choices, but also revisions throughout the process of comprehension (Sperber & Wilson, 1995). Therefore, having an established shared cognitive environment is only a prerequisite for a successful message transfer from one language into another, for the relevance of all translation choices is being constantly re-evaluated and is subject to change at any point (Stroińska & Drzazga, 2018). In addition, it is only logical to assume that we can communicate using more than one context at a given time. For example, in situations when a person's attention is divided between two tasks, (e.g. watching TV and talking to a friend), this person's attention is thought to be switching from one context to another, a quite different one, suggesting there is a conceptual short-term memory of sorts, where contexts that are not being used for the time being, get temporarily stored in this memory (Sperber & Wilson, 1995). Finally, the context is not restricted to information about the immediate physical environment or the preceding utterances; therefore, among other things it may also contain any general assumptions, cultural or religious beliefs, past memories, or expectations about the future events (Sperber & Wilson, 1995).

Another essential factor is the speaker's intention. With this in mind, interpreters rely on make-sense approach, where they assume that the utterances

they hear in the source language make sense (Uhlenbeck, 1974; Schweda-Nicholson, 1987). Particularly, when faced with a comprehension problem or an ambiguity, interpreters must employ this make-sense approach and make use of both linguistic and extralinguistic cues available to them at that particular moment, in order to be able to deliver an accurate and complete interpretation (Schweda-Nicholson, 1987).

Finally, interpreters have also to bear in mind the make-up of the audience, and adapt their rendition accordingly. Even if there are no comprehension or ambiguity issues, and even if terminology and background knowledge pose no challenge to the interpreter, it is possible that the audience does not share the same specialist knowledge with the speaker. Most scholars inform us that a message is aimed at a target audience. However, the notion of a target audience seems to imply that such persons are only passively participating in a communication event, whereas, in reality, this notion seems to be misleading for "those who must decode a text are fully as important in the communication event as the original encoder or translator" (Nida, 1978).

2.5.5.1 Bilingualism

Processes related to both short- and long-term memories have been found to be well-developed and better organized in *experts* (Ericsson & Kintsch, 1995). Someone whose performance is consistently of exceptional quality on "a specified set of representative tasks for a domain" is considered to be an expert (Ericsson & Lehmann, 1996: 277). An expert also manages to efficiently and successfully

function in very complex environments where comprehensive knowledge has to be used in an expert manner, notwithstanding a full capacity of working memory to handle only a very limited number of items or ideas at once (Cowan, 2014). Expert level of performance typically comes with continuous practice – which is sometimes deliberate and targeted. This practice, among other things, includes activities through which one acquires, applies and organizes knowledge into specialized categories ready for future use, but also the constant modification of these categories according to feedback one receives (Hiltunen, 2016). Therefore, by continuously refining and improving one's performance along the proficiency scale, expertise in any area can be achieved and perfected. Bilingual speakers, following this definition, can be considered to be experts for their superiority over monolinguals in linguistic processing and reasoning, but also executive functioning, metalinguistic awareness, and general cognitive development and flexibility (Fox et al., 2019). Bilingualism can be defined as the use and knowledge of two languages in everyday life, including, among other things, "the presentation of information in two languages, the need for two languages, the recognition of two or more languages" (Grosjean, 2012: 5). This definition applies to typically developing children, but also adults without known neurological disorders, who, to varying degrees, use two different languages in order to meet their communicative needs (Kohnert, 2021).

Across disciplines, bilingual memory is considered to be a multifaceted construct that involves lexical and semantic processes of both native and non-native languages in proficient bilinguals, and whose components have been

reported to interact and exchange information (Dong, Gui & Macwhinney, 2005; Kroll, Dussias, Bice & Perrotti, 2015). In the same way, subdomains of different language systems such as semantics or phonology have been found to rely on their somewhat autonomous networks (Teichmann, Turc, Nogues, Ferrieux & Dubois, 2012). The differences in performance between the two groups have been suggested to be due to retrieval rather than encoding (Morrison et al., 2018). However, some of the neural circuits engaged during bilingual processing, which includes translation from one language to another, are found to be different from those involved in processing of a single language, such as word reading (Borius, Giussani, Draper & Roux, 2012). Bilinguals have typically been found to outperform monolinguals in their ability to maintain attention, as well as their superior executive control in tasks requiring attentional shifts from one stimulus to another (Rämä et al., 2018; van den Noort et al., 2019). However, some scholars reported no difference between monolinguals and bilinguals whatsoever, most notably, Hilchey and colleagues (2015), and Paap and colleagues (2015).

In a typical communication scenario, a bilingual person needs to select one language, while suppressing the other one. Such a challenge has been known to improve the executive control function (Calvo, Ibáñez & García, 2016). The ability to inhibit words in one language, be it first or second, is controlled by executive functions, however, the same degree of control is not always required by both languages (Rodriguez-Fornells, De Diego Balaguer & Munte, 2006). Switching from the dominant language to the less dominant one has been reported to be more difficult than switching the other way round for bilinguals of lower

proficiency (Costa & Santesteban, 2004). However, with extensive interpreter training and experience this asymmetrical switching cost can not only be significantly reduced, but also completely neutralized (Proverbio, Leoni & Zani, 2004).

Interpreters normally achieve fluency in at least two languages that goes far beyond that of an average bilingual person. An interpreter is expected to have "a very extensive command of his working languages, constituting if not a 'perfect' grasp then at least an extremely surefooted mastery not only of the language itself but of all aspects, social, cultural, political, etc., of the linguistic community concerned" (Henderson, 1982: 151). The inconsistency in reported accounts of bilingual advantage across studies could be explained by different degrees of language switching as well as different degrees of competing bilingual activation, suggesting that bilingualism is not a uniform category, where professional interpreters evidently demonstrate superior performance on linguistic tasks but also on other cognitive tasks, such as memory and attention tasks. This is particularly true in light of the fact that "interpreting may be regarded as a dual-task situation because of the simultaneity of comprehension and production processes," and that "the requirement to manage two languages may be related to increased cognitive control" (Christoffels et al., 2006: 340). Finally, in addition to bilingual skills and extensive professional training and experience, the level of language proficiency encountered in simultaneous interpreters "occurs through real-life interactions and experiences that take place over a considerable period of time" (Valdés & Angelelli, 2003: 62).

2.5.5.2 Perception and Production

There are four major theories or models that attempt to account for the link between perception and production: the *motor theory*, the *direct realist theory*, the *exemplar-based model of learning*, and *schema theory*.

The *motor theory* postulates that "the objects of speech perception are the intended phonetic gestures of the speaker, represented in the brain as invariant motor commands that call for movements of the articulators through certain linguistically significant configurations" (Liberman & Mattingly, 1985: 2). In other words, we recognize speech sounds by "creating a motor representation of how sounds would be produced" (Moulin-Frier & Arbib, 2013: 421). These gestures are, of course, physical realizations of phonetic notions – such as lip rounding or protrusion, jaw raising or lowering, tongue backing or lowering – and, as such, they represent the rudimentary elements of speech perception and production. Secondly, the theory argues that if both perception and production share the same group of invariants, then they must be closely linked. This link, however, is not something we learn through association; thus, it is not "a result of the fact that what people hear when they listen to speech is what they do when they speak" (Liberman & Mattingly, 1985: 3). On the contrary, this link seems to be "innately specified, requiring only epigenetic development to bring it into play" (Liberman & Mattingly, 1985: 3). Thus, the way in which we perceive gestures is in several ways different from the auditory mode responsible for perceiving speech sounds. Finally, acoustic signals get converted to articulatory gestures automatically, so that the listener can perceive phonetic information

without any intervention by the auditory manifestations that speech sounds are known to have. Having undergone extensive revisions over the years, the theory has mostly since been discarded, however, some scholars maintain that this model can account for the analysis of speech in noise where the listener makes hypotheses about the word the speaker is uttering, connecting the sounds produced by the speaker to the inventory of phonemes in the listener's native language memory store (Moulin-Frier & Arbib, 2013).

The *direct realist theory* offers a somewhat different take on perception and production, according to which "listeners use structure in acoustic speech signals as information for its causal source, as they do in perceiving generally" (Fowler et al., 2003: 398). The theory, based on Gibson's (1966) direct realist theory of perception, maintains that what listeners perceive as information for causal sources are structures in acoustic speech signals, just as they would with perception in general. Gibson established a cause-and-effect relationship between various objects and events which informs us that the acoustic energy we hear specifies its various sources (Goldstein & Fowler, 2003). Furthermore, distinctive properties of an event or object are said to "structure the media distinctively," and that way "structure in media imparted to sensory systems serves as information for its distal source" (Goldstein & Fowler, 2003: 25). The theory sees phonetic gestures produced by the vocal tract as the causal sources, therefore, "the common currency of listening and speaking that allows both perceptually guided speaking and achievement of parity is provided by the phonetic gestures that occur publicly during speech" (Fowler et al., 2003: 398). The theory holds that articulatory

gestures are perceived as distal events that "causally structure acoustic speech signals," which are "in turn, specified by them" (Goldstein & Fowler, 2003: 25). In this, the proponents of the direct realist theory see the explanation as to how children and adults alike are able to cross-modally integrate speech information.

The *exemplar-based model* of speech perception and production offers a schematic description of both processes. This model assumes that "individual speech utterances are stored in the mind as separate exemplars," which get activated during speech perception and production (Hay & Drager, 2006: 370). During speech production, exemplars get activated based on their acoustic resemblance to the incoming signal, where the speaker first selects a label, and then randomly samples the exemplar distribution for that particular label. This, in turn, activates the neighborhood of the exemplar selected, while "the average properties of this neighborhood constitute the production goal" (Pierrehumbert, 2003: 132). As a result, exemplars that are indexed with relevant contexts and social features also get activated (Hay & Drager, 2006). The exemplar theory can account for a wide array of linguistic phenomena ranging from the function of frequency to acquisition processes, sound change, style-shifting, and even changes in the phonology of an individual (Walsh et al., 2010). The input the listener receives is simply an auditorily-coded speech signal transferred on the listener's cognitive map by way of its perceptual properties (Pierrehumbert, 2003). This map "provides an analog representation of the phonetic space, with the dimensions being the many phonetic parameters which are relevant to speech perception," also containing category nodes or labels, each of which is linked with

"a frequency distribution of remembered instances of that label" (Pierrehumbert, 2003: 132). The distributions are created by storing an encoded exemplar in memory, where the potency of the representation at a particular map location depends on "the number and recency of the exemplars at that location" (Pierrehumbert, 2003: 132). As specified by this theory, each incoming stimulus is classified through a statistical choice rule that compares all possible distributions in the proximity of the incoming stimulus, eventually selecting the most optimal one (Pierrehumbert, 2003). Given a sufficient exposure, any distribution can be learned at an arbitrary location on this phonetic map. Of course, dominant in perception are high-frequency labels since they are represented more prominently on the cognitive map (Pierrehumbert, 2003).

2.5.5.3 Schema Theory

Another similar school of thought that emphasizes the importance of preexisting knowledge and experiences is Schema Theory. Schema Theory can trace its origins back to the 1920's. It pertains to the processing of both visual and auditory data, including language. According to this theory, every time we have a new experience, this experience is being compared with similar experiences stored in our memory, and processed in relation to the prototypical version of it. Using the notion of *schemata* – which are thought of as the building blocks of cognition – the theory explains the organization of various representations in human memory, and how these representations get activated in the process of understanding (Rumelhart & Ortony, 1977; Brown & Yule, 1983). The theory

specifies that these abstract building blocks “are involved in the storing of information; they are data structures representing stereotyped situations; global patterns of knowledge or generalized events stored in situational memory” (Schank, 1985, 230).

The British psychologist Frederick Bartlett is credited with first developing Schema Theory in the early 1930’s when he started using the term schema to refer to our mental organization of past experiences. The idea came from an experiment he conducted, where subjects would read short stories and then retell them days, weeks, or months later. One of the stories, set two centuries ago, was about Native American boys hunting seals and their encounter with a group of people who came up the river in canoes. When they were retelling the story, the subjects who were natives of Cambridge, England, reconstructed some parts, while adapting the others to their contemporary setting. Thus, the boys became fishermen; the canoes became boats, and the general atmosphere of the story was more reminiscent of the England of that time. The longer the interval between reading and retelling to story was, the more adapted to the English culture the narrative became. Based on these observations, Bartlett (1932) came up with the term *schema*, to describe what he believed to be the way people generalized ideas and events, but also organized this information in their memories.

There are two ways to look at schemata - either as static data storage systems that contain information about individual stereotyped topic, or as active mechanisms facilitating the process of retrieval, but also the process of inference

– both processes responsible for the manipulation of the stored representation (Hayes 1979).

Regardless of the interpretation, these building blocks of our cognition have been proposed to play a central role in regulating overall attention, perception and memory processes, which include but are not limited to allocation of processing resources, interpretation of sensory data, information retrieval from memory, and organization of planning and actions (Morton & Bekerian, 1986). Therefore, during information processing, this systematically organized knowledge gets retrieved from our memory serving as the starting point for the interpretation of discourse, while the knowledge of a particular scene may be understood as a "single, easily accessible unit stored in memory, rather than as a scattered collection of individual facts which have to be assembled from different parts of memory each time the scene is mentioned" (Zou, 2014: 206). Schema Theory, with its notion of schemata as abstract building blocks of our cognition, proved to be an immensely valuable approach when it comes to information processing, including both perception and production. Finally, it can be argued that information processing based on schema-driven model accounts for a significant portion of our overall processing capacity (Zou, 2014).

2.5.5.3.1 Types of Schemata

Various kinds of knowledge can be represented by schemata, ranging from everyday knowledge to very specific knowledge. Four types of schemata have been distinguished: *content*, *formal*, *linguistic* and *image* (Zou, 2014). Factual

knowledge, conventional values, and cultural conventions are all represented by content schemata. These schemata are responsible for specific arrangements of things in the outside world that we know through our perceptions (Zou, 2014). Making up a significant portion of our stored world knowledge, they contain selections of particular qualities, items and events (Zou, 2014). When triggered by a stimulus, content schemata engage in cognitive processing, and it is through them that we perceive all our experiences (Rumelhart & McClelland, 1986).

Formal schemata, also sometimes referred to as textual schemata, include forms, structure and knowledge of different texts types, ranging from narrative and descriptive, to expository and argumentative. In addition, they "include the understanding that each text type uses organizational patterns, language structures, levels of formality and registers differently" (Zou, 2014: 206). Generally, these schemata facilitate comprehension by helping the reader grasp the overall composition and arrangement of the text and anticipating forthcoming input (Zou, 2014).

Unlike formal schemata, which cover items at the level of discourse, language schemata contain a range of linguistic features such as vocabulary, syntactic structures, inflections, punctuation and spelling. Some scholars maintain that this type of schemata play a crucial role in cognitive processing (Segalowitz, 1986). Language schemata are of immense value when it comes to simultaneous interpreting since they considerably reduce the cognitive load, allowing the interpreter a quicker access to both languages.

Image schemata contain both visual and kinesthetic information, in other words knowledge of both shape and motion. As such, they are used to process both physical and logical relations, therefore, these organizational structures help us perceive and understand complex concepts (Zou, 2014).

CHAPTER 3: Research Questions

The literature reviewed in the previous chapter demonstrates that the presence of background noise typically affects the listener's ability to accurately process incoming auditory signals, in which case, two or more auditory stimuli compete for the listener's limited processing resources. The study presented in this thesis investigated the effects of different types of background noise on speech perception and spoken word recognition. The aim of the study was to answer the following research questions:

- Are speech perception and spoken word comprehension equally affected by noise maskers?
- What type of commonly occurring noise maskers have the most detrimental effect on speech perception and spoken word comprehension?
- How can the present findings contribute to research related to interpreting studies?

To that end, three tasks were designed and conducted – a listening span working memory task, a listening comprehension task, and speech shadowing task – all using what the literature describes as six commonly encountered types of noise maskers, added to prerecorded spoken stimuli (Schneider, Li & Daneman, 2007; Mattys, Davis, Bradlow & Scott, 2012).

CHAPTER 4: Research Methodology

The research proposal was submitted to McMaster Research Ethics Board and received ethics clearance February 3, 2021 (MREB#: 5121). Participants were recruited through SONA linguistics research participation system. The experiments were originally planned to take place in-person and at different times. On March 13, 2020, McMaster University announced that due to the "spread of COVID-19 around the world and its arrival in our own region" it had made a "decision to suspend classes" and "transition to on-line and other types of solutions" - a restriction affecting the in-person testing as well, thus, all experiments were conducted online as three parts of one study (Farrar, 2020).

4.1 Participants

Fifty undergraduate students (academic years 1-3; 33 females and 17 males) from McMaster University participated in the study for class credit. All participants were native speakers of English and reported normal hearing at the time of testing.

4.2 Stimuli

In each experiment, participants listened to seven spoken audio clips – one of which contained no background noise (*clean* condition), while the remaining six contained either informational or energetic maskers (*construction* noise, *single babble energetic*, *single babble informational*, and *multi babble energetic*), or degraded sound (*phone* and *reverb* effects) (Lecumberri et al., 2010; Mattys et al.,

2012). All the clips were edits culled from an exercises audio CD that accompanies *Cambridge Vocabulary for IELTS Advanced Band 6.5+, with Answers* (Cullen, 2012). Noise maskers added to the clips were taken from the BBC Sound Effects Library available online (BBC Sound Effects, n.d.). All the maskers were added at the signal-to-noise ratio of -5 dB, based on the results of previous studies which found that this particular ratio kept "the average intelligibility in the 45%-65% range" (Smiljanić & Bradlow, 2009). Considered to be moderate, this ratio was preferred to "ensure that listeners would not perform at ceiling in an 'easy' listening condition or at the floor at the more difficult SNR" (Van der Feest, Blanco & Smiljanić, 2019). The energetic maskers that were used were *construction* noise (drills and jackhammers), *single babble* masker (news broadcast in Mandarin, with no music or jingles), and *multi babble* masker (4-voice bar chatter in Greek). None of the participants were familiar with these two languages. The informational masker was a news broadcast in English. The clips that featured degraded sound were made by narrowing down the speech frequency bandwidth to 350-3400 Hz range (in case of the *phone* condition), or creating a reverberation effect (in case of the *reverb* condition). The reverberation time was 1s, typically found in classrooms with unfavorable acoustics and in larger conference rooms (Labia et al., 2020), while the rest of the settings read as follows: pre-delay 47 ms, attack time medium, attack shape 60%, rear shape 40%, chorus depth 10%, diffusion 0%, spread 70%, early reflections -4.0 dB, early reflections pre-delay 0.0 ms, early reflections spread 0%, high-frequency room coloration 0%, low-frequency room coloration 0%. All the mixing and effects

were done using Pro Tools 2020.5. The audio files used in all three experiments were sampled at 44.1 kHz with 16 bits per sample. All the audio files were normalized for loudness by matching their root-mean-square power, so that no difference in intensity between the stimuli would affect the outcome of the experiments. The audio stimuli were counterbalanced across conditions – after every ten participants noise effects were shuffled across the speech tracks so that the next group would hear speech tracks accompanied by different noise. This was done in order to achieve randomness and ensure that no speech material would affect the results. Participants received instructions prior to each experiment, and they also completed two (or three, if needed) practice trials in order to become familiar with the experiment, but also to adjust the volume of their headphones. The experiments were conducted in the winter and spring terms of 2021 over Zoom due to the abovementioned restrictions to in-person testing. The stimuli in the listening span working memory task and the listening comprehension task were played from the researcher's laptop via the "share audio" function in Zoom. The inbuilt soundcard was bypassed by the external Steinberg UR44C audio interface. All the monitoring on the researcher's side was done through Morgan Audio RMS11 speakers. Prior to the experiments, participants were informed that they would be using their computers and wired headphones, while their mobile phones would only be used in the third experiment for recording their responses. Before the experiments started, participants were instructed to turn off their cameras as well as any background applications that could be consuming both computer memory and bandwidth, and

keep the Zoom audio setting set to high fidelity music mode. This way no decrease in audio bitrates was expected to occur, which, in turn, might have affected the outcome of the experiments. The researcher's laptop was connected to the Internet via an ethernet cable, with the Internet speed of 1Gbps. Importantly, no participant reported connectivity issues during the experiments, nor did Zoom issue any network or connectivity warnings. All participants listened to the stimuli through their own wired headphones or earbuds set at the level they felt comfortable at.

4.3 Tasks

Participants did all three tasks in one session – roughly taking about 45 minutes per participant – with short breaks between the tasks and an option to take a break any time between the trials. The tasks always proceeded in the following order: listening span task, listening comprehension task, speech shadowing. Prior to each task, participants were explained the nature of the task and received instructions what to do. No information about the tasks was ever concealed from participants – they knew that the study investigated their speech perception under adverse conditions. In addition, before starting experimental conditions participants completed two (or three, if needed) practice trials, in order to become familiar with the task and stimuli at hand. Practice trials only used stimuli without background noise. Experimental trials were shuffled across participants, so that the order in which participants were exposed to different types of noise was counterbalanced.

4.3.1 Task 1: Listening Span Task

Short-term memory is responsible for aspects of speech understanding. It enables the listener to access previously stored information and successfully integrate it with the incoming material (Pichora-Fuller, Schneider & Daneman, 1995). The listening span task is typically used for measuring one's working memory capacity, i.e., capacity to remember verbal material while comprehending spoken sentences. The task was first introduced in a now seminal 1980 paper by Daneman and Carpenter, and has since been used in either its original form or in modified versions. One of the practical applications of this testing technique lies in the fact that short-term memory for speech is assumed to support learning processes throughout one's lifetime (Baddeley, Gathercole, & Papagno, 1998; Gathercole et al., 2007). Subsequent studies on reading comprehension (Nouwens et al., 2017), vocabulary acquisition (Masrai, 2019), and word-problem solving (Fung & Swanson, 2017) justified such an assumption. Likewise, testing one's cognitive prowess, particularly working memory, has proved to be a reliable predictor of academic aptitude and achievement (Alloway & Alloway, 2010). The verbal working memory task in general can vary in content, including words, numbers, full sentences, or larger pieces of information. In a typical listening span setting, subjects are presented with a sequence of auditory sentence stimuli, and asked to recall specific embedded items from each other either in free or serial order. The maximum number of items correctly recalled determines one's listening span. The task has proved to be a valid

indicator of scholastic aptitude in general, and language comprehension in particular (Dong et al., 2018).

In the present experiment, participants listened to five-sentence audio clips (31 to 42 seconds long; mean = 36.5 s; SD = 7.77), one for each of the seven noise conditions. They were instructed to remember the last word of each sentence in order of presentation, having to remember five target words per condition. One of the clips contained no background noise (*clean* condition), while the rest contained six different types of noise maskers: *construction* noise, *single babble energetic*, *single babble informational*, *multi babble energetic*, *phone* effect, and *reverb* effect.

4.3.2 Task 2: Listening Comprehension Task

Listening comprehension is a process in which listeners formulate meaning of the input speech based on their linguistic competence and contextual cues. In addition, during this active process, listeners focus on particular aspects of auditory input and evaluate what they hear against their existing knowledge (O'Malley et al., 1989). Specifically, competent listeners must be able to deduce the meaning of words and phrases sometimes unknown to them, access their prior knowledge in order to be able to understand the message, dismiss irrelevant information, recognize and make use of discourse markers and cohesive devices, identify the prosodic patterns, and understand the attitude and intention of the speaker (Willis, 1981). Listening comprehension can be seen as a three-stage process; the first stage being the *perception* stage, during which the listener

decodes the received message; the second being the *parsing* stage, during which "the words in the message are transformed into a mental representation of the combined meaning of the words," and, finally, the *utilization* stage, during which listeners "use the mental representation of the sentence's meaning" in order to complete a task (Anderson, 2009: 358). The listening comprehension task assesses one's ability to understand material presented in auditory mode, make inferences, and draw logical conclusions. Listening comprehension is a skill that can be tested indirectly through reading and writing, but also directly – typically answering multiple-choice questions after auditorily presented material. In the present experiment, the latter modality was used.

A model of text comprehension proposed by Kintsch and Rawson (2005) suggests that comprehension exercises can either test information at the text-based level where comprehension means understanding information explicitly given in the text, or, such exercises can test information that is only implied in the text, where prior knowledge needs to be accessed in order to make inferences. All the questions used in the current comprehension task only tested the information explicitly given in the speech recordings. There were no ambiguous or trick questions, and no inferences needed to be made.

Participants listened to seven audio clips (62 to 92 seconds long; mean = 77 s; SD = 21.21), one in each noise condition. After each clip, they were asked to answer four multiple-choice questions related to the material they had heard. One of the clips contained no background noise (*clean* condition), while the rest contained 6 different types of maskers: *construction* noise, *single babble*

energetic, *single babble informational*, *multi babble energetic*, *phone* effect, and *reverb* effect.

4.3.3 Task 3: Speech Shadowing

Auditory shadowing, or simply *shadowing*, is defined as "a paced auditory tracking task which involves the immediate vocalization of auditorily presented stimuli, *i.e.*, word-for-word repetition, *in the same language*, parrot-style, of a message presented through headphones" (Lambert, 1992: 226). In other words, it is an activity during which subjects simultaneously listen to and repeat the model's speech, trying to reproduce the phonological representations from the perceived auditory input (Kadota, 2019). Shadowing was first used to study selective attention and memory in humans (Underwood & Moray, 1971). As an effective tool for improving listening to foreign language, shadowing was then introduced to foreign language classes where it helped the students approach listening in very systematic fashion, resulting in an overall improvement their listening skills (Sumarish, 2017). When shadowing, subjects do not have enough time and cognitive capacity to process both linguistic and extralinguistic information; likewise, continuous switching between perception and comprehension is not a viable option (Kadota, 2019). During shadowing exercises, learners, especially those less proficient in a language, are focusing on the incoming information without paying much attention to its meaning. In other words, their top-down processing is obstructed in favor of bottom-up processing. As for perception skills, shadowing practice helps students perceive a broader

range of phonetic segments and details than they were able to before shadowing. As the exercises continue, students gradually become able to incorporate these new phonemes into their repertoire. Students have also been reported to create new phonetic categories in order to be able to understand unfamiliar variations (Flege, 1995). One study, which investigated native speakers' perception of non-native accents, found that after about only forty minutes of shadowing, native listeners started developing perceptual adaptation, which helped them get used to the non-native accent (Sidaras, Alexander & Nygaard, 2009). The study demonstrated that even brief exposure to an unfamiliar accent through shadowing can facilitate better understanding. Shadowing has proved to have advantages over other techniques, some of which make it suitable for use in clinical trials (Liu et al., 1997). One of the most important ones is that shadowing is a fully auditory technique, which makes it apt for testing nonproficient readers. Also, shadowing has no metalinguistic components, unlike lexical decision and phoneme monitoring tasks.

In the present study, participants shadowed seven audio clips (59 to 72 seconds long; mean = 65 s; SD = 8.48), one in each noise condition. One of the clips contained no background noise (*clean* condition), while the rest contained 6 different types of maskers: *construction* noise, *single babble energetic*, *single babble informational*, *multi babble energetic*, *phone* effect, and *reverb* effect. In order to avoid any technical difficulties typical for online environments (such as echoes, delays, sound errors due to unstable internet connection, etc.), participants were emailed the shadowing stimuli immediately prior to the third task, and they

played the stimuli from their own computers. They were instructed to delete the stimuli after the experiment.

4.4 Data Analysis

The data analysis was performed using R Statistical Software (version 4.1.2). The aim of an exploratory statistical analysis was to investigate and compare error percentages and means obtained in the three tasks. In order to generalize the obtained results, inferential statistics analysis was conducted. Since the data were counts, and not normally distributed, traditional parametric tests such as ANOVA and linear regression analysis were not used. Instead, the Shapiro-Wilk test was used to check the data distribution, and non-parametric *Wilcoxon* matched-pairs *signed rank* tests were used to compute statistical significance in error count distributions. Used for small sample sizes of 50 and less, the Shapiro-Wilk test is a statistical test used for determining how to close to a normal distribution the sample data is, where the null hypothesis would state that the variable is distributed normally, and the alternative hypothesis would state it is not (King & Eckersley, 2019: 157). The Wilcoxon signed-rank test is a non-parametric test typically used for testing differences in the means or medians of paired units (Wilcoxon matched pairs signed-rank test: Principles, n.d.). By first assigning ranks to the scores that are considered to belong to the same group, and then summing these ranks for each group, the Wilcoxon signed-rank test assumes the two samples as coming from the same population, and any difference in rank sums to be a result or a sampling error (Kerby, 2014: 1). It transforms the data to

an ordinal scale, assigning each data point an ordinal rank. In addition, the Friedman test, as a nonparametric test that is typically used for analyzing related samples, was employed for assessing whether the distributions of three or more paired groups has any statistically significant differences between them (Eisinga et al., 2017). The Friedman test is typically recommended to be used when the date measured by one-way ANOVA test does not meet the normality assumptions, or when an ordinal scale is used for measuring the dependent variable (Friedman Test in R, 2019).

CHAPTER 5: Results

5.1 Listening Span Task

The mean, median, percentage of errors, and standard deviation were calculated for each condition, as given in Table 5.1. The mean here means the average number of errors out of five words per condition (e.g. in the clean condition, participants incorrectly recalled 3.62 words out of 5 words).

Table 5.1. Mean, median, percentage of errors and standard deviation in the listening span task.

condition	mean	median	percentage	sd
clean	3.62	4.0	72.4	1.40
construction	2.76	3.0	55.2	1.45
m.b.e.	3.60	4.0	72.0	1.41
phone	3.88	4.0	77.6	1.14
reverb	3.88	4.0	77.6	1.42
s.b.e.	3.64	4.0	72.8	1.41
s.b..i	3.24	3.0	64.8	1.41

Figure 5.1 shows a plot of the error percentages and standard deviations for each experimental condition.

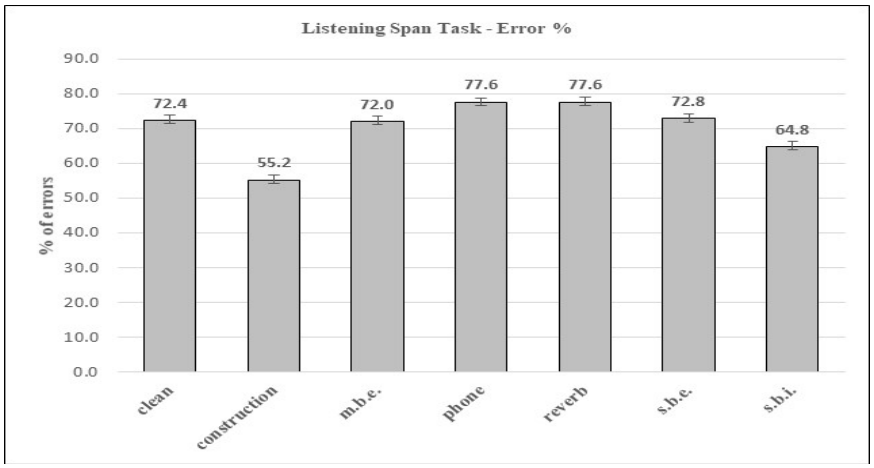

Figure 5.1. Mean error proportions out of 5 items as percentages and standard deviations in the seven noise conditions of the listening span task.

5.1.1 Comparing Error Rates in Clean vs. Other Conditions in the Listening Span Task

Using the Shapiro-Wilk test, the distribution of errors in each of the conditions was compared with a normal distribution. The results of this test are shown in Table 5.2.

Table 5.2. Results of the Shapiro-Wilk test of distribution normality in the listening span task.

condition	p-value
clean	0.000009703
reverb	0.0000002486
phone	0.000001951
construction	0.004388
single babble energetic	0.000002362
multi babble energetic	0.00000927
single babble informational	0.0004674

The Shapiro-Wilk test established that none of the conditions followed a normal distribution, which is why a non-parametric test – particularly the Wilcoxon matched-pairs signed rank test – was used to find out whether there were statistically significant differences in the error counts between the control condition (clean) and every other experimental condition. The results of this test are presented in Table 5.3.

Table 5.3. The results of the Wilcoxon matched-pairs signed rank test.

condition	p-value	z-value	p-adj[6]	p-adj-signif[7]
clean - reverb	0.2467	-1.158385	1.0000000	ns
clean - phone	0.4528	-0.750698	1.0000000	ns
clean - construction	0.0007351	-3.376159	0.0044106	**
clean - single babble energetic	0.7335	-0.340459	1.0000000	ns

[6] Adjusted p-value in multiple comparisons, arrived at by using the Bonferroni correction method available in R.
[7] Adjusted significance of the *p*-value, arrived at by using the Bonferroni correction method available in R.

clean - multi babble energetic	0.8087	-0.242152	1.0000000	ns
clean - single babble informational	0.09875	-1.650924	0.5925000	ns

Table 5.3 shows no statistically significant difference in error rate medians in the following experimental conditions: reverb, phone, single babble energetic, multi babble energetic, and single babble informational compared to the clean condition. On the other hand, a statistically significant difference was found between the clean and construction conditions, affecting the outcome of the listening span task. Table 5.1 shows that there was 17.2% fewer errors in the construction condition comparing to the control. Even though one may assume from these numbers that neither type of background noise affected the results of the listening span task, the extremely poor performance in the clean condition makes it hard to draw this conclusion. The fact that participants got only one or two items right without any noise leaves little room for noise to hurt performance. It does raise the question whether some of the participants were trying at all. The SD:s suggest that some were simply at floor in each condition, i.e. not really trying, except maybe in the construction noise condition. Alternatively, this was an extremely difficult task for this population.

5.1.2 Comparing Error Rates within the Energetic Noise Category

Since a normal distribution was not detected for the error rates, a Friedman non-parametric analysis was conducted to find out whether there were statistically significant differences in error rates between the three energetic noise conditions (construction, single babble energetic, and multi babble energetic). The Friedman test resulted in the probability value $p = 0.00001276$, below the level of

significance of 0.05, therefore, the null hypothesis assuming no differences between the energetic noise conditions was rejected. To identify the particular pairs that gave rise to the overall effect of kind of energetic noise condition the pairwise Wilcoxon signed-rank test was used for pairwise comparisons. The results of these comparisons are presented in Table 5.4.

Table 5.4. Results of the pairwise Wilcoxon signed-rank test comparing the different energetic noise conditions in the listening span task.

condition 1	condition 2	p-adj	p-adj-signif
construction	multi babble energetic	0.000558	***
construction	single babble energetic	0.00300	**
multi babble energetic	single babble energetic	1.00000	ns

Table 5.4 shows the presence of statistically significant difference between the construction and single babble energetic pair, as well as the construction and multi babble energetic pair. On the other hand, no statistically significant difference was found between the single babble energetic and multi babble energetic conditions. The statistically significant differences are visualized in Figure 5.2.

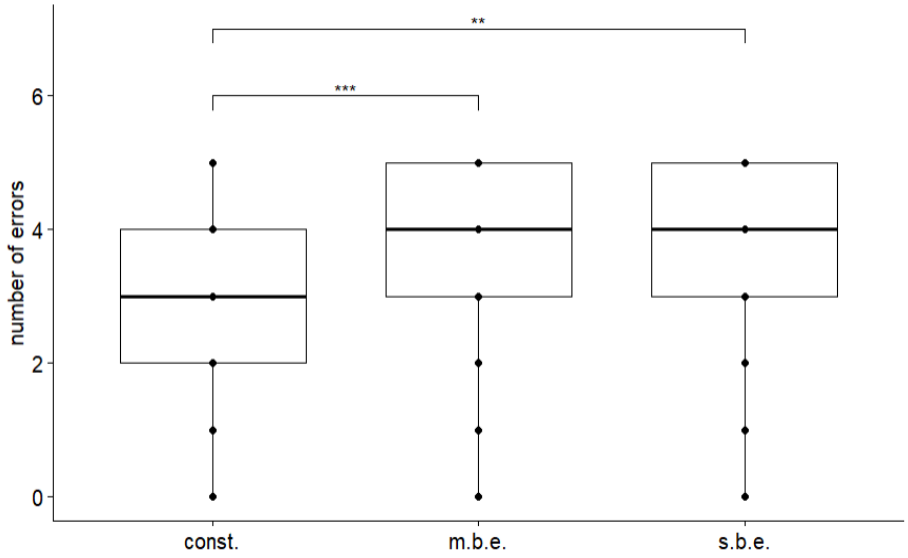

Figure 5.2. Statistically significant differences in recall error rates between energetic noise conditions revelaed by the pairwise Wilcoxon signed-rank test.

5.1.3 Comparing Error Rates within the Signal Degradation Noise Category

The Wilcoxon matched-pairs signed rank test was used to establish the presence of statistically significant differences in the error rates between the reverb and phone conditions. No statistically significant difference in the medians between the phone and reverb conditions was found as can be observed in Table 5.5.

Table 5.5. Results of the pairwise Wilcoxon signed-rank test comparing error rates between the two noise conditions with signal degradation.

condition 1	condition 2	p-value	z-value
reverb	phone	0.8201	0.2273

5.1.4 Comparing Error Rates between the Three Noise Categories

Finally, the Wilcoxon matched-pairs signed rank test was used in order to establish the presence of statistically significant differences in the recall error rates between the three categories of noise masking: degradation, energetic and informational.

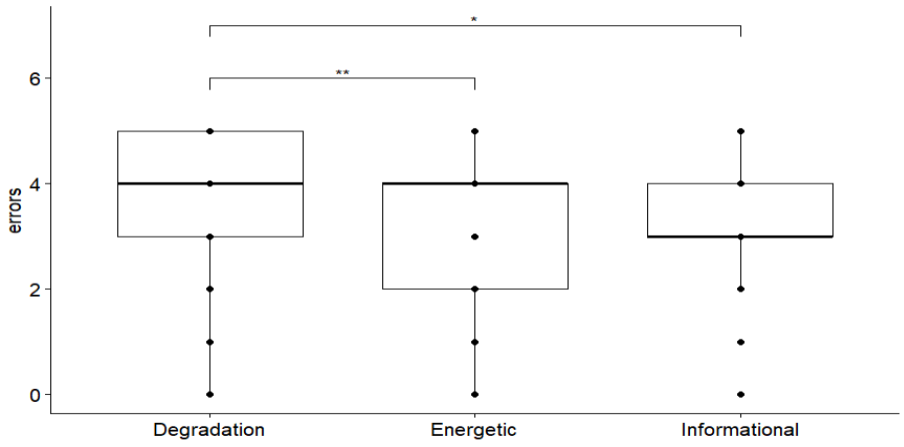

Figure 5.3. Statistically significant differences in error rates revealed by the pairwise Wilcoxon signed-rank test.

The test found statistically significant differences between the degradation and energetic categories, as well as between the degradation and informational categories. The error rate was higher when the noise masker belonged to the degradation category (reverb and phone) compared to the two other noise categories, which did not significantly differ between each other (Table 5.6).

Table 5.6. Results of the pairwise Wilcoxon signed-rank test for the between-group significance.

condition 1	condition 2	p-adj	p-adj-signif
degradation	energetic	0.002	**

degradation	informational	0.005	*
energetic	informational	0.536	ns

5.1.5 Discussion

None of the noise conditions in the listening span task had normal distribution of error rates ($p < 0.05$; see Table 5.2). Statistically, none of the maskers significantly lowered the performance compared to the clean baseline condition ($p > 0.05$; see Table 5.3). However, the uniform results across the conditions suggest that participants performed at a floor level, which could be due to the difficulty of the task itself. Surprisingly, participants scored 17.2% better in the construction noise condition compared to the clean condition (see Table 5.1). Overall, the phone and reverb conditions were most detrimental to the performance, and the participants made a greater number of errors in these two conditions (77.6% in both conditions; see Table 5.1). When the two conditions were compared to each other as the only representatives of the signal degradation category of noise maskers, no significant difference was found between them. As for the energetic maskers alone, both multi babble energetic and single babble energetic were more detrimental than the construction masker (see Table 5.4). The between-group comparison found the degraded sound significantly more detrimental than both energetic and informational masking (Table 5.6). Finally, as the construction noise condition had a significantly lower error rate than the clean condition, and also absolutely lower than the other noise conditions, it could be speculated that this type of noise forced the listener to pay attention in an otherwise boring task.

5.2 Listening Comprehension Task

The mean and median error rates, percentage of error and standard deviation of error rate were calculated for each noise condition (Table 5.7).

Table 5.7. Mean and median error rates, percentage of errors and standard deviations of error rates in the listening comprehension task.

condition	mean	median	percentage	sd
clean	1.34	1.0	33.5	0.75
construction	2.62	3.0	65.5	0.90
m.b.e.	1.56	1.5	39.0	0.93
phone	2.72	3.0	68.0	0.90
reverb	1.80	1.0	45.0	0.90
s.b.e.	2.16	2.0	54.0	0.84
s.b.i.	1.80	2.0	45.0	1.03

A plot showing the error percentages and standard deviations for each experimental condition is seen in Figure 5.4.

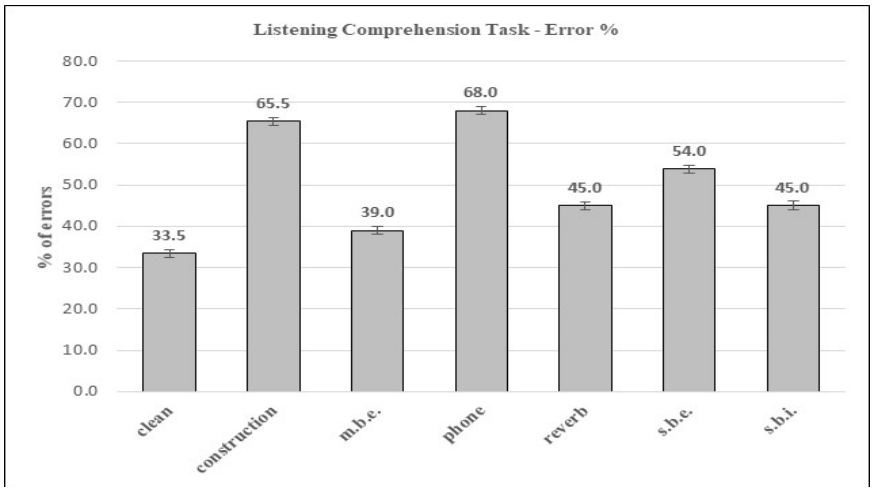

Figure 5.4. Mean error percentages and their standard deviations in the listening comprehension task.

5.2.1 Comparing Error Rates in the Clean Condition with Other Conditions in the Listening Comprehension Task

Using the Shapiro-Wilk test, the distributions of error rates in each of the conditions were checked. The results of this test are shown in Table 5.8.

Table 5.8. Results of the Shapiro-Wilk test for difference from normality in the listening comprehension task.

condition	p-value
clean	0.000003126
reverb	0.0001071
phone	0.00007214
construction	0.00009324
single babble energetic	0.00002763
multi babble energetic	0.00009585
single babble informational	0.001248

The Shapiro-Wilk test found that none of the experimental conditions had a normal distribution, so the Wilcoxon non-parametric paired tests for significance were conducted to test differences between the different noise conditions and the clean condition. Table 5.9 shows that statistically significant differences in medians were found in the phone, construction, and single babble energetic conditions when compared with the control (clean signal) condition[8]. The noise conditions were in all significant cases associated with higher error rates.

[8] Since multiple comparisons were performed, the Bonferroni correction was applied – resulting in the *p*-value being 0.05/6 = 0.0083.

Table 5.9. The results of the of the listening comprehension task: pairwise comparisons between error rates in the clean and six noise conditions using the Wilcoxon matched-pairs signed rank test.

condition	p-value	z-value	p-adj	p-adj-signif
clean - reverb	0.005418	-2.781099	0.032508	ns
clean - phone	0.00000006453	-5.405777	0.00000038	***
clean - construction	0.0000002061	-5.193769	0.0000012	***
clean - single babble energetic	0.000006952	-4.495166	0.000041	***
clean - multi babble energetic	0.1635	-1.393406	0.98100	ns
clean - single babble informational	0.002622	-3.008873	0.015732	ns

5.2.2 Comparing Error Rates within the Energetic Noise Category

Since no normal distribution was found, the Friedman non-parametric test was conducted to find out whether there were statistically significant differences in the error rates between the three energetic noise conditions (construction, single babble energetic, and multi babble energetic). The Friedman test resulted in the probability value p = 0.00001189, below the level of significance of 0.05. Therefore, the null hypothesis assuming no differences between the energetic noise conditions was rejected. In order to identify statistically significant differences between individual noise conditions, the pairwise Wilcoxon signed-rank test was used. The results are presented in Table 5.10.

Table 5.10. Results of pairwise comparisons beween energetic noise conditions in the listening comprehension task, using the Wilcoxon signed-rank test.

condition 1	condition 2	p-adj	p-adj-signif
construction	multi babble energetic	0.00000777	****
construction	single babble energetic	0.0670	ns
multi babble energetic	single babble energetic	0.00600	**

Table 5.10 lists the presence of statistically significant differences between the construction and multi babble energetic noise pair, as well as the single babble energetic and multi babble energetic noise pair. The statistically significant differences are illustrated in Figure 5.5.

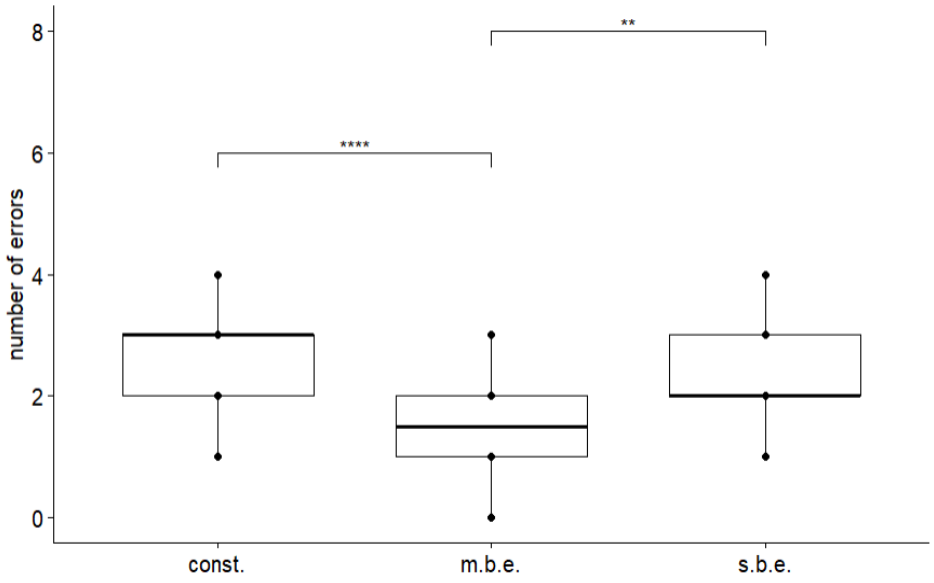

Figure 5.5. Error rates in energetic noise conditions in the listening comprehension task: Statistically significant differences revealeded by the pairwise Wilcoxon signed-rank test.

5.2.3 Comparing Error Rates within the Signal Degradation Noise Category

The Wilcoxon matched-pairs signed rank test was used to detect the presence of statistically significant differences in error rates between the reverb and phone conditions in the listening comprehension task. Statistically significant differences in the medians between the phone and reverb conditions were found as can be observed in Table 5.11.

Table 5.11. Results of the pairwise Wilcoxon signed-rank test between reverb and phone noise conditions the degradation category.

condition 1	condition 2	p-value	z-value
reverb	phone	0.00001031	-4.410494

5.2.4 Comparing Error Rates between the Three Noise Categories

Finally, the Wilcoxon matched-pairs signed rank test was used to probe for the presence of statistically significant differences in the error rates between the noise categories (Figure 5.6).

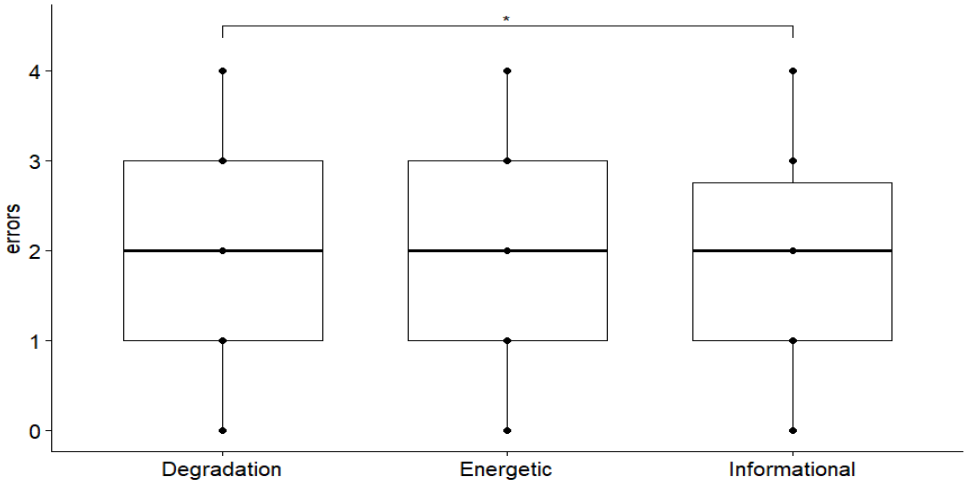

Figure 5.6. Statistically significant differences in error rates between noise categories based on the pairwise Wilcoxon signed-rank test.

Given that all three categories look almost the same in the figure, their means, medians and standard deviations are reported in the table 5.12.a.

Table 5.12.a. The means, medians and standard deviations for the pairwise Wilcoxon signed-rank test for the between-noise category significance in error rates.

condition	mean	median	sd
degradation	2.26	2	1.006
informational	1.80	2	1.02
energetic	2.11	2	0.98

The test found statistically significant differences between the degradation and informational noise categories, with better performance in the informational category (Table 5.12.b).

Table 5.12.b. Results of the pairwise Wilcoxon signed-rank test for the between-noise category significance in error rates.

condition 1	condition 2	p-adj	p-adj-signif
degradation	energetic	0.289	ns
degradation	informational	0.015	*
energetic	informational	0.068	ns

5.2.5 Discussion

In the analysis of the results from the listening comprehension task, none of the conditions had normal distribution of the error rates ($p < 0.05$; see Table 5.7). Importantly, all the maskers significantly impaired the performance ($p < 0.05$) except the multi babble energetic ($p > 0.05$; see Table 5.8). Overall, participants made most errors in the phone and construction noise conditions (68.0% and 65.5% respectively; see Table 5.7). Within the energetic noise category, statistically significantly poorer performance was found for the construction compared to the multi babble noise (26.5%); as well as for the single-babble compared to the multi-babble noise (15%; see Table 5.7). When it comes to the signal degradation noise category, participants made 23% fewer errors in the reverb condition than in the phone condition (Table 5.7). The

between-group comparison found statistical significance only between the degradation and informational noise categories, with lower overall performance in the degradation category (Table 5.12.b).

5.3 Speech Shadowing

Due to the difference in the possible maximum number of errors in each condition (for each stimulus had a different number of words), scaling to error percentages was first performed to fit all the error counts to the same 0 to 100 range. The mean, median, percentage of errors, and standard deviation were calculated for each condition, as given in Table 5.13.

Table 5.13. Mean and median percentages of errors and their standard deviations in the different noise conditions in the speech shadowing task.

condition	mean percentage	median	sd
clean	3.84	3.0	3.87
construction	7.85	6.2	7.61
m.b.e.	7.48	5.7	6.51
phone	5.68	4.0	5.04
reverb	9.62	6.7	8.68
s.b.e.	5.14	3.7	5.32
s.b.i.	10.39	7.9	8.74

A plot, showing the error percentages and standard deviations for each experimental condition is shown in Figure 5.7.

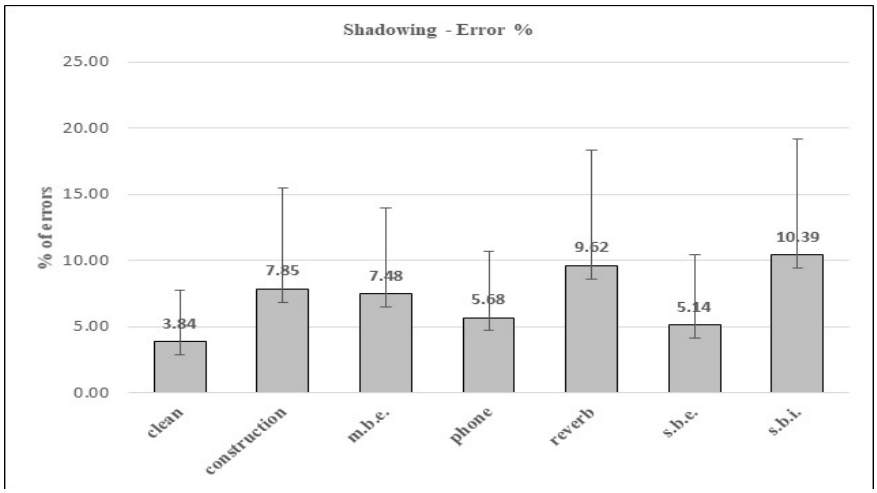

Figure 5.7. Mean percentages and standard deviations in the speech shadowing task.

Using the Shapiro-Wilk test, the normality of the distribution of errors in each of the conditions was checked. The results of this test are shown in Table 5.14.

Table 5.14. Results of the Shapiro-Wilk test for deviation from normality in the speech shadowing task.

condition	p-value
clean	0.000000124
reverb	0.0001012
phone	0.000002242
construction	0.000001214
single babble energetic	0.0000004236
multi babble energetic	0.0005912
single babble informational	0.00006694

The Shapiro-Wilk test found that none of the experimental conditions had normal error rate distributions, so the Wilcoxon non-parametric paired test for significance was used for comparisons between the clean condition that had the lowest error rate and the different noise conditions. The results of this test are presented in Table 5.15.

Table 5.15. Pairwise comparisons of error rates in the speech shadowing task between the noise conditions and the clean condition (Wilcoxon matched-pairs signed rank).

condition	p-value	z-value	p-adj	p-adj-signif
clean - reverb	0.00000085	-4.923258	0.0000051	***
clean - phone	0.005347	-4.121998	0.032082	ns
clean - construction	0.0000853	-3.928999	0.0005118	***
clean - single babble energetic	0.04175	-2.035978	0.25050	ns
clean - multi babble energetic	0.00003756	-4.121998	0.0002253	***
clean - single babble informational	0.000000026	-5.560321	0.00000016	***

Table 5.15 shows the results of the significance test. Only speech shadowing performance in the phone condition and single babble energetic condition was not significantly poorer than in the clean condition (with the Bonferroni-corrected *p*-value of 0.0083).

5.3.1 Comparing Error Rates within the Energetic Noise Category

Since a normal distribution of error rates was not found, the Friedman non-parametric test was conducted to test for significant differences between the three energetic conditions. The Friedman test indicated that statistically significant main effect of noise condition did exist. In order to successfully identify pairs of conditions contributing to this main effect, the pairwise Wilcoxon signed-rank test was used. The results are presented in Table 5.16.

Table 5.16. Results of the pairwise comparisons in error rates between noise conditions in the energetic noise category (Wilcoxon signed-rank test) in the shadowing task.

condition 1	condition 2	p-adj	p-adj-signif
construction	multi babble energetic	0.142	ns
construction	single babble energetic	0.009	**
multi babble energetic	single babble energetic	0.018	*

Table 5.16 illustrates the presence of a statistically significant difference between the construction and single babble energetic pair, as well as the single babble energetic and multi babble energetic pair. These differences reflect better shadowing performance in the single babble condition compared to the two other noise conditions. The statistically significant differences could also be observed in Figure 5.8.

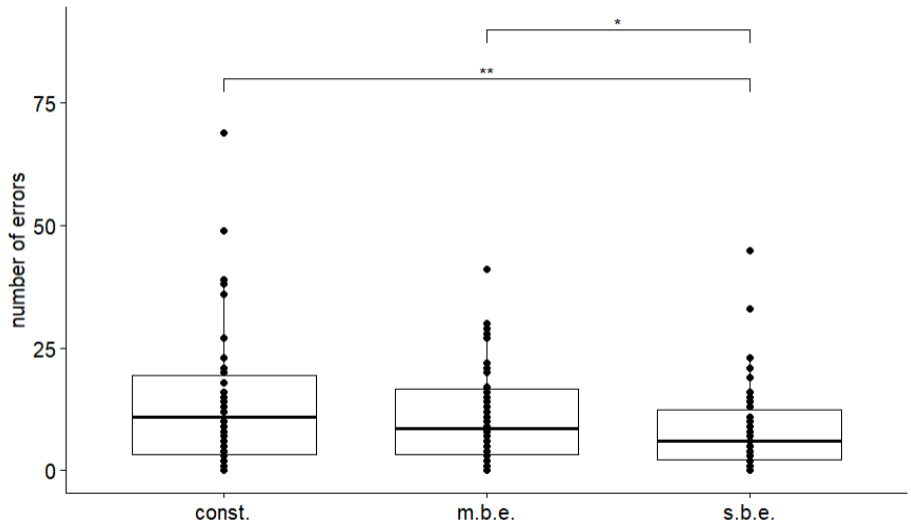

Figure 5.8. Statistically significant differences discovered by the pairwise Wilcoxon signed-rank test

5.3.2 Comparing Error Rates within the Signal Degradation Noise Category

As normal distributions of error rates were not found, the Wilcoxon matched-pairs signed rank test was used to detect the presence of statistically significant differences in the error rates between the reverb and phone conditions, which was confirmed by the *p*-value of 0.005849 (Table 5.17).

Table 5.17. Results of the pairwise Wilcoxon signed-rank test in the degradation group

condition 1	condition 2	p-value	z-value
reverb	phone	0.005849	2.52113

5.3.3 Comparing Error Rates between the Three Noise Categories

Finally, the Wilcoxon matched-pairs signed rank test was used to detect the presence of statistically significant differences in the error rates between the noise categories (Figure 5.9).

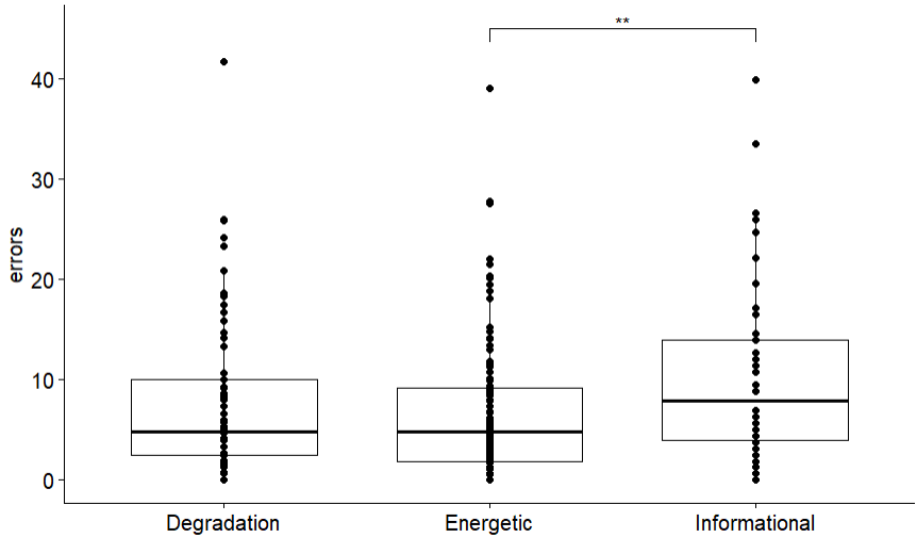

Figure 5.9. Statistically significant differences between error rates in the three noise categories (pairwise Wilcoxon signed-rank test).

The test found a statistically significant difference (with the Bonferroni-corrected *p*-value of 0.0167) between the energetic and informational noise categories, indicating better shadowing performance in the energetic noise conditions (Table 5.18).

Table 5.18. Results of the pairwise Wilcoxon signed-rank test for the between-noise categories error rates in speech shadowing.

condition 1	condition 2	p-adj	p-adj-signif
degradation	energetic	0.411	ns
degradation	informational	0.029	ns
energetic	informational	0.003	**

5.3.4 Discussion

In all experimental conditions maskers showed statistically significant effects in their ability to impair speech perception compared with the clean (control) condition ($p < 0.05$; see Table 5.15). Within the energetic noise category, statistically significant differences were observed between the construction – single-babble energetic, and multi-babble – single-babble energetic pairs (2.72% and 2.34% respectively; see Table 5.13). As for the degradation noise category, participants made 3.49% more errors in the reverb condition (Table 5.13). The between noise category comparison found statistical significance only between the informational and energetic noise categories, with poorer shadowing performance inthe informational noise conditions (Table 5.18).

5.4 Types of Errors in the Listening Span Task

Calculating errors in a span task is a rather straightforward operation, for the object of calculation is "the number of items that a person is able to recall from a list" (Imhof, 2018: 395). Three types of errors were observed in this experiment – omissions, items recalled in wrong order, and wrong target words. Furthermore, wrong target words were either words that were in the stimulus, but were not target words (e.g., the word *tractors* in the sentence *The use of heavy*

machinery like tractors can compact the ***soil.***), or words not appearing in the stimulus, but either semantically or phonologically related to the topic discussed (e.g., *The use of heavy* machines *like tractors can compact the soil.* instead of *The use of heavy* machinery *like tractors can compact the soil.*)

5.5 Types of Errors in the Listening Comprehension Task

Scoring a listening comprehension task with multiple-choice answers is a straightforward operation, where the overall score would reflect the number of correct answers. All the expected answers were explicitly given in the speech recordings, and no inferences needed to be made.

5.6 Types of Errors in Speech Shadowing

The coding of errors followed two nomenclatures: the first one used by Marslen-Wilson (1985), identifying omissions (or missing targets – further subdivided into the categories of *function* and *content* words for more detailed breakdown), constructive errors (substituted parts of or whole target words) and delivery errors (stutters, slurs and unintelligible pronunciations); and the second one used by Healey & Howe (1987) for identifying insertions. All the types of errors were found in all conditions, the only difference being their percentual representation. Table 5.19. shows the percentages of all errors in each condition. The overall lexical density (which is the number of content words divided by the total number of words in all seven speech excerpts) of the speech samples used in

the shadowing task was 51.53%, meaning that the ratio of content vs. function words across the samples was very close to being even.[9]

Table 5.19. Percentages of error types across the conditions in the speech shadowing task.

condition	omissions (function) %	omissions (content) %	constructive errors %	delivery errors %
clean	32.92	52.83	11.54	2.71
reverb	36.07	39.14	20.92	3.87
phone	34.10	42.53	19.43	3.94
construction	32.55	51.37	12.12	3.96
s.b.e.	45.17	38.57	13.47	2.79
m.b.e.	39.88	42.27	15.56	2.29
s.b.i.	36.53	40.42	19.22	3.83

The distribution of insertions across the conditions can also be seen in Figure 5.10.

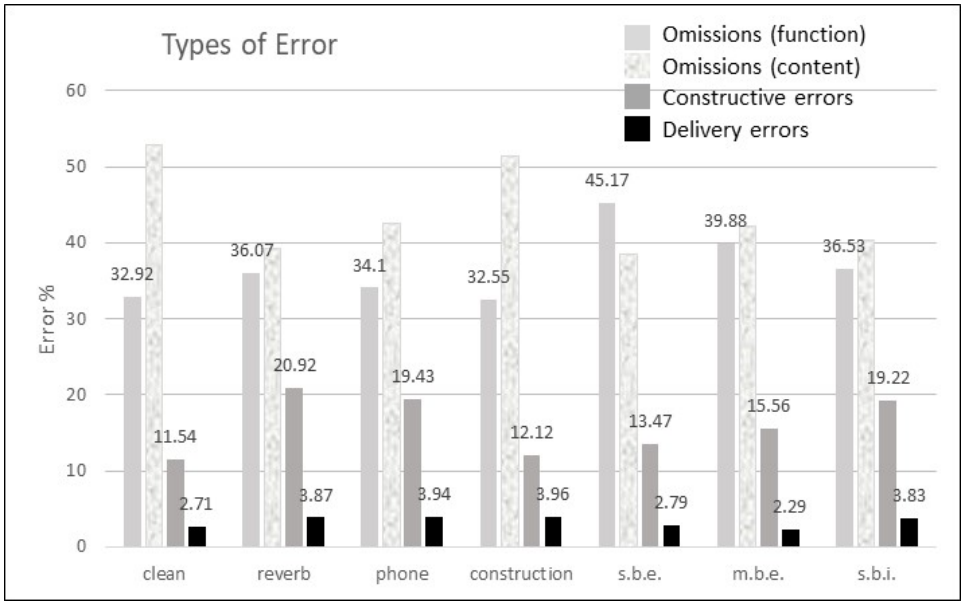

Figure 5.10. Distribution of error types across the conditions.

[9] The lexical density calculator used is available here: https://www.analyzemywriting.com/index.html

Table 5.20. Breakdown of content and function words for all seven stimulus excerpts used in the shadowing task.

CONTENT Words (%)				**FUNCTION Words (%)**			
nouns	**adjectives**	**verbs**	**adverbs**	**prepositions**	**pronouns**	**aux. verbs**	**determiners**
25.85	7.63	12.21	5.83	12.66	5.21	6.55	24.06

Thus, errors included in the analysis consisted of omitted target words, mispronounced target words, which were typically singular nouns instead of target plurals (*chair* instead of *chairs*) or vice versa (*theories* instead of *theory*), near homophones (*crash* instead of *crack*), or completely different words instead of targets (*jumpsuits* instead of *jackets*).

The next group of errors that deviated from the original stimuli were self-corrections. Two types of self-corrections were observed, and classified based on whether they were correct or incorrect. *Correct* self-corrections were repeated smaller chunks that included corrected renditions of previously mispronounced target words (*at the rear of the office* building*…of the office* block instead of *at the rear of the office block*) and as such were not counted as errors, since the target was successfully identified and the overall meaning not affected. On the other hand, any self-correction resulting in an incorrect target (such as *while the loud bang you* heard*… you* heard instead of *while the loud bang you hear*) was labeled as an error, and classified as either function or content word error based on the category it belonged to.

Finally, one type of errors which was analyzed separately belongs to the category of additions or insertions (Healey & Howe, 1987). Apart from four instances of content word insertions (Tables 5.62), this category contained only

function words (*we're going to expand* on *the budget* instead of *we're going to expand the budget*) and auxiliaries (*most people assume balloons* would *make a loud bang* instead of *most people assume balloons make a loud bang*). The reason for analyzing additions separately is a simple one: the total number of errors per condition was calculated by deducting omitted or mispronounced targets in that condition, in which case including any additional words into the analysis would not fit into this number, warranting a separate analysis of this type of errors. The following section provides examples of errors found across the conditions.

5.6.1 Omissions

5.6.1.1 Single Function Word Omissions

The following examples illustrate single function word omissions across the conditions (Tables 5.21-27).

Table 5.21. Examples of a single function word omission observed in the clean condition.

Original	Transcription (clean condition)
Another problem was that these planes couldn't Instead, physics has shown us and this was measured with an IQ test.	Another problem was that _____ planes couldn't Instead, physics ____ shown us and this was measured with __ IQ test.

Table 5.22. Examples of a single function word omission observed in the reverb condition.

Original	Transcription (reverb)
we notice the unwritten rules and you need a chair that can swivel the loud bang you hear is actually a sonic boom!	we notice ___ unwritten rules and you need a chair that ___swivel the loud bang you hear is actually ___ sonic boom!

Table 5.23. Examples of a single function word omission observed in the phone condition.

Original	Transcription (phone)
And then the final two intelligences are	And then the final two intelligences ___

interpersonal and intrapersonal	interpersonal and intrapersonal
In the past, people were seen as	In ___ past, people were seen as
One of the main factors in ensuring a harmonious society is that there are	One of the main factors in ensuring a harmonious society is ____ there are

Table 5.24. Examples of a single function word omission observed in the construction condition.

Original	**Transcription (construction)**
And despite the fact that our competitors	____ despite the fact that our competitors
The next kind of intelligence is logical	____ next kind of intelligence is logical
kind of intelligence is logical and this is used	kind of intelligence is logical and ____ is used

Table 5.25. Examples of a single function word omission observed in the s.b.e. condition.

Original	**Transcription (s.b.e.)**
whose strengths are in subjects such as	whose strengths are ____ subjects such as
I'd like to take a look at how it all began	I'd like to take a look ___ how it all began
But we believe a picture in the newspapers	____ we believe a picture in the newspapers

Table 5.26. Examples of a single function word omission observed in the m.b.e. condition.

Original	**Transcription (m.b.e.)**
where the skaters manage to spin incredibly fast	where ____ skaters manage to spin incredibly fast
but today I'd like to take a look at how it all began	____ today I'd like to take a look at how it all began
there is a blending of these customs	there is a blending of ____ customs

Table 5.27. Examples of a single function word omission observed in the s.b.i. condition.

Original	**Transcription (s.b.i.)**
Instead, physics has shown us the loud bang	Instead, physics has shown __ the loud bang
but don't tell the students you've done this	but ____ tell the students you've done this
you should always head towards the main stairs	you should always head towards ____ main stairs

5.6.1.2 Function Word Omissions as Parts of Larger Sequences

The missing function words illustrated in the following examples (Tables 5.28-34) were counted as single function word omissions, with each word in an

omitted sequence labeled individually as either omitted function word or omitted content word

Table 5.28. Examples of a function word omission as part of a larger sequence observed in the clean condition.

Original	Transcription (clean condition)
according to our accepted standards of behavior	according to our accepted standards __ _____ .
so the loud bang you hear is actually a sonic boom!	so the loud bang you hear is actually _____ boom!
We can observe this principle in the real world	We can observe this principle in _____ world

Table 5.29. Examples of a function word omission as part of a larger sequence observed in the reverb condition.

Original	Transcription (reverb)
Those who fail to observe these norms are	Those who fail _____ these norms are
as our statistics show that it's just not cost-effective.	as our statistics show that it's ___ ___ cost-effective.
the products are now selling well nationally in department stores,	the products _____ selling well nationally in department stores,

Table 5.30. Examples of a function word omission as part of a larger sequence observed in the phone condition.

Original	Transcription (phone)
if they hold them close to their body, making themselves narrower,	if they hold them close _____ _____, making themselves narrower,
rather than just in our local shop here	rather than just _____ local shop here
First of all, can I just remind you that if you hear the fire alarm,	First of all, can I just remind you that _____ the fire alarm,

Table 5.31. Examples of a function word omission as part of a larger sequence observed in the construction condition.

Original	Transcription (construction)
We can observe this principle in the real world in the sport of ice skating	We can observe this principle _____ in the sport of ice skating
different types of intelligence and these are reflected in your	different types of intelligence __ _____ reflected in your
we believe a picture in the newspapers will be	we believe _____ the newspapers

much more	will be much more

Table 5.32. Examples of a function word omission as part of a larger sequence observed in the s.b.e. condition.

Original	**Transcription (s.b.e.)**
they will slow down and if they hold them close to their body	they will slow down _____ _____ close to their body
But today I'd like to take a look at	But today _____ take a look at
rather than just in our local shop here	rather than just in _____ shop here

Table 5.33. Examples of a function word omission as part of a larger sequence observed in the m.b.e. condition.

Original	**Transcription (m.b.e.)**
a great experiment for demonstrating an important principle of	a great experiment for demonstrating _____ principle of
The first of these is termed linguistic	The first _____ is termed linguistic
who fail to observe these norms are inevitably excluded from that group.	who fail to observe these norms _____ excluded from that group.

Table 5.34. Examples of a function word omission as part of a larger sequence observed in the s.b.i. condition.

Original	**Transcription (s.b.i.)**
is used to describe people whose strengths are in subjects such as	is used to describe people _____ in subjects such as
Ask one of your students to sit on the chair holding the weights	Ask one of your students _____ _____ holding the weights
second experiment is always fun as it involves a balloon!	second experiment is always fun as _____

5.6.1.3 Single Content Word Omissions

The following examples (Tables 5.35-3.41) illustrate single content word omissions across the conditions.

Table 5.35. Examples of a single content word omission observed in the clean condition.

Original	**Transcription (clean condition)**
next experiment is called the arm engine and for this one	next experiment is called the arm _____ and for this one

the fact that our competitors advertise baby clothes on TV,	the fact that our competitors _____ baby clothes on TV,
enough seats to make passenger traffic profitable.	enough seats to make _____ traffic profitable.

Table 5.36. Examples of a single content word omission observed in the reverb condition.

Original	Transcription (reverb)
an important principle of energy and momentum.	an important principle of energy and _____.
in the real world in the sport of ice skating,	in the real world in the sport of __ skating,
our competitors advertise baby clothes on TV,	our competitors advertise baby _____ on TV,

Table 5.37. Examples of a single content word omission observed in the phone condition.

Original	Transcription (phone)
we won't be using this method,	we won't be _____ this method,
these people are often labeled conservative.	these people are often _____ conservative.
so the loud bang you hear is actually a sonic boom!	so the loud bang you hear is _____ a sonic boom!

Table 5.38. Examples of a single content word omission observed in the construction condition.

Original	Transcription (construction)
will be much more attractive to potential customers.	will be much more attractive to _____ customers.
The multiple intelligence theory first came to light	The _____ intelligence theory first came to light
through such people that our heritage is preserved,	through such people that our _____ is preserved,

Table 5.39. Examples of a single content word omission observed in the s.b.e. condition.

Original	Transcription (s.b.e.)
to see how elastic they are and	to see how _____ they are and
and how much stress can be put on them.	and how much _____ can be put on them.
patterns in the way we conduct ourselves.	patterns in the way we _____ ourselves.

Table 5.40. Examples of a single content word omission observed in the m.b.e. condition.

Original	Transcription (m.b.e.)
who fail to observe these norms are inevitably excluded	who fail to observe these _____ are inevitably excluded
the hole expands rapidly forming a catastrophic crack.	the hole expands _____ forming a catastrophic crack.
the hole expands rapidly forming a	the hole expands rapidly forming a

catastrophic crack.	_____ crack.

Table 5.41. Examples of a single content word omission observed in the s.b.i. condition.

Original	Transcription (s.b.i.)
who are more interested in the written word and reading.	who are more _____ in the written word and reading.
which describes people who are attracted by or drawn to images.	which describes _____ who are attracted by or drawn to images.
is always fun as it involves a balloon!	is always fun as it _____ a balloon!

5.6.1.4 Content Word Omissions as Parts of Larger Sequences

The following examples (Tables 5.42-49) illustrate content word omissions as part of a larger sequence. The rest of the words missing in these sequences were labeled individually as either omitted function word or omitted content word.

Table 5.42. Examples of a content word omission as part of a larger sequence observed in the clean condition.

Original	Transcription (clean condition)
There are those who observe these social mores religiously,	There are those who _____ _____ religiously,
The first of these were provided by some of the airmail services	The first of these _____ _____ airmail services
we believe a picture in the newspapers will be much more attractive	we believe _____ newspapers will be much more attractive

Table 5.43. Examples of a content word omission as part of a larger sequence observed in the reverb condition.

Original	Transcription (reverb)
this describes people who are more interested in the written word and reading.	this describes people who are _____ ___ the written word and reading.
these planes couldn't carry enough seats to make passenger traffic profitable.	these planes couldn't carry enough seats to make passenger _____.
we won't be using this method, as our statistics show that it's just not cost-effective.	we won't be using this method, _____ _____ it's just not cost-effective.

Table 5.44. Examples of a content word omission as part of a larger sequence observed in the phone condition.

Original	Transcription (phone)
We can observe this principle in the real world in the sport of ice skating,	We can observe this principle in the real world _____ _____,
the products are now selling well nationally in department stores,	the products are now selling well _____ stores,
our competitors advertise baby clothes on TV,	our competitors advertise ____ _____ on TV,

Table 5.45. Examples of a content word omission as part of a larger sequence observed in the construction condition.

Original	Transcription (construction)
we expect people to behave according to our accepted standards of behavior.	we expect people to behave according to our _____ _____.
Most people assume balloons make a loud bang when the air is released	Most people assume _____ _____ when the air is released
In the real world, this principle is used to test different materials	In the real world, _____ _____ different materials

Table 5.46. Examples of a content word omission as part of a larger sequence observed in the s.b.e. condition.

Original	Transcription (s.b.e.)
if you pierce the balloon through the sticky tape, instead of bursting it,	if you pierce the balloon _____ ___ _____ _____, instead of bursting it,
If we enter a new group, we notice the unwritten rules and social norms	If we enter a new group, we notice ___ _____ and social norms
couldn't fly over mountains, so passengers took trains for part of their journey.	couldn't fly over mountains, so passengers _____ part of their journey.

Table 5.47. Examples of a content word omission as part of a larger sequence observed in the m.b.e. condition.

Original	Transcription (m.b.e.)
Those who fail to observe these norms are inevitably excluded from that group.	Those who fail to observe _____ _____ ____ _____ excluded from that group.
Those who fail to observe these norms are inevitably excluded from that group.	Those who fail to _____ _____ _____ ___ inevitably excluded from that group.
But today I'd like to take a look at how it all began.	But today ___ ___ __ _____ __ ____ __ how it all began.

Table 5.48. Examples of a content word omission as part of a larger sequence observed in the s.b.i. condition.

Original	Transcription (s.b.i.)
It's actually through such people that our heritage is preserved,	It's actually through such _____ _____ _____ _____ is preserved,
different types of intelligence and these are reflected in your personality.	different types of intelligence and these are _____ ____ _____ _____.
The second experiment is always fun as it involves a balloon!	The second experiment is ____ ___ __ _____ involves a balloon!

5.6.1.5 Constructive Errors Substituting Parts of Words or Whole Words

The constructive errors substituting parts of words or whole words illustrated in the following examples (Tables 5.49-55) were counted as single constructive errors, while the missing words immediately preceding or following these constructive errors in some of these sequences were counted as either function or content word omissions depending on which category the omitted word belonged to.

Table 5.49. Examples of constructive errors substituting parts of words or whole words observed in the clean condition.

Original	Transcription (clean condition)
there are clear established patterns in the way	there are clear established answers in the way
in the sport of ice skating,	in the sport of figure skating
multiple intelligence theory first came to light in	multiple intelligence theory first came to life in

Table 5.50. Examples of constructive errors substituting parts of words or whole words observed in the reverb condition.

Original	Transcription (reverb)
and these are reflected in your personality.	and these ___ reflect __ your personality.
But we believe a picture in the newspapers will be	But we believe an ad in the newspapers will be
and some small hand weights.	and some small hands ____.

Table 5.51. Examples of constructive errors substituting parts of words or whole words observed in the phone condition.

Original	Transcription (phone)
physics has shown us the loud bang	physics has explained __ the loud bang
when the balloon does burst open, it does so faster than	when the balloon does burst open, it bursts ___ faster than
the significance of their new invention was	the significant of their new invention was

Table 5.52. Examples of constructive errors substituting parts of words or whole words observed in the construction condition.

Original	Transcription (construction)
as our society becomes more and more multicultural,	as our society becomes more and more _____cultural,
wearing fluorescent orange jackets.	wearing fluorescent orange jumpsuits.
the air will leak out quietly and slowly.	the air will let out quietly and slowly.

Table 5.53. Examples of constructive errors substituting parts of words or whole words observed in the s.b.e. condition.

Original	Transcription (s.b.e.)
but don't tell the students you've done this.	but don't tell the students you__ did this.
at the rear of the office block.	at the rear of the office building.
Then there is musical intelligence,	Then there is music intelligence,

Table 5.54. Examples of constructive errors substituting parts of words or whole words observed in the m.b.e. condition.

Original	Transcription (m.b.e.)
We can observe this principle in the real world	We can observe these principles in the real world
By this I mean the national newspapers, in order to	By this I mean the national news_____, in order to
balloons make a loud bang when	balloons make a loud sound when

Table 5.55. Examples of constructive errors substituting parts of words or whole words observed in the s.b.i. condition.

Original	Transcription (s.b.i.)
you should always head towards the main stairs	you should always head towards the main stair
so that everyone is clear on the streams for	so that everyone is clear about the streams for

each First you inflate the balloon and then you put the sticky tape	each First you inflate the balloon and then you inflate the sticky tape

5.6.2 Delivery Errors

Marslen-Wilson defines delivery errors as those target words lacking "articulatory clarity and fluency," and this category includes "slurrings, hesitations, stutterings, and unintelligible responses" (Marslen-Wilson, 1985: 58). There were only five instances of single delivery errors, while the rest were always followed by another delivery error or errors, or by omissions. This type of errors had the lowest number of occurrences, the proportion ranging from only 2.71% of the total number of errors in the clean condition to 3.96% of the total number of errors in the construction condition (Table 5.20). Since they constituted unintelligible attempts at target words, these errors could not be transcribed.

5.6.3 Insertions

As explained in 5.6, the category or insertions, sometimes also referred to as additions, was analyzed separately. The reason for this was that following Marslen-Wilson's convention of error categorization (1985) – in other words, counting all the errors in a stimulus and deducting that number from the total number of words in it – insertions would add to the total number of words in a stimulus and ultimately result in inaccurate error percentages.

5.7 Data Analysis

Since the total number of words differed from one stimulus to another, the counts were first transformed to proportions expressed as per cent, after which the mean, median, and standard deviation were calculated for each condition, as given in Table 5.56.

Table 5.56. Mean, median, and standard deviation in the shadowing task.

condition	mean	median	sd
clean	0.46	0.30	0.55
reverb	1.17	0.83	0.86
phone	1.09	0.67	0.92
construction	1.21	1.13	0.75
single babble energetic	0.90	0.62	0.80
multi babble energetic	1.02	0.67	0.86
single babble informational	1.10	1.26	0.75

Figure 5.11 shows the insertion percentages and standard deviations for each experimental condition.

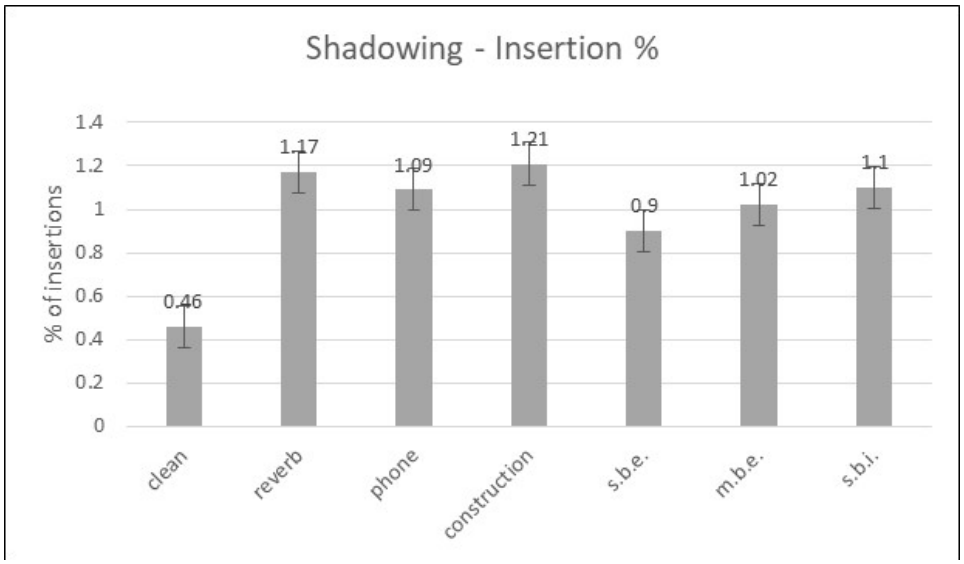

Figure 5.11. Mean percentages and standard deviations in the shadowing task.

5.7.1 Comparing Insertions in Clean vs. Other Conditions in the Speech Shadowing Task

Using the Shapiro-Wilk test, the distribution of insertions in each of the conditions was checked. The results of this test are shown in Table 5.57.

Table 5.57. Results of the Shapiro-Wilk test in the shadowing task.

Condition	p-value
clean	0.0000001675
reverb	0.0003712
phone	0.00001742
construction	0.005615
single babble energetic	0.0001019
multi babble energetic	0.0001472
single babble informational	0.0009749

None of the conditions followed a normal distribution. Signed rank test was used in pairwise comparisons to find out whether there were statistically significant differences in the insertion rates between the control condition (clean) and every other experimental condition. The results of these tests are presented in Table 5.58.

Table 5.58. The results of the Wilcoxon matched-pairs signed rank test comparing the clean (quiet) condition with each of the noise conditions.

condition	p-value	z-value	p-adjusted	p-adj-signif.
clean - reverb	0.00002344	-4.229301	0.00014064	***
clean - phone	0.00001179	-4.381391	0.000070740	***
clean - construction	0.0000007792	-4.940495	0.0000046752	***
clean - single babble energetic	0.0002211	-3.693601	0.0013266	**
clean - multi babble energetic	0.00001427	-4.33964	0.000085620	***
clean - single babble informational	0.0000009149	-4.909117	0.0000054894	***

Statistically significant differences were found between the clean condition and all other experimental conditions.

5.7.2 Comparing Insertion Rates within the Energetic Noise Category Conditions

The pairwise Wilcoxon signed-rank test was used to identify significant differences within the energetic noise category, the results of which are presented in Table 5.59.

Table 5.59. Results of the pairwise Wilcoxon signed-rank test in the shadowing task.

condition 1	condition 2	p-adj	p-adj-signif
construction	multi babble energetic	0.105	ns
construction	single babble energetic	0.023	*
multi babble energetic	single babble energetic	1.000	ns

A statistically significant difference was only found between the single babble energetic and construction conditions, as graphically illustrated in Figure 5.12.

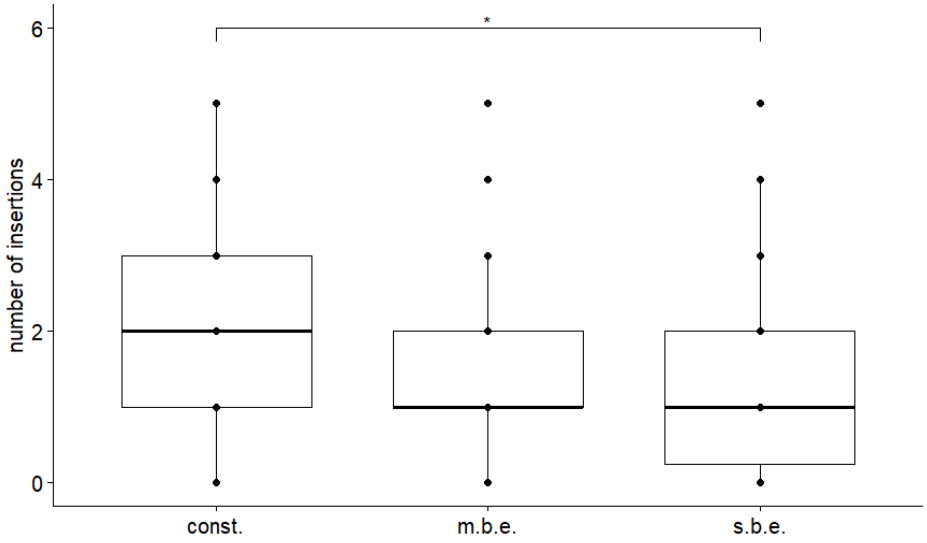

Figure 5.12. Statistically significant differences discovered by the pairwise Wilcoxon signed-rank test.

5.7.3 Comparing Insertion Rates within the Signal Degradation Noise Category

The Wilcoxon matched-pairs signed rank test was used in order to establish the presence of statistically significant differences in the insertion rates between the reverb and phone conditions. No statistically significant difference in the medians was found as can be observed in Table 5.60.

Table 5.60. Results of the pairwise Wilcoxon signed-rank test in the degradation group.

condition 1	condition 2	p-value	z-value
reverb	phone	0.6797	-0.4128414

5.7.4 Comparing Insertion Rates between the Noise Categories

Finally, the Wilcoxon matched-pairs signed rank test was used in order to establish the presence of statistically significant differences in the insertion rates between the noise categories.

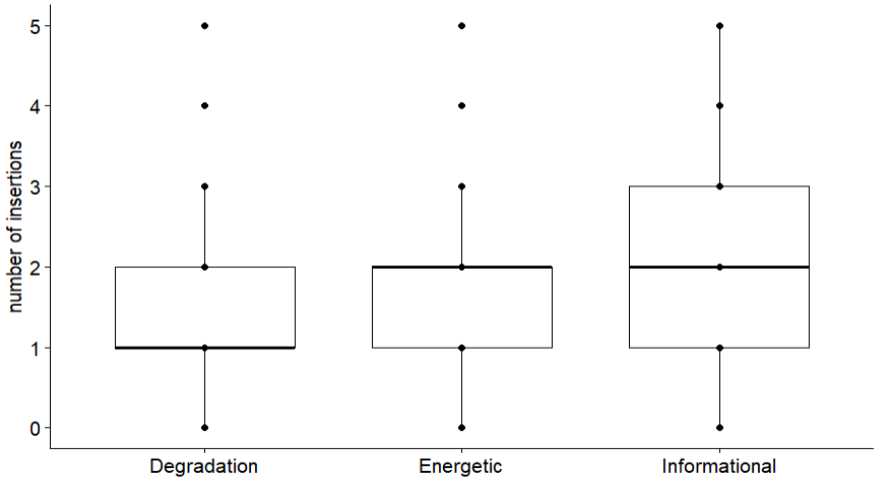

Figure 5.13. The pairwise Wilcoxon signed-rank test found no statistically significant differences between error rates in the three noise categories.

The test found no statistically significant differences between the noise categories (Table 5.61).

Table 5.61. Results of pairwise Wilcoxon signed-rank test for the between-noise category differences.

Condition 1	condition 2	p-adj	p-adj-signif
degradation	energetic	0.876	ns
degradation	informational	0.588	ns
energetic	informational	1.000	ns

5.7.5 Examples of Insertions in Speech Shadowing

The following section (Tables 5.62-69) provides illustrations of insertions found across the conditions. Interestingly, there were only four instances of content word insertions:

Table 5.62. Examples of a content word insertion observed in the clean condition.

Original	Transcription
Then there is musical intelligence, followed by kinesthetic,	Then there is musical intelligence, followed by kinesthetic intelligence, (phone)
the products are now selling well nationally in department stores,	the products are now selling well nationally in social department stores, (construction)
It's vital that you do not spend time collecting	It's somewhat vital that you do not spend time collecting (phone)
by tucking their hands	by tucking their own hands (clean)

All the other insertions throughout the trials were function word insertions. Their illustrations are presented in the following tables:

Table 5.63. Examples of a function word insertion observed in the clean condition.

Original	Transcription (clean condition)
You also need a pin and some sticky tape.	You will also need a pin and some sticky tape.

established patterns in the way we conduct ourselves	established patterns in the way that we conduct ourselves
We're going with this method.	And we're going with this method.

Table 5.64. Examples of a function word insertion observed in the reverb condition.

Original	**Transcription (reverb)**
except if it sounds at 11.00 a.m.	except for if it sounds at 11.00 a.m.
the loud bang occurs	the loud bang that occurs
but don't tell the students you've done this.	but don't tell the students that you've done this.

Table 5.65. Examples of a function word insertion observed in the phone condition.

Original	**Transcription (phone)**
However, psychologists now recognize	However, the psychologists now recognize
But today I'd like to take a look at	So but today I'd like to take a look at
gradually over time, as our society becomes	gradually over the time, as our society becomes

Table 5.66. Examples of a function word insertion observed in the construction condition.

Original	**Transcription (construction)**
and this describes people who are more interested	and this one describes people who are more interested
different types of intelligence and these are reflected	different types of intelligence and that these are reflected
for this one you need a chair that can	for this one you will need a chair that can

Table 5.67. Examples of a function word insertion observed in the s.b.e. condition.

Original	**Transcription (s.b.e.)**
so the loud bang you hear is actually	so the loud bang that you hear is actually
the air will leak out quietly and slowly	the air that will leak out quietly and slowly
in the real world in the sport of ice skating	in the real world and in the sport of ice skating

Table 5.68. Examples of a function word insertion observed in the m.b.e. condition.

Original	**Transcription (m.b.e.)**
Most people assume balloons make	Most people assume that balloons make
if you hear the fire alarm, you should always head towards	if you hear the fire alarm, that you should always head towards
so passengers took trains for part of their journey	so passengers took the trains for part of their journey

Table 5.69. Examples of a function word insertion observed in the s.b.i. condition.

Original	Transcription (s.b.i.)
You also need a pin and some sticky tape.	You would also need a pin and some sticky tape.
new marketing and advertising strategy so that everyone is	new marketing and advertising strategy and so that everyone is
I'd like to take a look at	I'd like to take on a look at

5.7.6 Summary of Error Types in Shadowing

Most of the errors made across the conditions were omissions, which is consistent with previous studies (Braun & Hahn, 2011). 43.88% of the errors across all conditions were content word omissions, while 36.75% of the errors were function word omissions (Table 5.20). 16.07% of the overall number of errors made were constructive errors, while only 3.35% were delivery errors (Table 5.20). Lastly, the category of insertions was almost irrelevant, with only 1% of insertions made across the conditions (Table 5.56.)

CHAPTER 6: General Discussion

6.1 Listening Span Task

In this task, participants were presented with seven five-sentence stimuli (one in the control and six in the experimental noise conditions). For each set of five sentences they were instructed to listen to the sentences and to remember the last word of each sentence, and recall them in the order of presentation – serial recall task (Conway et al., 2005). What was tested was their working memory storage, as well as their ability to process workloads that were fairly demanding due to not only sentence length, but also presence of noise maskers, which have been reported to interfere with working memory performance (Armstrong & Sopory, 1997; Brungart, 2001). This was a single-task design, with no other data collected (Conway et al., 2005). The responses were not timed. The amplitudes of the noise maskers were not varied, therefore, the same signal-to-noise ratio of -5 dB was preserved throughout the trials. The study was not interested in determining a tolerance threshold for each individual masker, but rather comparing performance across different maskers. Overall, the results show that participants found the listening span task quite challenging (Table 5.1); with an average error rate of 71.02% across all conditions noticeably high. This error rate indicates that on average only 1.45 words out of 5 were recalled in their correct serial position. The obtained results correspond with the nature of a serial recall task, known to be quite demanding on short-term memory (Conway et al., 2005). The results revealed no significant differences in error rate across noise conditions with one exception. The construction masker surprisingly yielded fewer errors

than the control quiet condition. Thus, the number of errors made in the construction condition was 17.2% lower compared to the control condition.

Such an outcome is not completely unheard of, since it has been reported that construction workers do not find the noise typically found in their working environment (e.g. constructions sites) to have an effect on their hearing and communications (Yang et al., 2021). This usually results in "the familiarization to the on-site environment," which, as a result, improves one's "ability to capture the target sound" (Yang et al., 2021: 10). However, it is highly unlikely for a group of first- to third-year business students to have had an experience of working at a construction site which would have ultimately made them comfortable being around this type of noise.

In single babble energetic and multi babble energetic conditions participants made 72.8% and 72.0% errors respectively. While such informational maskers are typically thought of as more detrimental to speech perception that their energetic counterparts, in the listening span task, informational masking yielded 6% fewer errors than energetic maskers. Both noise conditions from the degradation noise category – reverb and phone – were found to result in the highest percentage of errors (77.6% each). For the reverb condition, this is consistent with Rogers and colleagues (2006) who found the performance of their participants considerably poorer in the reverb condition. For the phone condition, such a score was anticipated since it has been well established that "the reduced bandwidth of the telephone speech accounts for a significant amount of performance deterioration" (Hu et al., 2013: 189).

Overall, the findings suggest that when engaged in a relatively short memory task, such as the listening span task, with no embedded tasks or simultaneous secondary tasks, short-term memory capacity could be equally unaffected by both energetic and informational maskers, as well as degradation of the speech sound signal. This was clearly not an anticipated outcome, given the reported detrimental effects of noise maskers on short-term memory (Conway et al., 2005). The predicted outcome was not observed in this experiment; likewise no definitive explanation could be provided as to the number of errors made in the control condition, which was similar to the number of errors in the experimental conditions. A possible explanation for the uniformity of results across conditions may be the fact that in adverse listening conditions word processing is enhanced by predictable contextual information (van der Feest et al., 2019). In addition, just about any type of noise between words disturbs the short-term memory storing, which could also account for the relatively low and uniform results (Ljung & Kjellberg, 2009). Previous studies investigating effects of reverberation on speech perception reported that individuals with lower working memory "experienced a more precipitous decline in speech intelligibility as a function of reverberation" (Reinhart & Souza, 2016: 1549). This finding could possibly add yet another explanation as for the uniformity of results across all trials, not just the reverb, for a group of university students should most definitely possess high working memory capacity, which has been proved to provide "a form of cognitive compensation in cases of degraded speech acoustics" (Reinhart & Souza, 2016: 1549). Additional work, including free recall modality and secondary tasks would

most probably help explain such high and uniform error rates across all conditions, and hopefully account for the unexpectedly poor scores in the clean condition.

6.2 Listening Comprehension Task

Participants were presented with seven five-sentence stimuli (one control set and six sets in experimental noise conditions) and asked to answer four multiple-choice questions after each condition. Tested was their ability to process and comprehend auditorily presented sentences in adverse conditions, and understand the main ideas discussed in the recordings. Numerous studies have documented detrimental effects of background noise on listening comprehension (Berne, 2004; Picou et al., 2016; Yang et al., 2017; Rudner et al., 2018). With that in mind, it was predicted that the outcomes of the listening comprehension test would be adversely affected by background noise maskers. Similarly to the previous experiment, this was a single-task design, with no other data collected. Responses were not timed. Amplitudes of the maskers were not varied, and the same signal-to-noise ratio of -5 dB was preserved throughout the trials, because the study was not interested in determining a tolerance threshold for each individual masker, but rather comparing performance across different maskers. Overall, the results show the average of 50.00%, error rate across all conditions (Table 5.7). Statistically significant differences from the control condition were found for all experimental conditions, except for multi babble energetic. The greatest number of errors was measured in the phone condition, which had a

narrowed frequency bandwidth ranging from 350 to 3400 Hz. This result corresponds with findings from studies which established that the quality of speech perception suffers dramatically when high frequencies are removed (Moore & Tan, 2003; Monson et al., 2014). This also implies that relevant linguistic information is contained in the high-frequency band, which is something that should be tested and expanded upon in future research. The two energetic maskers that significantly affected the results were construction noise and single babble, with the error rate of 65.5% and 54.0% respectively. Surprisingly, the single babble informational masker yielded 9.0% fewer errors than the single babble energetic. This is in contrast to what is typically reported in the literature that has compared the two maskers – by and large, it is informational masking that ultimately leads to poorer performance in listening comprehension tasks (Klatte, Bergström & Lachmann, 2013; Prodi & Visentin, 2017), but also speech perception and spoken word recognition in general (Brungart et al., 2001; Wightman et al., 2006; Schneider et al., 2007; Lecumberri et al., 2010; Snyder et al., 2012). Even though statistically significant, the reverb condition proved not to be too detrimental in this particular task, with only 11.5% more errors than the control. Overall, the findings suggest that noise maskers do have negative effects on listening comprehension – the outcome that was expected in this experiment, which which resembled participation in an oral training session or listening to a lecture.

6.3 Speech Shadowing

After completing two to three practice trials (with no noise maskers added), participants were presented with seven five-sentence stimuli (one control and six experimental noise conditions) and instructed to shadow each recording at a pace they felt comfortable with. Tested was their ability to accurately perceive speech sounds and words presented in noise, and to repeat them immediately. Responses were audio recorded. The accuracy of shadowing was measured. The final scores were obtained by subtracting all the missing and mispronounced words from the total number of words for each audio recording. Overall, the scaled results show an average 7.13 error rate per 100 words (Table 5.13).

The analysis found statistically significant differences from the control condition were found for all experimental noise conditions. As expected, the greatest number of errors (10.39%) was found in the single babble informational condition. This result corresponds to previous studies, which found the effects of background speech in a language familiar to the listener to be the most detrimental to speech perception (Brungart et al., 2001; Schneider et al., 2007; Lecumberri et al., 2010; Snyder et al., 2012; Mattys et al., 2012). Furthermore, in line with the results of previous studies, background speech in a language with which the listener is not familiar – single babble energetic masker – was found to affect performance less (5.26%) than its informational counterpart (Brungart et al., 2001; Lecumberri et al., 2010; Snyder et al., 2012; Mattys et al., 2012). The two remaining energetic maskers – construction and multi babble energetic – resulted in 7.85% and 7.47% error rate respectively. Also expected was a high

error rate in the reverb condition, which at 9.61% was only 0.78% less detrimental than the single babble informational masker. This result is also in line with what has previously been reported (Rogers et al., 2006; Snyder et al., 2012). Finally, the phone condition yielded an error rate of 5.68%, resulting in 1.84% more errors compared with the clean condition.

At this point, an important distinction needs to be made between continuous shadowing and simple repeating. When engaged in shadowing, the subject follows a continual flow of discourse, while in simple repetition, the subject is expected to repeat the heard utterance during a pause in the model's speech. In addition, shadowing is an online process requiring subjects to focus on the model's speech they are listening to and vocalize it, having almost no time left to think about meaning (Hamada, 2016). Also, when shadowing, subjects are primarily focused on the phonological properties of the model's speech. Simple repetition, on the other hand, is an off-line process which does allow for silent pauses during which participants have time to engage in cognitive activities, accessing the meanings in particular (Kadota, 2007). Therefore, while repeating, subjects have to temporarily store a complete chunk of input information, whereas during shadowing, there is neither time nor the requirement to do so (Hamada, 2016). Instead of focusing solely on the phonology of input speech, subjects engaged in repetition have sufficient time to analyze higher-level grammatical features while getting ready to start repeating the model's speech. They may perform a number of cognitive tasks, such as phonological, syntactic and semantic processing. The two tasks also have different goal – shadowing in the context of

language learning is expected to improve active listening and prosody; with repeating, on the other hand, the focus is on extra time given to the learner to analyze the input and carefully construct the output (Sumarish, 2017).

The obtained results correspond with previous findings linking the effects of background noise to speech perception (Brungart et al., 2001; Schneider et al., 2007; Lecumberri et al., 2010; Snyder et al., 2012; Mattys et al., 2012). The aim of this experiment was to examine the effects of background noise on an on-line listening and speaking process by measuring the error rate and comparing it to the error rate obtained in the same task with no background noise. The results showed that noise maskers did significantly impair performance.

Finally, the analysis found that participants made overall only 1% of insertions across the conditions, ranging from 0.46% in the clean condition to 1.17% in the reverb condition (Table 5.56). Except for the clean condition, the distribution of insertions across the experimental condition was very uniform, with significant differences between all the experimental conditions when compared with the control (Table 5.58).

While the first part of the thesis investigated the effects of different noise maskers on human participants, the second part will examine how the same noise maskers affect speech-to-text technology, which converts spoken input into text output. Designed to enable fast and accurate speech transcription in different settings, ranging from academic to clinical and industry, this technology has been found sensitive to background noise (Fontan et al., 2022).

CHAPTER 7: Automatic Speech Recognition

The second part of this thesis tested the same types of maskers with two speech-to-text processing models, Ava and Otter, to find out whether such technology would be more resilient to adverse noise conditions.

7.1 Background

Automatic speech recognition, or speech-to-text technology, was originally developed as an aid for the deaf and hard of hearing. Ever since the inception of the speech recognition systems, scholars have been addressing their various aspects – most notably accuracy and filtering of background noise – to improve their performance. Gradually, these systems have found their way into not only medical, banking, and military applications, but also to classrooms (Reddy, 1990).

In its most elementary form, speech recognition is "a conversion from an acoustic waveform to a written equivalent of the message information" (Mary, 2018: 20). What is missing from this deceptively simple definition is the fact that this task if often quite challenging because of background noise as well as other external distractions (O'Shaughnessy, 2008). The ultimate goal of speech recognition is to process the speech signal and match it with "a sequence of stored patterns that have previously been learned" (Roe & Wilpon, 1993: 55). During the conversion, the speech signal is being transformed into a sequence of symbols that correspond to speech phonemes and syllables, which, then, gets transformed into a text (Mary, 2018). Therefore, two distinct stages of the

process can be observed – speech signal-to-symbol (phonetic or syllabic) transformation, and symbol-to-text conversion (Mary, 2018). Of equal importance is the pre-processing stage, during which background noise is filtered out, and the overall signal-to-noise ratio improved, making the speech signal "more acceptable for feature extraction analysis" (Ibrahim et al., 2017: 186).

The beginnings of automatic speech recognition technology can be traced back to the early 1950's when Bell Labs introduced a program called Audrey, capable of understanding and transcribing simple numbers (Furui, 2004). As computing technologies advanced in the 1960s and 1970s, the first practical recognizers were developed, which utilized filter banks together with the process of dynamic programming (O'Shaughnessy, 2008). These were capable of performing simple recognition tasks of isolated words (with pauses between each word). This was an era when research on automatic speech recognition gained its momentum, with numerous labs in Japan, England, the Soviet Union and the U.S. diligently working on hardware for spoken sound recognition, capable of supporting four vowels and nine consonants (Sarma & Sarma, 2014). Considering the limitations of the computer technology of the time, this hardware was truly impressive! This novel technology was now able to take small fragments of the speech signal and successfully determine phonemes, or the smallest units of speech that differentiate meanings, then feeding them into another program responsible for word guessing, based on a probability function. This function would search endless transcriptions, looking for the perfect match for the

target word. Computing multiple sets of specific templates, target units were now being compared with testing units that would eventually select the closest match for the corresponding input (O'Shaughnessy, 2008). The subsequent milestone came in the 1970s, with Hidden Markov Models being successfully integrated into speech-to-text technology (Ferguson, 1980). One of the reasons why these models quickly gained popularity was that they were quite simple and that their computations were feasible, so that their parameters could be automatically estimated from a large quantity of data (Huang & Deng, 2009). With the introduction of Hidden Markov Models, statistical models replaced templates, also referred to as probability density functions or PDF's, so that "instead of seeking the template closest to a test frame, test data are evaluated against sets of PDFs, selecting the PDF with the highest probability" (O'Shaughnessy, 2008: 2967). This approach, thus, signified a major shift from "simple pattern recognition methods, based on templates and a spectral distance measure, to a statistical method for speech processing," resulting in much higher accuracy (Rabiner, n.d.). The main advantage of Hidden Markov Models over the previous systems was their ability to "retain more statistical information about the speech patterns than templates" (Roe & Wilpon, 1993: 58). The next big development in automatic speech recognition came in the late 1980s and early 1990s when neural networks started being incorporated into this technology, inducing significant improvements. Featuring a multi-layer architecture capable of performing massive amounts of computation, these brain-inspired networks were characterized by "their unprecedented energy-efficiency and rapid information

processing" (Wu et al., 2020: 1). The neural networks approach was essentially an attempt to replicate the human brain architecture, employing a multilayered network of artificial nodes. The advancements in theory and application of statistical modeling, such as neural networks, gave rise to artificial intelligence-based solutions, so far the most successful model in automatic speech recognition (Jiang et al., 2019). However, speech recognition systems are still far from perfect, being especially susceptible to inaccuracies due to individual accents they are not trained for and background noise.

One of the main issues with automatic speech recognition models since their inception has been the issue of accuracy. For instance, while one model may be calibrated or trained to work with a particular speaker in a quiet environment, when tested with a different speaker, or a different microphone, or in the presence of background noise, reduced or lower accuracy is usually obtained (Kathania et al., 2020). This is referred to as the *mismatch problem* (Xu et al., 2018). The mismatch problem of lower accuracy is typically the result of these systems being affected by speech variance, or the variability in the speech signal, which can be the result of different speaking rates, different acoustic conditions, and even different contexts (Roe & Wilpon, 1993). While accentedness and speaker variance are beyond the scope of this research project, testing speech-to-text software in adverse conditions offers a valuable perspective into how accurate present-day technology is when exposed to a signal combined with background noise.

The present study aimed to analyze the accuracy of two varieties of speech-to-text software – Otter and Ava – which are commonly used in both academic and professional settings. In particular, the extent to which different types of background noise mixed with the input speech signal affects the word error rates (WER) of these two applications was investigated. In order to control for a number of variables typically reported to cause inaccurate renditions of the input speech signal – such as speech rate, accent, the quality of microphone used, the distance between the speaker and microphone – professional recordings were used (Xu et al., 2018, Kathania et al., 2020). Another reason for using professional recordings is disfluencies common in spontaneous speech (Deng et al., 2020). Even though both Ava and Otter applications are programmed to disregard vocal disfluencies such as *uh*, *um* and *ah*, it could not be anticipated how such fillers would be treated by either application when mixed with background noise. In both cases, free versions of the programs were used, since these are readily available to most customers. The study also expands on previous research that has investigated the accuracy of the two applications in a university classroom environment (Weigel, 2021). The results of this analysis are expected not only to be of benefit to those writing speech-to-text algorithms, but also educators and interpreters relying on such software in their work.

7.2 Method

7.2.1 Materials and Procedures

Nine speech recordings were used as stimuli – seven previously used in the speech shadowing task with human participants, and two additional ones. Each of them was tested on the speech-to-text applications in seven different modes – clean, with no background noise added, and in noisy conditions with the following types of background noise added to the speech signal: *construction* noise, *single babble energetic*, *single babble informational*, *multi babble energetic*, *phone*, and *reverb*. Noise maskers added to the clips were taken from the BBC Sound Effects Library. The maskers were added at the signal-to-noise ratio of -5 dB, based on the results of previous studies with human participants, which found that this particular ratio kept the average intelligibility at between 45% and 65% (Smiljanić & Bradlow, 2009).

Following the experiments with human participants, the energetic maskers used were *construction* noise (drills and jackhammers), *single babble* (one voice) and *multi babble* (multiple voices) maskers. The single babble masker resembled a real language and was found in the BBC Sound Effects Library. The multi babble masker was created using an online babble noise generator.[10] The settings for the multi babble noise were as follows: sub-bass -32, low bass -31, bass -32, high bass -26, low mids -20, mids -19, high mids -22, low treble -21, treble -21, high treble -25 (all in dBFS)**.** Since Otter announced in the fall of 2021 that its captions would be available in more than thirty languages, and the list expanding

[10] https://mynoise.net/NoiseMachines/babbleNoiseGenerator.php

to other languages in the future, had a real language been used as a masker, it would have actually acted as an informational masker. This was also empirically confirmed prior to the experiment, where recordings of Mandarin and German speech were played into the software, which rendered the output captions in English. The informational masker was a news broadcast in English. The clips that featured degraded sound were made by narrowing down the speech frequency bandwidth to 350-3400 Hz range for the *phone* condition, and creating a reverb effect for the *reverb* condition. The reverberation time was 1s (pre-delay 47 ms) – values typically found in classrooms with unfavorable acoustics and in larger conference rooms (Labia et al., 2020).

The mixing and effects were done using Pro Tools 2020.5 software. The audio files were sampled at 44.1 kHz with 16 bits per sample, and normalized for loudness by matching their root-mean-square power, so that no difference in intensity between the stimuli would affect the outcome of the experiments. The presentation of speech material was also counterbalanced between conditions by randomizing the order in which the recordings were played across the trials. Importantly, the recordings were not fed to the programs through external microphones, but directly thorough the researcher's computer, therefore, no outside (and therefore, uncontrolled) noise could mix with the speech signal. Following the methodology employed by Weigel and in order to demonstrate the usefulness of the programs out of the box neither application was trained on a particular speaker (Weigel, 2021). For the same reason, the order in which the

stimuli were played was randomized to avoid the software learning effect. Furthermore, prior to each trial, the applications were closed and restarted.

The obtained transcripts were analyzed, and word error rates calculated. Speech rate for each excerpt was calculated using an online speech rate calculator[11]. The speech recordings contained between 149 and 177 words (mean = 161.77, SD = 10.89), and their speech rate (here calculated as the number of words per minute) was between slow and moderate (mean = 160, SD =14.94). Each recording contained a 5-second silence gap in the beginning and end. The 5-second gaps in the beginning and end of the recordings were not taken into the speech rate calculation. No recording contained longer pauses, typically found in spontaneous speech. All the clips were edits culled from an exercises audio CD that accompanies *Cambridge Vocabulary for IELTS Advanced Band 6.5+*.

7.3 Results

Following the methodology established by Errattahi et al. (2018) and repeated by Weigel (2021), word error rate (WER) scores were calculated based on three types of errors found in the transcripts – substitutions (marked in red in the examples below), omissions (marked with red lines in the examples below) and additions (marked in blue in the examples below). Also following the previous two studies, the present analysis did not take punctuation errors into account. At a glance, numerous punctuation errors were produced by both programs even in the

[11] The speech rate calculator can be found here: http://www.speech-topics-help.net/speed-of-speech.html

clean condition, and the number increased with noise maskers added. Detailed analysis of such errors was beyond the scope of this project.

7.3.1 Transcription Errors by Otter

For each observed condition, the mean, median, and standard deviation were calculated.

Table 7.1. Mean and median percentages of word errors, and standard deviations produced by the Otter software for each condition.

Otter Data condition	**mean**	**median**	**sd**
clean	2.59	2.20	1.47
reverb	3.74	2.70	2.74
phone	13.09	12.70	4.85
construction	33.49	35.50	5.49
single babble energetic	26.20	22.30	10.19
multi babble energetic	7.34	6.90	2.71
single babble informational	35.41	36.40	8.57

Figure 7.1 shows Otter performance expressed as the overall accuracy percentages.

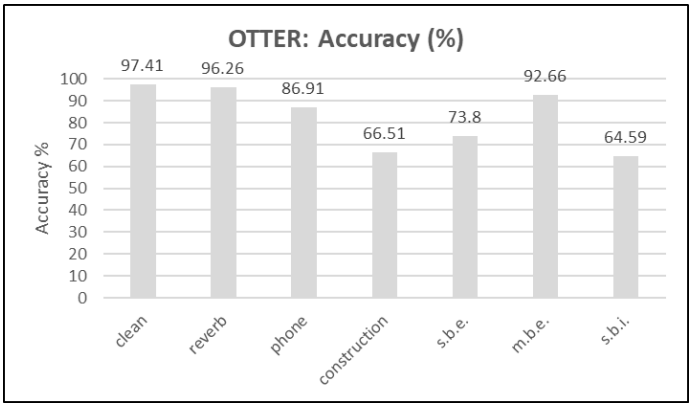

Figure 7.1. Overall word accuracy in Otter transcripts across the noise conditions.

7.3.1.1 Discussion

As expected, the transcript of the clean condition remained closest to the original, with the overall accuracy rate of 97.41%. The errors found in this condition can be labeled relatively insignificant, including an occasional misspelling or a missed word, not influencing the understanding of the overall message to a significant degree. Table 7.2 shows examples of errors obtained in the clean condition when using Otter.

Table 7.2. Examples of errors observed in the clean condition, the accuracy of which was 97.41%.

Otter: Clean	
Original	**Transcription (with errors marked in red)**
in a small airplane for 33 hours from New York to Paris.	in a small airplane for 33 ____ from New York to Paris.
Thanks to the weights, the student sitting in the chair will be	Thanks to the weights, the students sitting in the chair will be
So it can't be the air escaping that causes the noise	So it can't be the air escaping but causes the noise
the exposure to these products. And despite the fact that our	the exposure to these products___despite the fact that our
to remain with the company for the foreseeable future	to remain with the company for the facility future.

The reverb condition fared similarly with the overall accuracy of 96.26%. Again, an occasional misspelling or a missed word was observed in the analysis, mostly having no significant impact on understanding.

Table 7.3. Examples of errors observed in the reverb condition, the accuracy of which was 96.26%.

Otter: Reverb	
Original	**Transcription (with errors marked in red)**
but their planes were noisy, cold and uncomfortable.	their planes were normal, cold and uncomfortable.
Most people assume balloons make a loud bang when the air is released through the hole.	Most people assume balloons, make a loud bang when the air is released. Through the hull.

it does so faster than the speed of sound, so the loud bang you hear is actually a sonic boom. In the real world, this principle is used to test	it does so faster than the speed of sound. So the loud bang you hear, is actually a sonic boom.__ The real world, this principle is used to test
I think it's time for us to divide up parts of	____ think it's time for us to divide up parts of
After that there is visual intelligence, which describes people	_____ that there is visual intelligence which describes people

The next best performance on the accuracy scale was in the multi babble energetic condition at 92.66%. In addition to omissions and misspellings, in this condition for the first time substitutions could be observed. Even though phonetically similar to original words, these substitutions bore no semantic connection to the rest of the message, and would likely have caused comprehension problems.

Table 7.4. Examples of errors observed in the m.b.e. condition, the accuracy of which was 92.66%.

Otter: Multi Babble Energetic	
Original	**Transcription (with errors marked in red)**
the first passenger planes did little to change that.	the first passenger planes delivered to change, that.
transatlantic flight captured America's imagination.	transatlantic light to cast in America's imagination.
Thanks to the weights, the student sitting in the chair will be able	Thanks to the waste, the students sitting in the chair will be able
Instead, physics has shown us the loud bang occurs because the hole expands rapidly	instead physics has shown us the large bangs and purse, because the hole expands rapidly,
most of the advertising was done through leaflets posted through	most of the advertising was done through leafless posted through

Of the two degraded signals, the phone condition presented a more significant challenge to the software, resulting in omission of significant parts of sentences and reducing the accuracy rate to 88.91%.

Table 7.5. Examples of errors observed in the phone condition, the accuracy of which was 88.91%

Otter: Phone **Original**	**Transcription (with errors marked in red)**
Twenty years later, by 1923 the first passenger planes did little to change that. The first of these were provided by some of the airmail services flying mail around the country.	20 years later, by 1923 _____ _____ Did little to change. That _____ _____ male services flying mail around the country.
They couldn't fly over mountains, so passengers took trains for part of their journey.	They couldn't fly over mountains so passing took trains but part of their journey.
One of the main factors in ensuring a harmonious society is that there are clear established patterns in the way we conduct ourselves.	One of the main factors in in sharing_____ a harmonious society _____ _____ in the way we conduct ourselves
Those who fail to observe these norms are inevitably excluded from that group.	Those who fail to observe these norms are inevitably _____from that group.
I want to run through the fire evacuation procedure now that we're in a new building. First of all, can I just remind you	I want to run through the fire of action _____ in a new building _____ of all. Can. I just remind you

Even larger chunks were missing from the transcripts obtained in the single babble energetic condition. Coupled with a greater number of insertions and the overall accuracy rate of 73.80%, these transcripts would be relatively useless to anyone who would have to rely on them for comprehending the message (the blue color is used for the insertions).

Table 7.6. Examples of errors observed in the s.b.e. condition, the accuracy of which was 73.80%.

Otter: Single Babble Energetic **Original**	**Transcription (with errors marked in red)**
there are clear established patterns in the way we conduct ourselves. And we expect people to behave according to our accepted standards of behavior. There are those who observe these social mores religiously, and these people are often labeled *conservative*. It's actually through such people that our heritage is preserved, but then, gradually over time,	they're _____ established _____ in the way we conduct ourselves. _____ _____ _____ _____ _____ _____ preserved.___ Then gradually over time,
Thanks to the weights, the student sitting in the chair will be able to control their own speed. If they hold the weights out, they will slow down and if they hold them close to their body,	Thanks to the way ____ we're ___ sitting in the chair will be able to control their _____ _____ way to slow down and _____ close to their body,
Most people assume balloons make a loud bang when the air is released through the hole. However, if you pierce the balloon through the	most people assume balloons maker love back __ when the error is released with the whole. However, if you pass the blame through the

sticky tape, instead of bursting it, the air will leak out quietly and slowly.	stickers. Instead of busting it. The I will live, how about why _____ slowly.
You might think that the list of negative factors would include discrimination, but it doesn't because discrimination comes under the larger category of fear. Now, what you should also notice is that the external factors are not labeled in this way. It's much more difficult to know how to measure the effects of external factors and whether they actually are external or not. The influence of family relationships, climate, beliefs and values,	_____ The list of negative, _____ discrimination, but it doesn't because discrimination comes under the larger category of fear. Now, _____ _____ is that the external factors are not legal _____ difficult to know how to measure the effects of external factors, ____ whether they actually are external _____ ____ influence of family relationships _____ beliefs and values
Which leads me to the next point for future development – that of increasing our workforce. It's become clear that all our departments are understaffed, so we'll be taking on more employees over the next year.	Which leads me to the next point for future development _____ of increasing our workforce _____ and all other places,_____ understaffed, so we'll be taking on new employees over the next year.

Of similarly poor quality were the transcripts from the construction condition, where the software was just skipping fairly large chunks of sentences. This condition proved to be one of the most difficult ones for the software, which scored at only 66.51% accuracy.

Table 7.7. Examples of errors observed in the construction condition, the accuracy of which was 66.51%.

Otter: Construction	
Original	**Transcription (with errors marked in red)**
the significance of their new invention was of course not yet apparent. Twenty years later, by 1923 the first passenger planes did little to change that. The first of these were provided by some of the airmail services flying mail around the country.	the significance of their new invention was _____ a 20 years later by 1,923. The first passenger planes _____ to change that _____ _____ male services. Flying mail around the country.
And we expect people to behave according to our accepted standards of behavior. There are those who observe these social mores religiously, and these people are often labeled conservative. It's actually through such people that our heritage is preserved, but then, gradually over time, as our society becomes	And we expect people ___ behave according to our _____ standards of behavior _____ religiously, and these people are often labeled conservative _____ such a through such people.____ But our heritage _____ gradually over time, as our society becomes
The next experiment is called the arm engine and for this one you need a chair that can swivel or rotate and some small hand weights. This is a great experiment for demonstrating an	_____ The arm engine. ___ For this one you need a chair that can swivel or rotate, and some small hand weights _____ demonstrates

important principle of energy and momentum. However, if you pierce the balloon through the sticky tape, instead of bursting it, the air will leak out quietly and slowly. So it can't be the air escaping that causes the noise. Instead, physics has shown us the loud bang occurs because the hole expands rapidly forming a catastrophic crack. You can also tell your students, when the balloon does burst open, it does so faster than the speed of sound, this was measured with an IQ test. However psychologists now recognize that there are many different types of intelligence and these are reflected in your personality.	_____ the energy and momentum. However,_____ _____ instead of bursting ___ _____ quietly slowly, so it can't be the air escaping that causes the noise._____ _____ to show us the lamp bag _____ forming a cat ___ you can all _____ the the balloon does burst open. it does so faster than the speed of sound. this was measured with the ___Q test. _____ _____ Recognize that there are many different types of _____ _____ reflecting ___ your personality.

At 64.59% accuracy, a similarly low score was obtained in the single babble informational condition. In addition to significant number of omissions and misspelled words, this condition also saw the greatest number of insertions – an outcome that could be anticipated given the nature of the masker.

Table 7.8. Examples of errors observed in the s.b.i. condition, the accuracy of which was 64.59%.

Otter: Single Babble Informational **Original**	**Transcription (with errors marked in red)**
and these people are often labeled conservative. It's actually through such people that our heritage is preserved, but then, gradually over time, as our society becomes more and more multicultural, there is a blending of these customs and we gradually come to redefine the norm. If we enter a new group, we notice the unwritten rules and social norms of that group. Those who fail to observe these norms are inevitably excluded from that group. Of course, there will always be This is a great experiment for demonstrating an important principle of energy and momentum. Ask one of your students to sit on the chair holding the weights in their hands. Then get another student to spin the chair as fast as they can. Thanks to the weights, the student sitting in the chair will be able to control their own speed. If they hold the weights out, they will slow down and if they hold them close to their body, making themselves narrower, they will	And these people are often labeled__affected a week it's actually _____our heritage is preserved Mexico the, but then gradually over time as our society becomes more and more multicultural, where is the_____customs or and we naturally come to, you know, just a little bit_____ we enter a new category when we notice the American writers ___ and social norms of that group then takes the, those who fail to observe these norms are inevitably _____exploded at 16, of course, there will always be the latest is a great experiment for demonstrating the _____ principle of your energy and momentum early tomorrow. Last ____ one of your students to sit on the chat and holding away.____ It is extend____ get another student but just a reason the chat ___as fast as they can over the gold thanks to the ___ way to the student sitting in the chair will be able to control their own speed going to be conducive, they hold the way vacation,_____ slow down

accelerate the speed of their rotation. We can observe this principle

but don't tell the students you've done this. Now ask the students what makes a balloon burst. Most people assume balloons make a loud bang when the air is released through the hole. However, if you pierce the balloon through the sticky tape, instead of bursting it, the air will leak out quietly and slowly. So it can't be the air escaping that causes the noise. Instead, physics has shown us the loud bang occurs because the hole expands rapidly forming a catastrophic crack.

And despite the fact that our competitors advertise baby clothes on TV, we won't be using this method, as our statistics show that it's just not cost-effective. People don't pay much attention to TV ads for baby clothes. But we believe a picture in the newspapers will be much more attractive to potential customers. We're going with this method.

and create a new life are considerably more complex. Let’s start with an overview of the issues as shown on this diagram. You might think that the list of negative factors would include discrimination, but it doesn’t because discrimination comes under the larger category of fear. Now, what you should also notice is that the external factors are not labeled in this way.

and if they hold them _____and then They will accelerate the speed ____rotation real land for, we can observe this principle

But don't tell the students _____ that way through now ask the students what makes a balloon ____ and and as it does the most people assume balloons make a loud bang, when the air is released._____ Once it reemerge however, the goal. Here's _____ the balloon through the sticky Taiko instead of busting it or is the I ___ will leak out quietly and slowly for a reason and copy___ the air escaping ___ causes the noise, and then as it near instead water physics has shown us the loudest bang occurs. Now,____ the whole expands rapidly and forming a catastrophic crime.

And then, despite the fact that our competitors advertise baby clothes on TV. We will be using this method._____
_____
_____
_____ a picture in the newspapers will be much more attractive to potential customers ____ going with this. Early

and create a new life. _____The goal of _____ let's start with an overview of the issues as shown on this diagram which is really really you might think that the list of negative factors____ like discrimination to a category but it doesn't because discrimination. _____ _____ Larger category of fear a little bit so it's now a category what we ____ should also _____ the external factors are not layered in this one takes this hard ____

7.3.2 Transcription Errors by Ava

For each observed condition, the mean, median, and standard deviation were calculated.

Table 7.9. Mean and median percentages of word errors, and standard deviations produced by the Ava software for each condition.

Ava Data condition	mean	median	sd
clean	2.87	2.90	0.69
reverb	4.90	4.00	2.87
phone	3.17	3.00	1.24

construction	30.88	29.30	7.31
single babble energetic	16.44	16.30	4.56
multi babble energetic	9.08	6.80	4.37
single babble informational	22.18	19.70	5.86

Figure 7.2 shows Ava performance expressed as the overall accuracy percentages.

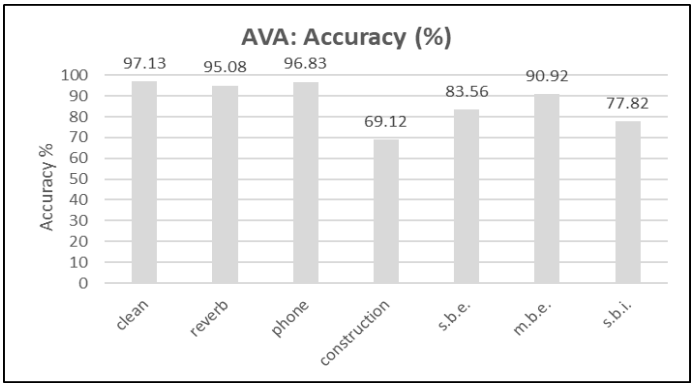

Figure 7.2. Overall word accuracy in Ava transcripts across the noise conditions.

7.3.2.1 Discussion

As with the Otter software, the transcripts obtained in the clean condition remained closest to the original, with the overall accuracy rate of 97.13%. Likewise, the errors were mostly insignificant, such as misspelled words and occasionally missed words, which did not influence the understanding of the overall message.

Table 7.10. Examples of errors observed in the Ava transcript obtained in the clean condition, the accuracy of which was 97.13%.

Ava: Clean **Original**	**Transcription (with errors marked in red)**
flew history's first airplane in North Carolina in 1903	flew his. _____ airplane in North Carolina in 1903,
but their planes were noisy, cold and uncomfortable.	but their planes would noisy cold. Cold and uncomfortable.
Then get another student to spin the chair as fast as they can.	Then get another student to spin the chair as fast as. As they can
do not spend time collecting your bags or	do not spend time collecting your bags or

personal belongings because this wastes valuable evacuation time.	personal. _____ Because this wastes valuable evacuation time.
is used to describe people whose strengths are	is used to describe people who strengths are

Also scoring very high in accuracy was the phone condition, with the overall accuracy of 96.83%. Again, an occasional misspelling or a missed word was observed in the analysis, having no impact on understanding.

Table 7.11. Examples of errors observed in the phone condition, the accuracy of which was 96.83%

Ava: Phone	
Original	**Transcription (with errors marked in red)**
but their planes were noisy, cold and uncomfortable.	but their planes would noisy, cold and uncomfortable.
there is a blending of these customs and we gradually come to redefine the norm.	there is a blending of these customs. And we _____ come to redefine the norm.
If they hold the weights out, they will slow down and	if they hold the way. Out, they will slow down and
The next experiment is called the arm engine and for this one	The next experiment is called the arm.____ Ian and for this one
the final two intelligences are interpersonal and intrapersonal.	the final two intelligences, our interpersonal and intrapersonal

The third-ranked performance was in the reverb condition at 95.08% accuracy. Just like the previous two, this condition contained misspellings and omissions, however, and occasional insertion could be observed as well.

Table 7.12. Examples of errors observed in the reverb condition, the accuracy of which was 95.08%.

Ava: Reverb	
Original	**Transcription (with errors marked in red)**
take a look at how it all began. When	take a look at how it all._____Again, when
for demonstrating an important principle of energy and momentum. Ask one of your	for demonstrating an important principle of energy. _____ Ask one of your
Now ask the students what makes a balloon burst.	Now. I asked the students, what makes a balloon burst.
you should always head towards the main stairs	you should always head._____Main stairs. In

in order to leave the building. Now, last year most of the advertising was done through leaflets posted through people's letterboxes across the city.	order to leave the building, Now, last year, most of the _____ _____ Posted through people's letter boxes across the city.

Multi babble energetic noise was responsible for reducing the accuracy of the next condition down to 90.92%. This condition saw even more insertions, and the resulting transcripts were sometimes impossible to understand. Interestingly, in this condition, one of the transcripts had four asterisks (****) instead of the original word *crack*. This was the only occurrence of asterisks in the entire study, and further research is recommended in order to investigate whether asterisks are something that the software would use instead of a word it had mistaken for a swear word or there was another reason for that.

Table 7.13. Examples of errors observed in the m.b.e. condition, the accuracy of which was 90.92%.

Ava: Multi Babble Energetic **Original**	**Transcription (with errors marked in red)**
And we expect people to behave according to our accepted standards of behavior.	and we expect people to behave. _____ ____ Accept__ sweets standards of behavior.
If we enter a new group, we notice the unwritten rules and social norms of that group.	If we enter a new group, We noticed the unwritten rules and social norms of that growth.
Thanks to the weights, the student sitting in the chair	Thanks to the waves, the students sitting in the chair
Instead, physics has shown us the loud bang occurs because the hole expands rapidly forming a catastrophic crack.	instead physics has shown us the loud, bangs, a person with was the ____ Expands rapidly forming a catastrophic _____****.
do not spend time collecting your bags or personal belongings because this wastes valuable evacuation time. When you have left the building, please look for the fire marshals, who will be	do not spend time collecting your bags or personal belongings._____ _____ _____ Please look for the Fire. Marshals, who will be
And despite the fact that our competitors advertise baby clothes on TV, we won't be using	__ Despite the fact that our competitors at the tide David kids _____on TV, we won't be using
Let's start with an overview of the issues as shown on this diagram.	Let's start with an overview of the issues as ____ And it's diagram,

The single babble energetic masker lowered the accuracy rate of Ava to 83.56%. In addition to misspellings and omissions, it also raised the number of insertions in the transcript.

Table 7.14. Examples of errors observed in the s.b.e. condition, the accuracy of which was 83.56%.

Ava: Single Babble Energetic	
Original	**Transcription (with errors marked in red)**
did little to change that. The first of these were provided by some of the	did little to change that Louisa were the first of these _____ provided by some of the
so passengers took trains for part of their journey. Another problem was that these planes couldn't	So passengers took trains,___most of the part of their Journey. Why should we do another problem? Was that these planes couldn't
where the skaters manage to spin incredibly fast by tucking their hands in close to their body.	Where the skaters managed to spin in with relatively fast by tucking their hands in close to their body.
physics has shown us the loud bang occurs because the hole expands rapidly forming a catastrophic crack. You can also tell your students,	physics has shown us the law. Lads, ___ bang occurs because the whole expands rapidly forming a catastrophic crash, don't you? Don't you don't. You can also tell your students
it's the park at the rear of the office block. Your department has a fire safety officer,	it's the park at the rear of the apis block. You want your department has a fire safety officer
and these are reflected in your personality. The multiple intelligence theory first came to light in	and these are reflected in your personality emotion. Would also the multiple intelligence theory first came to light in
people whose strengths are in subjects such as maths and science. Then there is musical intelligence,	people who strengths are in subjects such as math __ science. Don't you? Don't you don't then? There is musical intelligence

The remaining two conditions, single babble informational and construction, fared the poorest on the accuracy scale. The former achieved the accuracy score of 77.85% with numerous insertions which made the resulting transcript incomprehensible.

Table 7.15. Examples of errors observed in the s.b.i. condition, the accuracy of which was 77.82%.

Ava: Single Babble Informational	
Original	**Transcription (with errors marked in red)**
as our society becomes more and more multicultural, there is a blending of these customs and we gradually come to redefine the norm. If we enter a new	as our society becomes more and more Multicultural, there is _____ application denied,___these customs and we gradually come to redefine, ____ you know, if we enter a new
Most people assume balloons make a loud bang when the air is released through the hole. However, if you pierce the balloon through the sticky tape, instead of	Most people assume balloons make a loud bang. When the air is released through the hole, once it reemerged. However, _____the go past the balloon through the sticky tape instead of
Then get another student to spin the chair as fast as they can. Thanks to the weights, the student sitting in the chair will be able to	Then get another student. _____In the chat as wants to stake a the the gold thanks to the weights in the students sitting in the chair, will be able to
different types of intelligence and these are reflected in your personality. The multiple intelligence theory first came to light in	different types of intelligence.__ These are reflected in your personality. Again, overnight the multiple intelligence theory first came to light in
being able to adapt to the new culture and create a new life are considerably more complex. Let's start with an overview of the issues as shown on this diagram. You might think that the list of	being able to adapt to the new culture and create a new life.__ Probably more complex that the Gulf of Mexico,___ start with an overview of the issues. As shown on this diagram is really, really. You might think, but the list of

The construction masker proved to be the most challenging one for Ava, bringing its accuracy down to 69.12%. These transcripts were characterized mostly by significant chunks of text missing.

Table 7.16. Examples of errors observed in the construction condition, the accuracy of which was 69.12%.

Ava: Construction	
Original	**Transcription (with errors marked in red)**
The US Post Office Department added a few seats for extra revenue, but their planes were noisy, cold and uncomfortable.	The US Post Office department _____ _____ _____and uncomfortable.
as our society becomes more and more multicultural, there is a blending of these customs and we gradually come to redefine the norm.	as ___ a society, becomes more and more Multicultural _____ costumes and we gradually come to redefine the norm.
This is a great experiment for demonstrating an important principle of energy and momentum.	This is a great experiment. _____ _____

Ask one of your students to sit on the chair holding the weights in their hands. Then get another student to spin the chair as fast as they can. Thanks to the weights, the student sitting in the chair will be able to control their own speed. If they hold the weights out, they will slow down and if they hold them dose to their body, making themselves narrower, they will accelerate the speed of their rotation. We can observe this principle in the real world in the sport of ice skating, where the skaters manage to

_____
_____
_____
_____
_____
_____
_____
_____
_____
_____Where the skaters managed to

First you inflate the balloon and then you put the sticky tape on it, but don't tell the students you've done this. Now ask the students what makes a balloon burst. Most people assume balloons make a loud bang when the air is released through the hole. However, if you pierce the balloon through the sticky tape, instead of bursting it, the air will leak out quietly and slowly. So it can't be the air escaping chat causes the noise. Instead, physics has shown us the loud bang occurs because the hole expands rapidly forming a catastrophic crack. You can also

first, you inflate the balloon and then you put sticky._____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____ You can also

Let’s start with an overview of the issues as shown on this diagram. You might think that the list of negative factors would include discrimination, but it doesn’t because discrimination comes under the larger category of fear.

Let's start with an overview of the issues. As should,_____ you might think that the Vista _____
_____ discrimination, but it doesn't because discrimination _____ under the large_____sphere.

Therefore, over the next five years, I aim to set up two small subsidiary companies in order to focus on international expansion in Europe and Asia. There are many organizations in emerging markets which could benefit from our experience and skills.

are for over the next five years. I aim to Apple trees smaller subsidiary._____
Hold. International expansion. _____There are many organizations in Emerging Markets which could benefit from our experiences ___skills,

7.5 General Discussion

Figure 7.3 shows the construction noise being the most detrimental masker for both Otter and Ava, reducing the accuracy rate down to 66.51% and 69.12% respectively.

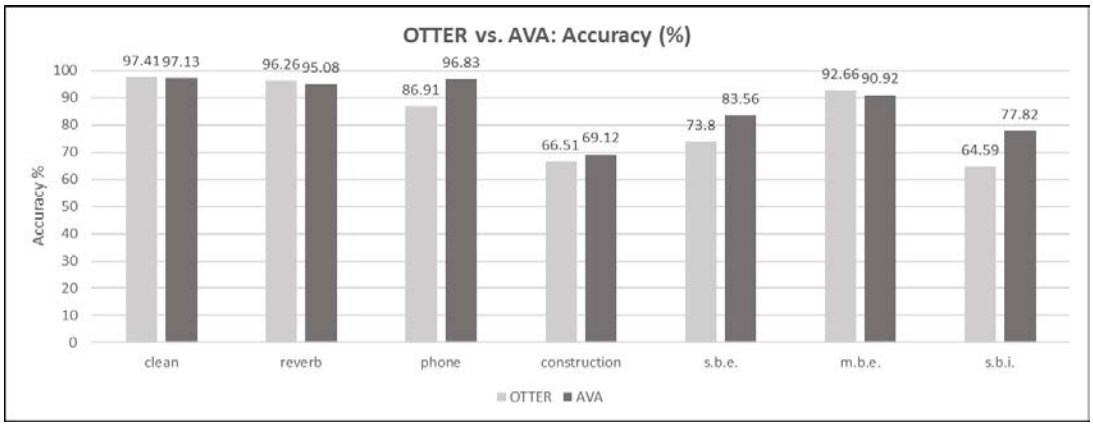

Figure 7.3. A comparison of overall accuracy of Otter and Ava transcripts across the noise conditions.

Also highly detrimental were the single babble informational noise (64.59% for Otter and 77.82% for Ava), which in this study was speech in a real language, and single babble energetic noise (73.80% for Otter and 83.56% for Ava). Importantly, both single babble energetic and multi babble energetic maskers were created by a babble generator not using a real language, but language-like sounds and syllables. The reason these were used instead of real languages was that the two programs would otherwise try to recognize the noise elements and translate them. In that case they would, in fact, act as informational maskers. Interestingly, both programs seemed to be reasonably resistant to an acoustically degraded speech signal, with only Otter dropping below the 90% accuracy mark in the phone condition. The greatest difference in performance, that of 13.23% in Ava's favor, was observed in the single babble informational condition, which proved to be the most detrimental for Otter. One of the most remarkable findings of this study is that under the present testing conditions, the findings were not consistent with Weigel (2021), who found Otter to be more reliable than Ava in

read and spontaneous academic settings. In the present study, both programs performed quite similarly, and, understandably, their performance considerably deteriorated only when presented with the most challenging energetic and informational maskers. Unfortunately, it is not possible to compare these scores with the accuracy rates claimed by the two manufacturers, since they never provided any data measured in noise. However, Otter did provide a disclaimer stating that "background noise is generally the biggest factor in inaccurate transcriptions or missing audio," where a high level of background noise can cause the program to "misspell what that person is saying or ignore it completely" (Troubleshooting Audio Problems, n.d.). In addition, Ava claimed to have "powerful filters for background noise," but also warned its accuracy can decrease with noise, "especially with conversation chatter (restaurant, coffee shop or bar)" (Help – The Ava Captions are not Accurate, n.d.). Thus, it is safe to say that both programs performed quite well, with Otter being a bit more sensitive to the narrowband speech signal. However, the transcripts show that both programs could make up to ten successive omissions in this condition, rendering parts of the message impossible to understand. Therefore, given the rise of popularity of online environments in which these two programs can be expected to be used, it is a priority that their filtering algorithms be enhanced. Finally, based on the results obtained in this study, it is opinion of this author that either program would be a welcome asset primarily in an academic setting, where online (real-time) transcription services can improve accessibility. Given the still largely unresolved issue of inaccurate punctuation, should an academic or professional institution

rely on these programs as their source of lecture or meeting notes, the transcripts must be proofread and edited prior to being used. Even though such transcripts "eliminate the subjective nature of notes when they are written by others," at the moment, "human involvement seems necessary to achieve accuracy high enough for academia" (Weigel, 2021: 58). When it comes to using these two programs as interpreting aids, careful selection of microphones and speaker placement is a must, as well as a noise-free environment.

Finally, future studies should investigate the ability of human participants to tolerate inaccurate transcriptions – something that could vary remarkably from one type of text to another, but also, from one context to another.

CHAPTER 8: Conclusion, Implications, and Limitations

8.1 Conclusion

The theoretical framework presented in the literature review described speech perception and spoken-word recognition as both auditory and cognitive processes primarily relying on working memory, which gets particularly taxed in adverse listening environments. In addition, reviewed were the most common types of noise found in typical listening environments. The present study investigated the effects of noise masking on speech perception and spoken word recognition found in such environments by conducting three experiments: listening span task, listening comprehension task, and shadowing. All three experiments included six different types of noise maskers presented with the stimuli. By doing so, the study tried to answer the following research questions:

- Are speech perception and spoken word comprehension equally affected by noise maskers?
- What type of commonly occurring noise maskers has the most detrimental effect on speech perception and spoken word comprehension?
- How can the present findings contribute to research on interpreting?

If shadowing is considered primarily a task relying on speech perception, as commonly described in the literature, with the other two tasks being spoken word comprehension and semantic inference tasks, the results indicate that speech perception and spoken word comprehension were not equally affected by noise maskers. The results also indicate that informational masking is the most detrimental to speech perception, while energetic masking and sound degradation

are most detrimental to spoken word recognition. Importantly, the categories of energetic and informational masking, as well as degraded speech signal seem to be too broad and must be used with caution, since not all maskers belonging to one category had the same effect on performance. Finally, the present findings could potentially be of significance for interpreting studies, for they propose useful exercises which could benefit students of interpreting.

The study cannot offer any generalization of reliable predictors of performance on speech perception and spoken word recognition tasks in adverse noise conditions due to individual differences in perceptual and cognitive abilities which are reflected in the variance in the results. However, the results of the study suggest that these types of tasks – commonly used in audiology and cognitive psychology – be incorporated in future training of interpreting for their ability to improve memory function and help students become accustomed to some of the most commonly occurring noise maskers in a professional environment. This study will hopefully make a valuable contribution to the growing body of literature on interpreting studies.

Finally, while the results of this study suggested that noisy backgrounds affect speech perception and spoken-word recognition – both processes significantly relying on working memory – it may be possible that auditory noise maskers also affect other cognitive processes (in non-auditory tasks) relying on attention and working memory. Further research should be conducted in order to find out more about this relationship.

8.2 Implications for Interpreter Training

While the role of working memory capacity in the context of simultaneous interpreting has been addressed in detail in Chapter 2, it is worth asserting that memory exercises such as the listening span task and the listening comprehension task could be beneficial if included in interpreter training programs to help students increase memory function and improve focus. Due to its limited capacity, working memory needs to be constantly updated through a process that involves "selecting and maintaining available relevant information (…) and removing it once no longer relevant" (Palladino & Artuso, 2018: 45). When engaged in a challenging cognitive task of integrating new information with what has already been stored, it is the updating process that "allows modification of part of a representation in memory, whilst keeping the rest of the representation unaltered" (Palladino & Artuso, 2018: 45). The author of this thesis is of the opinion that introducing a noise component to any memory task adds another cognitively stimulating real-life dimension to the task. This would, in the long run, not only help students of interpreting improve their memory skills, but also get accustomed to working in a noisy environment, which is an inevitable part of this profession. Similarly to listening comprehension, simultaneous interpreting is based on the process of active listening where the activated mechanisms for comprehension allow the listener to see beyond the mere word meanings (Herrero, 2017). At the same time, the interpreter is engaged in the process of decoding the message through both nonverbal and verbal channels (Herrero, 2017). Like the previously described modified listening span task, a modified version of the listening

comprehension task employed in this study is believed to bring a real-life challenge to this otherwise valuable exercise.

Finally, given the major challenge of simultaneous interpreting – that of being accustomed to speaking and listening at the same time – the importance of speech shadowing exercises as part of the training period cannot be stressed enough. However, the challenge of having to alternate between perception and production processes can be successfully overcome with regular practice of shadowing (Schweda-Nicholson, 1990). Some proponents of shadowing suggest that the task be enhanced with additional load elements: Schweda-Nicholson (1990) proposes that students should write numbers from one to one hundred, or the days of the week while shadowing; Kalina (2000) suggests that shadowing content should contain mistakes, so that students would be required to correct them; Guichot de Fortis (2020) suggests the time lag be varied in order to increase difficulty. The present study used an enhanced version of the shadowing exercise, introducing an additional auditory component in the form of noise maskers. The author is of the opinion that students of interpreting would most certainly benefit from such a modified exercise as the kind of maskers introduced in this study are the ones commonly encountered in real-life professional settings. This opinion is in line with Leek and Watson (1984) who found that the detrimental effects of auditory masking can be overcome with training. The opinion also reflects the fact that the perceptual-learning mechanisms were found to improve speech perception after engaging in shadowing (Kraljic & Samuel, 2005; Mitterer & Müsseler, 2013; Dias et al., 2021).

8.3 Limitations

The results of the present study are limited due to several factors, and these factors should be borne in mind when conducting follow-up research.

In the first experiment, participants performed almost identically across all seven conditions. More research is needed to precisely understand what caused such unexpectedly uniform results; including a free recall task should rule out the nature of the task or its difficulty being responsible for the results, while recruiting participants from diverse professional backgrounds should determine whether participant selection was responsible for such uniform performance.

This study deliberately avoided recruiting students of interpreting or professional interpreters as they have been proved to possess memory advantage over non-interpreters outperforming them on all memory and spoken word recognition tasks (Köpke, et al., 2011; Yudes et al., 2013; Elmer et al., 2014; Hiltunen et al., 2016). The obtained scores would not have necessarily reflected performance differences between the two (or more) groups as a result of adverse conditions.

Given the online environment, necessitated by the restrictions to in-person testing in place at the McMaster University during the experimental phase of the project, several technical imperfections could not be avoided: variables such as sound volume and room acoustics could not be controlled; similarly, the quality of headphones most probably varied across participants; and, ideally, participants' hearing would have been assessed by audiometric tests prior to experiment to ensure normal-hearing thresholds. Were it not for the restrictions, the tasks would

not have been conducted in one session, which would have allowed multiple trials in each of the noise conditions.

Lastly, it is recommended that the study be replicated with a subject population with a different level of education in order to find out whether, and if yes, to what degree, educational background affects performance.

References

Adams-Goertel, R. (2013). Prosodic elements to improve pronunciation in English language learners: A short report. *Applied Research on English Language*, 2, 117-128.

Ahangari, S., Rahbar, S., & Maleki, S.E. (2015). Pronunciation or listening enhancement: Two birds with one stone *International Journal of Language and Applied Linguistics, 1,* 13-19.

Alloway, T. P., & Alloway, R. G. (2010). Investigating the predictive roles of working memory and IQ in academic attainment. *Journal of Experimental Child Psychology, 106*(1), 20-29.

Alonzo, C. N., Yeomans-Maldonado, G., Murphy, K. A., & Bevens, B. (2016). Predicting second grade listening comprehension using Prekindergarten measures. *Topics in Language Disorders, 36*(4), 312-333.

Andéol, G., Suied, C., Scannella, S., & Dehais, F. (2017). The spatial release of cognitive load in cocktail party is determined by the relative levels of the talkers. *Journal of the Association for Research in Otolaryngology, 18*(3), 457-464.

Anderson, J. R. (2009). *Cognitive psychology and its implications, seventh edition.* New York: Worth Publishers.

Anderson, S., & Kraus, N. (2012). Neural encoding of speech and music: Implications for hearing speech in noise. *Seminars in Hearing, 33*(02), 207-212.

Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentais, prosody, and syllable structure. *Language Learning, 42*(4), 529-555.

Armstrong, G. B., & Sopory, P. (1997). Effects of background television on phonological and visuo-spatial working memory. *Communication Research, 24*(5), 459-480.

Assmann, P. & Summerfield, Q. (2006). The perception of speech under adverse conditions. In S. Greenberg, W.A. Ainsworth & R.R. Fay, (eds.) *Speech processing in the auditory system*, 231-308. New York: Springer.

Auer, E. T. (2009). Spoken word recognition by eye. *Scandinavian Journal of Psychology, 50*(5), 419-425.

Babcock, L., Capizzi, M., Arbula, S., & Vallesi, A. (2017). Short-term memory improvement after simultaneous interpretation training. *Journal of Cognitive Enhancement, 1*(3), 254-267.

Babcock, L., & Vallesi, A. (2015). Are simultaneous interpreters expert bilinguals, unique bilinguals, or both? *Bilingualism: Language and Cognition, 20*(2), 403-417.

Babel, M., & Bulatov, D. (2011). The role of fundamental frequency in phonetic accommodation. *Language and Speech, 55*(2), 231-248.

Baddeley, A. D. (2017). Modularity, working memory and language acquisition. *Second Language Research, 33*(3), 299-311.

Baddeley, A. (2007). *Working memory, thought, and action.* Oxford: Oxford University Press.

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105*(1), 158-173.

Baddeley, A. D., & Dale, H. C. A. (1966). The effect of semantic similarity on retroactive interference in long- and short-term memory. *Journal of Verbal Learning & Verbal Behavior, 5*(5), 417-420.

Baigorri-Jalón, J. (2004). *De Paris à Nuremberg: naissance de l'interprétation de conférence.* Les Presses de l'Universite d'Ottawa.

Baigorri-Jalón, J., & Takeda, K. (2016). Introduction. In J. Baigorri-Jalón & K. Takeda (eds.), *New insights in the history of interpreting*, vii-xvi. Amsterdam: John Benjamins.

Bajo,M. T., Padilla, F., & Padilla, P. (2000). Comprehension processes in simultaneous interpreting. In A. Chesterman, N. Gallardo San Salvador & Y. Gambier (eds.), *Translation in context*, 127-142. Amsterdam: John Benjamins.

Banks, B., Gowen, E., Munro, K. J., & Adank, P. (2015). Audiovisual cues benefit recognition of accented speech in noise but not perceptual adaptation. *Frontiers in Human Neuroscience, 9,* 422. https://doi.org/10.3389/fnhum.2015.00422

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge: Cambridge University Press.

BBC Sound Effects. (n.d.). Retrieved from https://sound-effects.bbcrewind.co.uk

Berne, J. E. (2004). Listening comprehension strategies: A review of the literature. *Foreign Language Annals*, *37*(4), 521-531.

Best, C. (1995). A direct realist view of cross-language speech. In W. Strange (ed.), *Speech perception and linguistic experience,* 171-204. Baltimore: York Press.

Blauert, J., & Braasch, J. (2005). Acoustic communication: The precedence effect. In *Proceedings of the Forum Acusticum 2005,* Budapest.

Bodie, G. D., & Worthington, D. L. (2018). Measuring listening. In G. D. Bodie & D. L. Worthington (eds.), *The sourcebook of listening research: Methodology and measures*, 21-44. Hoboken: John Wiley & Sons.

Borius, P. Y., Giussani, C., Draper, L., & Roux, F. E. (2012). Sentence translation in proficient bilinguals: a direct electrostimulation brain mapping. *Cortex, 48,* 614-622.

Bradlow, A., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, *106*(2), 707-729.

Braun, A., & Hahn, H. (2011). The Effect of voice similarity on stream segregation. *ICPhS 2011,* 360-363.

Brazil, D. (1995). *A grammar of speech*. Oxford: Oxford University Press.

Bregman, A.S. (1990). Auditory scene analysis: The perceptual organization of sound. Cambridge: MIT Press.

Bregman, A. S., & Campbell, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology, 89*(2), 244-249.

Bronkhorst A. W. (2015). The cocktail-party problem revisited: early processing and selection of multi-talker speech. *Attention, Perception & Psychophysics, 77*(5), 1465-1487.

Brouwer, S., & Bradlow, A. R. (2016). The temporal dynamics of spoken word recognition in adverse listening conditions. Journal of *Psycholinguistic Research, 45*(5), 1151-1160.

Brouwer, S., Van Engen, K.J., Calandruccio, L. & Bradlow, A.R. (2012). Linguistic contributions to speech-on-speech masking for native and non-native listeners: language familiarity and semantic content. *The Journal of the Acoustical Society of America, 131*(2), 1449-1464.

Brown, G. (2008). Selective listening. *System, 36*(1), 10-21.

Brown, G. (1995). *Speakers, listeners and communication.* Cambridge: Cambridge University Press.

Brown, G., Gillian, B., & Yule, G. (1983). *Discourse analysis*. Cambridge University Press.

Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *The Journal of the Acoustical Society of America, 109*(3), 1101-1109.

Brungart, D. S., Simpson, B. D., Ericson, M. A., & Scott, K. R. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *The Journal of the Acoustical Society of America, 110*(5), 2527-2538.

Buck, G. (2001). *Assessing listening*. Cambridge: Cambridge University Press.

Bürki-Cohen, J., Miller, J. L., & Eimas, P. D. (2001). Perceiving non-native speech. *Language and Speech, 44*(2), 149-169.

Byrne, J. H., & Menzel, R. (2017). *Learning and memory: A comprehensive reference. Learning theory and behavior*. Amsterdam: Elsevier.

Calvo, N., Ibáñez, A., & García, A. M. (2016). The Impact of bilingualism on working memory: A null effect on the whole may not be so on the parts. *Frontiers in Psychology, 7*, 265. doi: 10.3389/fpsyg.2016.00265

Carey, R. (2010). Hard to ignore: English native speakers in ELF research. *Helsinki English Studies*, *6*, 88-101.

Carlet, A., & Cebrian, J. (2014). Training Catalan speakers to identify L2 consonants and vowels: A short-term high variability training study. *Concordia Working Papers in Applied Linguistics, 5*, 85-98.

Carlile, S., & Corkhill, C. (2015). Selective spatial attention modulates bottom-up informational masking of speech. *Scientific Reports*, *5*(1). doi:10.1038/srep08662

Carsten, S. (2017). A different shade of shadowing: Source text to source text as efficient simultaneous processing exercise. *Vertimo studijos, 6*(6), 9-33.

Cary, E. (1956) *La traduction dans le monde moderne*. Georg.

Cassidy, G., & MacDonald, R. A. (2007). The effect of background music and background noise on the task performance of introverts and extraverts. *Psychology of Music, 35*(3), 517-537.

Cerezo Herrero, E. (2017). A critical review of listening comprehension in interpreter training: The case of Spanish translation and interpreting degrees. *Porta Linguarum Revista Interuniversitaria de Didáctica de las Lenguas Extranjeras*, 7-22. doi:10.30827/digibug.54000

Chakalov, I., Draganova, R., Wollbrink, A., Preissl, H., & Pantev, C. (2013). Perceptual organization of auditory streaming-task relies on neural entrainment of the stimulus-presentation rate: MEG evidence. *BMC Neuroscience, 14*(120). https://doi.org/10.1186/1471-2202-14-120

Chernov, S. (2016). “At the dawn of simultaneous interpreting in the USSR.” In Takeda, K., & Baigorri-Jalón, J. (eds.), *New insights in the history of interpreting*, 135-166. Amsterdam: John Benjamins Publishing Company.

Chmiel, A. (2018). In search of the working memory advantage in conference interpreting – Training, experience and task effects. *International Journal of Bilingualism, 22*(3), 371-384.

Christoffels, I., Degroot, A., & Kroll, J. (2006). Memory and language skills in simultaneous interpreters: The role of expertise and language proficiency. *Journal of Memory and Language, 54*(3), 324-345.

Ciocca, V. (2008). The auditory organization of complex sounds. *Frontiers in Bioscience, 13,* 148-169.

Coltheart, M., Besner, D., Jonasson, J. T., & Davelaar, E. (1979). Phonological encoding in the lexical decision task. *Quarterly Journal of Experimental Psychology*, *31*(3), 489-507.

Colzato, L. S., Bajo, M. T., van den Wildenberg, W., Paolieri, D., Nieuwenhuis, S., & La Heij, W. (2008). How does bilingualism improve executive control? A comparison of active and reactive inhibition mechanisms, *Journal of Experimental Psychology*, *34*(2), 302-312.

Conrad, L. (1985). Semantic versus syntactic cues in listening comprehension. *Studies in Second Language Acquisition, 7*(1), 59-69.

Conrad, R. (1964). Acoustic confusions in immediate memory. *British Journal of Psychology*, *55*(1), 75-84.

Conway, A.R.A., Kane, M.J., Bunting, M.F., Hambrick, D. Z., Wilhelm, O., & Engle, R.W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review 12*, 769-786.

Cooke, M., García Lecumberri, M.L., & Barker, J., (2008). The foreign language cocktail party problem: energetic and informational masking effects in non-native speech perception. *Journal of the Acoustical Society of America, 123*(1), 414-427.

Costa, A., & Santesteban,M. (2004). Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *Journal of Memory and Language, 50,* 491-511.

Cowan N. (2014). Working Memory Underpins Cognitive Development, Learning, and Education. *Educational psychology review*, *26*(2), 197-223.

Cristia, A., Seidl, A., Vaughn, C., Schmale, R., Bradlow, A., & Floccia, C. (2012). Linguistic processing of accented speech across the lifespan. *Frontiers in Psychology, 3,* 1-12. doi:10.3389/fpsyg.2012.00479

Cutler, A. (2012*). Native listening: Language experience and the recognition of spoken words*. Cambridge: MIT Press.

Cusack, R., Decks, J., Aikman, G., & Carlyon, R. P. (2004). Effects of location, frequency region, and time course of selective attention on auditory scene analysis. *Journal of Experimental Psychology: Human Perception and Performance, 30*(4), 643-656.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior, 19*(4), 450-466.

Daró, V. (2002). Experimental studies on memory in conference interpretation. *Meta, 42*(4), 622-628.

Del Maschio, N., Sulpizio, S., Fedeli, D., Ramanujan, K., Ding, G., Weekes, B. S., Cachia, A., & Abutalebi, J. (2018). ACC sulcal patterns and their modulation on cognitive control efficiency across lifespan: A neuroanatomical study on bilinguals and monolinguals. *Cerebral Cortex, 29*(7), 3091-3101.

Delisle, J. (2020) Canadian parliamentary review. Retrieved from Fifty Years of Simultaneous Interpretation https://www.revparl.ca/english/issue.asp?param=193&art=1333

Delisle, J., & Woodsworth, J. (2012). *Translators through history*. John Benjamins Publishing.

Dias, J. W., Vazquez, T. C., & Rosenblum, L. D. (2021). Perceptual learning of phonetic convergence. *Speech Communication, 133*, 1-8.

Díaz-Galaz, S. (2014). Individual factors of listening comprehension in a second language: Implications for interpreter training. *Synergies Chili, 10*, 31-40.

Di Paolo, E., & De Jaegher, H. (2012). The interactive brain hypothesis. *Frontiers in Human Neuroscience, 6*, Article 163. doi:10.3389/fnhum.2012.00163

Derry, S. J., & Murphy, D. A. (1986). Designing systems that train learning ability: From theory to practice. *Review of Educational Research*, *56*(1), 1-39.

Dong, Y., Liu, Y., & Cai, R. (2018). How does consecutive interpreting training influence working memory: A longitudinal study of potential links

between the two. *Frontiers in Psychology, 9*.1-12. doi:10.3389/fpsyg.2018.00875

Dong, Y., Gui, S., & Macwhinney, B. (2005). Shared and separate meanings in the bilingual mental lexicon. *Bilingualism: Language and Cognition, 8,* 221-238.

Droop, M., & Verhoeven, L. (1998). Background knowledge, linguistic complexity, and second-language reading comprehension. *Journal of Literacy Research, 30*(2), 253-71.

Drozdova, P., van Hout, R., & Scharenborg,O. (2016). Lexically-guided perceptual learning in non-native listening. *Bilingualism: Language and Cognition, 19*(5), 914-920.

Duranti, A., Goodwin, C., & Professor of Applied Linguistics Charles Goodwin. (1992). *Rethinking context: Language as an interactive phenomenon*. Cambridge: Cambridge University Press.

Elmer, S., Hänggi, J., & Jäncke, L. (2014). Processing demands upon cognitive, linguistic, and articulatory functions promote grey matter plasticity in the adult multilingual brain: Insights from simultaneous interpreters. *Cortex, 54*, 179-189.

EMCI. (n.d.). Retrieved from https://www.emcinterpreting.org

Engel, A.K. (2010). Directive minds: how dynamics shapes cognition. In J. Stewart, O. Gapenne & E. Di Paolo (eds.), *Enaction: Towards a new paradigm for cognitive science*, 219-243. Cambridge: MIT Press.

Ericsson K.A, & Lehmann A.C. (1996). Expert and exceptional performance: evidence of maximal adaptation to task constraints. *Annual Review of Psychology, 47*, 273-305.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102*(2), 211-245.

Ezzatian, P., Li, L., Pichora-Fuller. M. K. & Schneider, B. A. (2012) The effect of energetic and informational masking on the time-course of stream segregation: Evidence that streaming depends on vocal fine structure cues, *Language and Cognitive Processes, 27*(7-8), 1056-1088.

Faris, M. M., Best, C. T., & Tyler, M. D. (2018). Discrimination of uncategorized non-native vowel contrasts is modulated by perceived overlap with native phonological categories. *Journal of Phonetics, 70*, 1-19.

Farrar, D. (2020). No more in-person classes and exams: President's letter. Retrieved from https://covid19.mcmaster.ca/a-letter-from-our-president/

Feldman, J. (2003). The simplicity principle in human concept learning. *Current Directions in Psychological Science, 12*(6), 227-232.

Feng, Z. (2020). Effects of identification and pronunciation training methods on L2 speech perception and production: Training adult Japanese speakers to perceive and produce English /r/-/l/. *Studies in Applied Linguistics and TESOL, 20*(2), 57-83.

Flege, J. E. (1995). Second language speech learning: Theory, findings and problems. In W. Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research*, 233-277. Walmgate: York Press.

Florit, E., Roch, M., Altoè, G., & Levorato, M. C. (2009). Listening comprehension in preschoolers: The role of memory. *British Journal of Developmental Psychology, 27*(4), 935-951.

Fontan, L., Farinas, J., Segura, B., Stone, M. A., & Füllgrabe, C. (2020). Using automatic speech recognition to predict aided speech-in-noise intelligibility. In *Proceedings of the Speech-In-Noise Workshop*, Toulouse.

Fowler, C. (2003). Rapid access to speech gestures in perception: Evidence from choice and simple response time tasks. *Journal of Memory and Language, 49*(3), 396-413.

Fox, R., Corretjer, O., & Webb, K. (2019). Benefits of foreign language learning and bilingualism: An analysis of published empirical research 2012–2019. *Foreign Language Annals, 52*(4), 699-726.

Friedrich, C. K., Felder, V., Lahiri, A., & Eulitz, C. (2013). Activation of words with phonological overlap. *Frontiers in Psychology, 4*. doi:10.3389/fpsyg.2013.00556

Fung, W., & Swanson, H. L. (2017). Working memory components that predict word problem solving: Is it merely a function of reading, calculation, and fluid intelligence? *Memory & Cognition, 45*(5), 804-823.

Gao, X., Yan, T., Huang, T., Li, X., & Zhang, Y. X. (2020). Speech in noise perception improved by training fine auditory discrimination: far and applicable transfer of perceptual learning. *Scientific Reports, 10*(1), 19320. https://doi.org/10.1038/s41598-020-76295-9

Galloway, N., & Rose, H. (2015). *Introducing Global Englishes*. Milton Park: Routledge.

García-Morales, M. G. (2016). Translators and interpreters during the Spanish Inquisition. *Lebende Sprachen, 61*(2), 353-367.

Gathercole, S. E., Durling, E., Evans, M., Jeffcock, S., & Stone, S. (2008). Working memory abilities and children's performance in laboratory analogues of classroom activities. *Applied Cognitive Psychology, 22*(8), 1019-1037.

Gerver, D. (1975). A psychological approach to simultaneous interpretation. *Meta: Journal des traducteurs*, *20*(2), 119-128.

Gerver, D., Longley, P. E., Long, J., & Lambert, S. (2002). Selection tests for trainee conference interpreters. *Meta*, *34*(4), 724-735.

Gibson, J. J. (1966). *The Senses Considered as Perceptual Systems.* Boston: Houghton Mifflin.

Goldstone, R.L. (1998). Perceptual learning. *Annual Review of Psychology, 49*, 585-612.

Goldstein, L., & Fowler, C. A. (2003). "Articulatory phonology: A phonology for public language use." In N.O. Schiller & A.S. Meyer (eds.), *Phonetics and phonology in language comprehension and production*, 159-207. Berlin: Mouton de Gruyter.

Golestani, N., Rosen, S., & Scott, S.K. (2009). Native-language benefit for understanding speech-in-noise: the contribution of semantics. *Bilingualism, 12*(3), 385-392.

Grant, K. W., & Seitz, P. (2000). The use of visible speech cues for improving auditory detection of spoken sentences. *The Journal of the Acoustical Society of America, 108*(3), 1197-1208.

Grosjean, F. (2012). Bilingualism: A short introduction. In F. Grosjean & P. Li (eds.), *The psycholinguistics of bilingualism*, 5-25. Hoboken: John Wiley & Sons.

Guediche, S., Blumstein, S. E., Fiez, J. A., & Holt, L. L. (2014). Speech perception under adverse conditions: Insights from behavioral, computational, and neuroscience research. *Frontiers in Systems Neuroscience, 7*. doi:10.3389/fnsys.2013.00126

Guichot de Fortis, C. (2020). Shadowing. *Interpreter Rraining Resources*. Retrieved from https://interpretertrainingresources.eu/language/shadowing/

Guise, K. (2020). Translating and interpreting the Nuremberg trials. Retrieved from https://www.nationalww2museum.org/war/articles/translating-and-interpreting-nuremberg-trials

Hamada, Y. (2017). *Teaching and Learning Shadowing for Listening: Developing Bottom-Up Listening Skills for Language Learners*. Milton Park: Routledge.

Hamada, Y. (2016). Shadowing: who benefits and how? Uncovering a booming EFL teaching technique for listening comprehension. *Language Teaching Research*, *20*(1), 35-52.

Hamada, Y., & Suzuki, S. (2020). Listening to global Englishes: Script-assisted shadowing. *International Journal of Applied Linguistics*, *31*(1), 31-47.

Hambrick, D. Z., & Engle, R. W. (2002). Effects of domain knowledge, working memory capacity, and age on cognitive performance: An investigation of the knowledge-is-Power hypothesis. *Cognitive Psychology, 44*(4), 339-387.

Harley, B. (2000). Listening strategies in ESL: Do age and L1 make a difference? *TESOL Quarterly*, *34*(4), 769.

Hay, J., Nolan, A., & Drager, K. (2006). From *fush* to *feesh*: Exemplar priming in speech perception. *The Linguistic Review, 23*(3), 351-379.

Hayes, J. R. (1979). The Complete problem-solver. Philadelphia: Franklin Institute.

Healey, E., & Howe, S. W. (1987). Speech shadowing characteristics of stutterers under diotic and dichotic conditions. *Journal of Communication Disorders, 20*(6), 493-506.

Helfer, K. S., & Freyman, R. L. (2005). The role of visual speech cues in reducing energetic and informational masking. *The Journal of the Acoustical Society of America, 117*(2), 842-849.

Henderson, J. (1982). Some psychological aspects of simultaneous interpretation. *Incorporated Linguist, 21*(4), 149-152.

Heynold, C. (1994). "Interpreting at the European Commission." In C. Dollerup & A. Lindegaard (eds.), *Teaching translation and interpreting 2: Insights, aims and visions*, 11-18. Amsterdam: John Benjamins.

Hilchey, M. D., Saint-Aubin, J., & Klein, R. M. (2015). Does bilingual exercise enhance cognitive fitness in traditional non-linguistic executive processing tasks?. In J. Schwieter (ed.). *The Cambridge Handbook of Bilingual Processing*, 586-613. Cambridge: Cambridge University Press.

Hiltunen, S., Pääkkönen, R., Vik, G., & Krause, C. M. (2016). On interpreters' working memory and executive control. *International Journal of Bilingualism, 20*(3), 297-314.

Hiramatsu, S. (2000). A Differentiated/integrated approach to shadowing and repeating. https://core.ac.uk/download/pdf/143641870.pdf

Hsia, H. J. (1977). Redundancy: Is it the lost key to better communication?. *Educational Communication & Technology, 25*(1), 63-85.

Hu, Y., Tahmina, Q., Runge, C., & Friedland, D. R. (2013). The perception of telephone-processed speech by combined electric and acoustic stimulation. *Trends in amplification, 17*(3), 189–196.

Imhof, M. (2018). Listening Span Tests. In D. L. Worthington & G. D. Bodie (eds.), *The Sourcebook of Listening Research: Methodology and Measures*, 394-401. Hoboken: John Wiley & Sons.

Jackson, C. N., & O'Brien, M. G. (2011). The interaction between prosody and meaning in second language speech Production1. *Die Unterrichtspraxis/Teaching German*, *44*(1), 1-11.

Jackson-Eade, J.A.B. (2018). The slave-interpreter system in the fifteenth-century Atlantic world. *Global Histories: A Student Journal*, *4*(2), 3-24.

James, C. J. (1984). Are you listening? The practical components of listening comprehension. *Foreign Language Annals*, *17*(2), 129-134.

Jenkins, J. (2009). English as a lingua franca: Interpretations and attitudes. *World Englishes*, *28*(2), 200-207.

Jin, S.H., & Liu, C. (2012). English sentence recognition in speech-shaped noise and multi-talker babble for English-, Chinese-, and Korean-native listeners. *The Journal of the Acoustical Society of America*, *132*(5), 391-397.

Johnson S.P. (1997). Young Infants' Perception of Object Unity: Implications for Development of Attentional and Cognitive Skills. *Current Directions in Psychological Science*, *6*(1), 5-11.

Kaandorp, M. W., De Groot, A. M., Festen, J. M., Smits, C., & Goverts, S. T. (2015). The influence of lexical-access ability and vocabulary knowledge on measures of speech recognition in noise. *International Journal of Audiology*, *55*(3), 157-167.

Kadota, S. (2019). *Shadowing as a practice in second language acquisition: Connecting inputs and outputs*. Milton Park: Routledge.

Keiser, W. (2004). L'interprétation de conférence en tant que profession et les précurseurs de l'Association Internationale des Interprètes de Conférence (AIIC) 1918-1953. *Meta*, *49*(3), 576-608.

Keskin, H. K., Arı, G., & Baştuğ, M. (2019). Role of prosodic reading in listening comprehension. *International Journal of Education and Literacy Studies*, *7*(1), 59.

Khaghaninejad, M. S., & Maleki, A. (2015). The effect of explicit pronunciation instruction on listening comprehension: Evidence from Iranian English learners. *Theory and Practice in Language Studies*, *5*, 1249-1256.

Kidd, G. J., Mason, C., Rohtla, T., & Deliwala, P. (1998). Release from informational masking due to the spatial separation of sources in the identification of nonspeech auditory patterns. *The Journal of the Acoustical Society of America*, *104*(1), 422-431.

King, A., & Eckersley, R. (2019). *Statistics for biomedical engineers and scientists: How to visualize and analyze data*. London: Academic Press.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge: Cambridge University Press.

Kintsch, W., & Rawson, K. A. (2005). Comprehension. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook*, 209-226. Malden, MA: Blackwell.

Klatte, M., Bergström, K., & Lachmann, T. (2013). Does noise affect learning? A short review on noise effects on cognitive performance in children. *Frontiers in Psychology, 4*. doi:10.3389/fpsyg.2013.00578

Koelega, H. S., & Brinkman, J. (1986). Noise and vigilance: An evaluative review. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *28*(4), 465-481.

Kohnert, K., Ebert, K. D., & Pham, G. T. (2021). Language disorders in bilingual children and adults (3rd ed.). San Diego: Plural Publishing.

Köpke, B., & Signorelli, T. M. (2011). Methodological aspects of working memory assessment in simultaneous interpreters. *International Journal of Bilingualism*, *16*(2), 183-197.

Köpke, B., & Nespoulous, J. (2006). Working memory performance in expert and novice interpreters. Interpreting. *International Journal of Research and Practice in Interpreting*, *8*(1), 1-23.

Kraljic, T., & Samuel, A.G. (2007). Perceptual adjustments to multiple speakers. *Journal of Memory and Language*, *56*(1), 1–15.

Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology, 51*(2), 141-178.

Kraus, N. (1999). Speech-sound perceptual learning. *The Hearing Journal, 52*(11), 64-66.

Kroll, J. F., Dussias, P. E., Bice, K., & Perrotti, L. (2015). Bilingualism, Mind, and Brain. *Annual Review of Linguistics, 1*, 377-394.

Labia, L., Shtrepi, L., & Astolfi, A. (2020). Improved room acoustics quality in meeting rooms: Investigation on the optimal configurations of sound-absorptive and sound-diffusive panels. *Acoustics, 2*(3), 451-473.

Lambert, S. (2004). Shared attention during sight translation, sight interpretation and simultaneous interpretation. *Meta, 49*(2), 294-306.

Lambert, S. (1991). Aptitude testing for simultaneous interpretation at the University of Ottawa. *Meta, 36*(4), 586-594.

Lambert, S., Daró, V., & Fabbro, F. (2002). Focalized attention on input vs. output during simultaneous interpretation: Possibly a waste of effort! *Meta, 40*(1), 39-46.

Laver, J., & John, L. (1994). *Principles of phonetics*. Cambridge: Cambridge University Press.

Lecumberri, M. L., Cooke, M., & Cutler, A. (2010). Non-native speech perception in adverse conditions: A review. *Speech Communication, 52*(11-12), 864-886.

Lee, Y. R., Trung, T. Q., Hwang, B., & Lee, N. (2020). A flexible artificial intrinsic-synaptic tactile sensory organ. *Nature Communications, 11*(1). doi:10.1038/s41467-020-16606-w

Leek, M. R., & Watson, C. S. (1984). Learning to detect auditory pattern components. *The Journal of the Acoustical Society of America, 76*(4), 1037-1044.

Leslie, L., & Caldwell, J. (2010). *Qualitative reading inventory (5th ed.)*. Bloomington: Pearson Assessments.

Lesser, M. J. (2007). Learner-based factors in L2 reading comprehension and processing grammatical form: Topic familiarity and working memory. *Language Learning, 57*(2), 229-270.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*(1), 1-36.

Litovsky, R.Y., Colburn, H.S., Yost, W.A., & Guzman, S.J. (1999). The precedence effect. *The Journal of the Acoustical Society of America, 106*(4 Pt 1).1633-1654.

Liu, H., Bates, E., Powell, T., & Wulfeck, B. (1997). Single-word shadowing and the study of lexical access. *Applied Psycholinguistics*, *18*(2), 157-180.

Liu, H. T., Squires, B., & Liu, C. J. (2016). Articulatory suppression effects on short-term memory of signed digits and lexical items in hearing bimodal-bilingual adults. *Journal of Deaf Studies and Deaf Education*, *21*(4), 362-372.

Ljung, R., & Kjellberg, A. (2009). Long reverberation time decreases recall of spoken information. *Building Acoustics, 16*(4), 301-311.

Logan, G. D. (1992). Attention and preattention in theories of automaticity. *American Journal of Psychology, 105*(2), 317-339.

Long, D. (1990). What you don't know can't help you: An exploratory study of background knowledge and second language listening comprehension. *Studies in Second Language Acquisition, 12*(1), 65-80.

Luca, F. X. (1999). Re-'interpreting' the role of the cultural broker in the conquest of La Florida. Florida International University. Retrieved from: http://www.kislakfoundation.org/prize/199901.html

Lynch, T. (1998). Theoretical perspectives on listening. *Annual Review of Applied Linguistics*, *18*, 3-19.

Macaro, E. (2006). Strategies for language learning and for language use: Revising the theoretical framework. *The Modern Language Journal*, *90*(3), 320-337.

Mackintosh, J. (1999). Interpreters are made not born. *Interpreting. International Journal of Research and Practice in Interpreting, 4*(1), 67-80.

Macnamara, B. N., & Conway, A. R. (2016). Working memory capacity as a predictor of simultaneous language interpreting performance. *Journal of Applied Research in Memory and Cognition*, *5*(4), 434-444.

Marslen-Wilson, W. D. (1985). Speech shadowing and speech comprehension. *Speech Communication*, *4*(1-3), 55-73.

Masrai, A. (2019). Exploring the impact of individual differences in aural vocabulary knowledge, written vocabulary knowledge and working memory capacity on explaining L2 learners' listening comprehension. *Applied Linguistics Review, 11*(3), 423-447.

Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes, 27*(7-8), 953-978.

McDonald, J. L. (2006). Beyond the critical period: Processing-based explanations for poor grammaticality judgment performance by late second language learners. *Journal of Memory and Language*, *55*(3), 381-401.

McLean, J. (1890). *James Evans, inventor of the syllabic system of the Cree language*. Toronto: Methodist Mission Rooms.

Mehrpour, S., & Rahimi, M. (2010). The impact of general and specific vocabulary knowledge on reading and listening comprehension: A case of Iranian EFL learners. *System*, *38*(2), 292-300.

Mendelsohn, D. (1995). Applying learning strategies in the second/foreign language listening comprehension lesson. In D. Mendelsohn & J. Rubin (eds.), *A Guide for the teaching of second language listening*, 132-150. San Diego: Dominie Press.

Micheyl, C., Shamma, S. A., & Oxenham, A.J. (2007). Hearing out repeating elements in randomly varying multitone sequences: A case of streaming?. In B. Kollmeier, G. Klump, V. Hohmann, U. Langemann, M. Mauermann, S. Uppenkamp & J.Verhey (eds.), *Hearing: From sensory processing to perception*, 267-274. Berlin: Springer.

Mitterer, H., & Müsseler, J. (2013). Regional accent variation in the shadowing task: Evidence for a Loose perception–action coupling in speech. *Attention, Perception, & Psychophysics, 75*(3), 557-575.

Monson, B. B., Hunter, E. J., Lotto, A. J., & Story, B. H. (2014). The perceptual significance of high-frequency energy in the human voice. *Frontiers in Psychology, 5*. doi:10.3389/fpsyg.2014.00587

Moore, B. C. (1986). Parallels between frequency selectivity measured psychophysically and in cochlear mechanics. *Scandinavian Audiology, Supplementum, 25*, 139-152.

Moore, B. C., & Gockel, H. E. (2012). Properties of auditory stream formation. *Philosophical Transactions of the Royal Society B: Biological Sciences, 367*(1591), 919-931.

Moore, B. C., & Tan, C. (2003). Perceived naturalness of spectrally distorted speech and music. *The Journal of the Acoustical Society of America, 114*(1), 408-419.

Morales, J., Padilla, F., Gómez-Ariza, C. J., & Bajo, M. T. (2015). Simultaneous interpretation selectively influences working memory and attentional networks. *Acta Psychologica, 155*, 82-91.

Morrison, C., Kamal, F., & Taler, V. (2018). The influence of bilingualism on working memory event-related potentials. *Bilingualism: Language and Cognition, 22*(1), 191-199.

Morton, J., and D. Bekerian. (1986). Three Ways of Looking at Memory. *Advances in Cognitive Science, 1*, 43-71.

Moulin-Frier, C., & Arbib, M. A. (2013). Recognizing speech in a novel accent: The motor theory of speech perception reframed. *Biological Cybernetics, 107*(4), 421-447.

Nees, M. A. (2016). Have we forgotten auditory sensory memory? Retention intervals in studies of nonverbal auditory working memory. *Frontiers in Psychology, 7*, 1-6. doi:10.3389/fpsyg.2016.01892

Nelson, D. G., Hirsh-Pasek, K., Jusczyk, P. W., & Cassidy, K. W. (1989). How the prosodic cues in motherese might assist language learning. *Journal of Child Language*, *16*(1), 55-68.

Nida, E. A. **(**1978). The setting of communication: A largely overlooked factor in translation. *Babel, XXIV*(3-4),114-17.

Nouwens, S., Groen, M. A., & Verhoeven, L. (2016). How working memory relates to children's reading comprehension: The importance of domain-specificity in storage and processing. *Reading and Writing, 30*(1), 105-120.

Nusbaum, H. C., & Morin, T.M. (1992). Paying attention to differences among talkers. In Y. Tohkura, E. Vatikiotis-Bateson, & Y. Sagisaka (eds.), *Speech perception, production and linguistic structure*, 113-134. Amsterdam: IOS Press.

O'Bryan, A., & Hegelheimer, V. (2009). Using a mixed methods approach to explore strategies, metacognitive awareness and the effects of task design on listening development. *Canadian Journal of Applied Linguistics*, *12*(1), 9-38.

O'Malley, J. M., Chamot, A. U., & Kupper, L. (1989). Listening comprehension strategies in second language acquisition. *Applied Linguistics*, *10*(4), 418-437.

Paap, K. R., Johnson, H. A., & Sawi, O. (2015). Bilingual advantages in executive functioning either do not exist or are restricted to very specific and undetermined circumstances. *Cortex, 69*, 265-278.

Padilla, F., Bajo, M. T., & Macizo, P. (2005). Articulatory suppression in language interpretation: Working memory capacity, dual tasking and word knowledge. Bilingualism: *Language and Cognition, 8*(3), 207-219.

Padilla, P., Bajo, M.T., & Cañas, J.J. (1995). "Cognitive processes of memory in simultaneous interpretation," In J. Tommola (ed.) *Topics in Interpreting Research*, 61-71. Turku: Centre for Translation and Interpreting.

Palladino, P., & Artuso, C. (2018). Working memory updating: Load and binding. *The Journal of General Psychology, 145*(1), 45-63.

Peters, A. M. (1977). Language learning strategies: Does the whole equal the sum of the parts? *Language, 53*, 560-73.

Phelan, M. (2001). The Interpreter's Resource, Bristol: Multilingual Matters LTD.

Pichora-Fuller, M. K., Schneider, B. A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *The Journal of the Acoustical Society of America*, *97*(1), 593-608.

Pickett, J. M., & Morris, S. R. (2000). The acoustics of speech communication: Fundamentals, speech perception theory, and technology. *The Journal of the Acoustical Society of America, 108*(4), 1373-1374.

Picou, E. M., Gordon, J., & Ricketts, T. A. (2016). The effects of noise and reverberation on listening effort in adults with normal hearing. *Ear & Hearing, 37*(1), 1-13.

Pierrehumbert J.B. (2003). Phonetic Diversity, Statistical Learning, and Acquisition of Phonology. *Language and Speech*. *46*(2-3), 115-154.

Pöchhacker, F. (2004). *Introducing interpreting studies*. Hove: Psychology Press.

Prell, C. G., Henderson, D., Fay, R. R., & Popper, A. N. (2011). Noise-induced hearing loss: Scientific advances. Berlin: Springer Science & Business Media.

Prodi, N., & Visentin, C. (2017). On the relationship between a short-term objective metric and listening efficiency data for different noise types. *The Journal of the Acoustical Society of America, 141*(5), 3972-3972.

Proverbio, A. M., Leoni, G., & Zani, A. (2004). Language switching mechanisms in simultaneous interpreters: An ERP study. *Neuropsychologia, 42,* 1636-1656.

Pulakka, H., Laaksonen, L., Yrttiaho, S., Myllylä, V., & Alku, P. (2012). Conversational quality evaluation of artificial bandwidth extension of telephone speech. *The Journal of the Acoustical Society of America, 132*(2), 848-861.

Rabinowitz, M., & Chi, M. T. H. (1987). "An interactive model of strategic processing." In S. J. Ceci (ed.), *Handbook of cognitive, social, and*

neuropsychological aspects of learning disabilities, 83-102. Hillsdale, NJ: Erlbaum.

Raichle , M. E. (2006). The brain's dark energy. *Science, 314*, 1249-1250.

Raichle , M. E., & Mintun , M. A. (2006). Brain work and brain imaging. *Annual Review of Neuroscience, 29*, 449-476.

Rämä P., Leminen, A., Koskenoja-Vainikka, S., Leminen, M., Alho, K., Kujala, T. (2018). Effect of language experience on selective auditory attention: an event-related potential study. *International Journal of Psychophysiology, 127*, 38-45.

Reinhart, P. N., & Souza, P. E. (2016). Intelligibility and clarity of reverberant speech: Effects of wide dynamic range compression release time and working memory. *Journal of Speech, Language, and Hearing Research, 59*(6), 1543-1554.

Restif-Filliozat, M. (2019). The Jesuit contribution to the geographical knowledge of India in the eighteenth century. *Journal of Jesuit Studies, 6*(1), 71-84.

Riccardi, A. (2005). On the evolution of interpreting strategies in simultaneous interpreting. *Volet interprétation, 50*(2), 753-767.

Ricker, T. J., Vergauwe, E., & Cowan, N. (2016). Decay theory of immediate memory: From Brown (1958) to today (2014). *Quarterly Journal of Experimental Psychology*, *69*(10), 1969-1995.

Robinson, P., Mackey, A., Gass, S. M. & Schmidt, R. (2012). Attention and awareness in second language acquisition. In S. Gass & A. Mackey (eds.), *The Routledge handbook of second language acquisition*. London: Routledge, 247-267.

Rodriguez-Fornells, A., De Diego Balaguer, R., & Munte, T. F. (2006). Executive control in bilingual language processing. *Language Learning, 56*, 133-190.

Rogers, C.L., Lister, J.J., Febo, D.M., Besing, J.M., & Abrams, H.B. (2006). Effects of bilingualism, noise, and reverberation on speech perception by listeners with normal hearing. *Applied Psycholinguistics, 27*(3), 465-485.

Rost, M. (2011). *Teaching and researching listening*. Boston: Allyn & Bacon.

Rost, M. (1990). *Listening in language learning*. Harlow: Longman.

Rubin, D. L. (1992). Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants. *Research in Higher Education, 33*(4), 511-531.

Rudner, M., Lyberg-Åhlander, V., Brännström, J., Nirme, J., Pichora-Fuller, M. K., & Sahlén, B. (2018). Listening comprehension and listening effort in the primary school classroom. *Frontiers in Psychology, 9*. doi:10.3389/fpsyg.2018.01193

Rumelhart, D., & McClelland, J. (1986). Parallel distributed processing. Cambridge: MIT Press.

Rumelhart, D., & McClelland, J. (1981). Interactive processing through spreading activation. In: A. M. Lesgold, C. A. Perfetti (eds.), *Interactive processes in reading*, 37-60. Milton Park: Routledge.

Rumelhart, D. E., & Ortony, A. (1977). "The representation of knowledge in memory." In R. C. Anderson, R. J. Spiro, & W. E. Montague (eds.), *Schooling and the acquisition of knowledge*, 99-135. Mahwah: Lawrence Erlbaum.

Saeki E, & Saito S. (2004). Effect of articulatory suppression on task-switching performance: implications for models of working memory. *Memory, 12*(3), 257-271.

Saito, K., Tran, M., Suzukida, Y., Sun, H., Magne, V., & Ilkan, M. (2019). How do L2 listeners perceive the comprehensibility of foreign-accented speech? Roles of L1 profiles, L2 proficiency, age, experience, familiarity and metacognition. *Studies in Second Language Acquisition 41*(5), 1133-1149.

Schank, R. (1985). Reminding and memory organization. New York: Academic Press.

Scharenborg, O., & Van Os, M. (2019). Why listening in background noise is harder in a non-native language than in a native language: A review. *Speech Communication, 108*, 53-64.

Scharenborg, O., Coumans, J. M., & Van Hout, R. (2018). The effect of background noise on the word activation process in nonnative spoken-

word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*(2), 233-249.

Schneider, B. A., Li, L., & Daneman, M. (2007). How competing speech interferes with speech comprehension in everyday listening situations. *Journal of the American Academy of Audiology, 18*(07), 559-572.

Schnell, J. (1995). Effective listening: More than just hearing. Retrieved from Educational Resources Information Center, https://eric.ed.gov/?id=ED379691

Schweda-Nicholson, N. (1990). The Role of shadowing in interpreter training. *The Interpreters' Newsletter, 3*, 33-37.

Schweda-Nicholson, N. (1987). Linguistic and extralinguistic aspects of simultaneous interpretation. *Applied Linguistics, 8*(2), 194-205.

Seeber, K. G., & Arbona, E. (2020). What's load got to do with it? A cognitive-ergonomic training model of simultaneous interpreting. *The Interpreter and Translator Trainer*, *14*(4), 369-385.

Segalowitz, D. (1986). "Skilled reading in the second language." In J. Vaid (ed.), *Language processing in bilinguals: Psycholinguistic and neuropsychological perspectives*, 3-19. Mahwah: Lawrence Erlbaum Associates.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review, 84*(2), 127-190.

Shockley, K., Sabadini, L., & Fowler, C. A. (2004). Imitation in shadowing words. *Perception & Psychophysics*, *66*(3), 422-429.

Sidaras, S., Alexander, J., & Nygaard, L. (2009). Perceptual learning of systematic variation in Spanish-accented speech. *The Journal of the Acoustical Society of America*, *125*(5), 3306–3316.

Slowiaczek, L. M. (1994). Semantic priming in a single-word shadowing task. *The American Journal of Psychology, 107*(2), 245-260.

Smiljanić, R., & Bradlow, A. R. (2008). Temporal organization of English clear and conversational speech. *The Journal of the Acoustical Society of America*, 124(5), 3171-3182.

Smiljanić, R., & Sladen, D. (2013). Acoustic and semantic enhancements for children with cochlear implants. *Journal of Speech, Language, and Hearing Research, 56*(4), 1085-1096.

Smiljanić, R., & Sladen, D. (2010). Effect of acoustic-phonetic and semantic enhancements on speech recognition for children with cochlear implants. *The Journal of the Acoustical Society of America, 128*(4), 2425-2425.

Snyder, J. S., Gregg, M. K., Weintraub, D. M., & Alain, C. (2012). Attention, awareness, and the perception of auditory scenes. *Frontiers in Psychology, (3)*, 1-17.

Snyder, J. S., & Alain, C. (2007). Toward a neurophysiological theory of auditory stream segregation. *Psychological Bulletin, 133*(5), 780-799.

Speaks, C. E. (2018). Introduction to sound: Acoustics for the hearing and speech sciences (4th ed.). San Diego: Plural Publishing.

Sperber, D., & Wilson, D. (1986). Relevance: Communication and cognition. Oxford: Blackwell.

Stæhr, L. S. (2009). Vocabulary knowledge and advanced listening comprehension in English as a foreign language. *Studies in Second Language Acquisition, 31*(04), 577.

Stansfeld, S. A., & Matheson, M. P. (2003). Noise pollution: Non-auditory effects on health. *British Medical Bulletin, 68*(1), 243-257.

Stavrakaki, S., Megari, K., Kosmidis, M. H., Apostolidou, M., & Takou, E. (2012). Working memory and verbal fluency in simultaneous interpreters. *Journal of Clinical and Experimental Neuropsychology, 34*(6), 624-633.

Stroińska, M., & Drzazga, G. (2018). "Relevance Theory, interpreting, and translation." In K. Malmkjaer (ed.), *The Routledge Handbook of Translation Studies and Linguistics*, 95-106. Milton Park: Routledge.

Sumarsih, S. (2017). The impact of shadowing technique on tertiary EFL learners' listening skill achievements. *International Journal of English Linguistics, 7*(5), 184-189.

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *The Journal of the Acoustical Society of America, 26*(2), 212-215.

Summerfield, A.Q., Culling, J. F., & Assmann, P. F. (2012). The perception of speech under adverse conditions: Contributions of spectro-temporal peaks, periodicity, and interaural timing to perceptual robustness. In S. Greenberg, & W. Ainsworth (eds.), *Listening to speech: An auditory perspective*, 223-235. Hove: Psychology Press.

Sussman, E.S. (2017). Auditory scene analysis: An attention perspective. *Journal of Speech, Language, and Hearing Research, 60*(10), 2989-3000.

Sussman, E., Ritter, W., & Vaughan, J. H. G. (1999). An investigation of the auditory streaming effect using eventrelated brain potentials. *Psychophysiology, 36*, 22-34.

Sutherland, R., Pipe, M., Schick, K., Murray, J., & Gobbo, C. (2003). Knowing in advance: The impact of prior event information on memory and event knowledge. *Journal of Experimental Child Psychology, 84*(3), 244-263.

Swerts, M., & Krahmer, E. (2008). Facial expression and prosodic prominence: Effects of modality and facial area. *Journal of Phonetics, 36*(2), 219-238.

Szabó, B. T., Denham, S. L., & Winkler, I. (2016). Computational Models of Auditory Scene Analysis: A Review. *Frontiers in neuroscience, 10*, 524. https://doi.org/10.3389/fnins.2016.005241-16

Szalárdy, O., Tóth, B., Farkas, D., György, E., & Winkler, I. (2019). Neuronal correlates of informational and energetic masking in the human brain in a multi-talker situation. *Frontiers in Psychology, 10*. doi:10.3389/fpsyg.2019.00786

Szalma, J. L., & Hancock, P. A. (2011). Noise effects on human performance: A meta-analytic synthesis. *Psychological Bulletin, 137*(4), 682-707.

Tao, L., Wang, G., Zhu, M., & Cai, Q. (2021). Bilingualism and domain-general cognitive functions from a neural perspective: A systematic review. *Neuroscience & Biobehavioral Reviews, 125*, 264-295.

Teichmann, M., Turc, G., Nogues, M., Ferrieux, S., & Dubois, B. (2012). A mental lexicon without semantics. *Neurology,79,* 606-607.

Teki, S., Chait, M., Kumar, S., Shamma, S., & Griffiths, T. D. (2013). Segregation of complex acoustic scenes based on temporal coherence. *eLife, 2*. 1-16. doi:10.7554/elife.00699

Text Compare! (n.d.). An online diff tool that can find the difference between two text files. Retrieved from https://text-compare.com

Thompson, I., & Rubin, J. (1996). Can Strategy Instruction Improve Listening Comprehension?. *Foreign Language Annals, 29,* 331-342.

Tulving, E. (1972). "Episodic and semantic memory." In E. Tulving & W. Donaldson (eds.), *Organization of memory,* 381-403. New York: Academic Press.

Uhlenbeck E.M. (1978) On the distinction between linguistics and pragmatics. In: Gerver D., Sinaiko H.W. (eds.) *Language interpretation and communication*. NATO conference series, vol 6. Boston: Springer. https://doi.org/10.1007/978-1-4615-9077-4_17

Underwood, G., & Moray, N. (1971). Shadowing and monitoring for selective attention. *Quarterly Journal of Experimental Psychology, 23*(3), 284-295.

Underwood, M. (1989). *Teaching listening*. Boston: Addison-Wesley Longman.

Valdés, G., & Angelelli, C. (2003). 4. Interpreters, interpreting, and the study of bilingualism. *Annual Review of Applied Linguistics, 23*, 58-78.

Van den Noort, M., Struys, E., Bosch, P., Jaswetz, L., Perriard, B., Yeo, S., Barisch, P., Vermeire, K., Lee, S-H., & Lim, S. (2019). Does the bilingual advantage in cognitive control exist and if so, what are its modulating factors? A systematic review. *Behavioral Sciences, 9*(27), 1-30.

Van der Feest, S. V., Blanco, C. P., & Smiljanić, R. (2019). Influence of speaking style adaptations and semantic context on the time course of word recognition in quiet and in noise. *Journal of Phonetics, 73*, 158-177.

Van Engen, K. J., & Peelle, J. E. (2014). Listening effort and accented speech. Frontiers in Human Neuroscience, 8. 577. doi: 10.3389/fnhum.2014.00577

Van Engen, K.J. (2010). Similarity and familiarity: second language sentence recognition in first- and second-language multi-talker babble. *Speech Communication, 52*(11-12), 943-953.

Van Noorden, L. (1975). Temporal coherence in the perception of tone sequences. Technische Hogeschool Eindhoven. https://doi.org/10.6100/IR152538

Vandergrift, L., & Goh, C.C. (2012). *Teaching and learning second language listening: Metacognition in action*. Milton Park: Routledge.

Vandergrift (2002). Listening: Theory and practice in modern foreign language competence. Retrieved from https://www.llas.ac.uk/resources/gpg/67

Vicari, S., Marotta, L., & Carlesimo, G. A. (2004). Verbal short-term memory in Down's syndrome: An articulatory loop deficit? *Journal of Intellectual Disability Research, 48*(2), 80-92.

Vismann, C. (2011). Breaks in language at the Nuremberg trials. Retrieved from https://law.unimelb.edu.au/__data/assets/pdf_file/0007/2088844/cornelia-vismann-on-language-breaks-at-nuremberg.pdf

Veenendaal, N. J., Groen, M. A., & Verhoeven, L. (2014). The role of speech prosody and text reading prosody in children's reading comprehension. *British Journal of Educational Psychology, 84*(4), 521-536.

Wagner, M., & Watson, D. G. (2010). Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes*, *25*(7-9), 905-945.

Wagner, R. K., Muse, A. E., & Tannenbaum, K. R. (eds.). (2007). *Vocabulary acquisition: Implications for reading comprehension*. New York: Guilford Press.

Wang, X., & Xu, L. (2021). Speech perception in noise: Masking and unmasking. *Journal of Otology, 16*(2), 109-119.

Walsh, M., Möbius, B., Wade, T., & Schütze, H. (2010). Multilevel exemplar theory. *Cognitive Science*, *34*(4), 537-582.

Warren, R. M., & Obusek, C. J. (1971). Speech perception and phonemic restorations. *Perception & Psychophysics, 9*(3), 358-362.

Wickelgren, W. A. (1965). Acoustic similarity and intrusion errors in short-term memory. *Journal of Experimental Psychology, 70*(1), 102-108.

Wightman, F., Kistler, D., & Brungart, D. (2006). Informational masking of speech in children: Auditory-visual integration. *The Journal of the Acoustical Society of America, 119*(6), 3940-3949.

Worthington, D. L., & Bodie, G. D. (2018). Defining listening: A historical, theoretical, and pragmatic assessment. In G. D. Bodie & D. L. Worthington (eds.), *The sourcebook of listening research: Methodology and measures*, 3-20. Hoboken: John Wiley & Sons.

Yang, X., Wang, Y., Zhang, R., & Zhang, Y. (2021). Physical and psychoacoustic characteristics of typical noise on construction site: "How does noise impact construction workers' experience?". *Frontiers in Psychology, 12*, 1-13. doi:10.3389/fpsyg.2021.707868

Yang, X., Jiang, M., & Zhao, Y. (2017). Effects of noise on English listening comprehension among Chinese college students with different learning styles. *Frontiers in Psychology, 8*. doi:10.3389/fpsyg.2017.01764

Yenkimaleki, M., & Heuven, V. J. (2016). The effect of prosody teaching on developing word recognition skills for interpreter trainees: An experimental study. *Journal of Advances in Linguistics, 7*(1), 1101-1107.

Yudes, C., Macizo, P., & Bajo, T. (2012). Coordinating comprehension and production in simultaneous interpreters: Evidence from the articulatory suppression effect. *Bilingualism: Language and Cognition, 15*(2), 329-339.

Zajdler, E. (2020). Speech shadowing as a teaching technique in the CFL classroom. *Lingua Posnaniensis, 62*(1), 77-88.

Zola D. (1981). The effect of redundancy on the perception of words in reading. *Center or the Study of Reading Technical Report no. 216*.

Appendix 1: Transcripts of the Stimuli Used in the Study

1. Listening Span Task

1
Certain human activities also have a negative impact on agriculture. Urban development also takes its toll as trees are cleared to make way for houses. Deforestation is one of the main causes of soil degradation in the world today. It seems that housing our growing population comes at the cost of providing much needed food. So it is not surprising that farmers are turning to genetically modified crops to try to boost productivity and grow crops in more ecologically healthy fields, while allowing more efficient use of resources.

2
Climate change also produces more extreme weather patterns. These can range from long stretches of drought, and also, conversely, extreme heavy rain, which can cause floods. The destruction of food crops can result from both a lack or a surfeit water. Certain human activities also have a negative impact on agriculture. The use of heavy machinery like tractors can compact the soil.

3
In 2001, we focused on monkeys and their capability to either fashion crude tools or take advantage of naturally occurring ones. Then in 2007, we turned our thoughts to higher-thinking and, in particular, numeracy. We conducted a significant piece of research to find out whether birds are in fact able to count. The findings amazed everyone and caused quite a stir around the world. This helped to spur us on and allowed us to expand the department, making it the world-class facility it is today.

4
I was recently in charge of a government-funded study looking into the impact that prison sentences have on criminals. For our study, we found 96 pairs of convicted burglars and 406 pairs of offenders who had been charged with assault. One member of each pair had been given a prison sentence for their crimes, while the other had received some form of non-custodial penalty. The findings of our study were interesting. Our research team found that offenders who were given a prison sentence were slightly more likely to re-offend than those who did not go to jail.

5
Various parts of New York have changed radically in their ethnic make-up over the last 200 years. Communities became wealthier, governments introduced new laws, and employment opportunities came and went. These factors affect where people choose to live or force them to move somewhere different. For example, most people think that the population has changed in Manhattan due to the rise of its importance as a financial trade center, which is true to some extent. Brooklyn is an interesting example too, and we'll be looking at it as our case study later in the lecture.

6
We used to think that not having a degree would condemn you to a job in the service sector. But now, the job market is extremely competitive and trainees are finding that it is the qualifications they gain thought technical courses rather than degree courses that can help make them employable. The fact is that nowadays there are plenty of jobs that offer a living wage and that don't require a degree. Some of these occupations are familiar. For example, a carpenter, creating things for the home.

7
Over the years, researchers have put forward other types of intelligence to add to this list, but these are usually ignored as they tend to be rather complex and less easily defined. So, how can we use this information in education? Well, these intelligences basically refer to your strengths and weaknesses. Once you have identified these you can build on your strengths by choosing activities that match your intelligence type. For example, a kinesthetic learner is a typical fidgeter who will struggle to learn from a lecture.

2. Listening Comprehension Task

1.
Good morning. In the last few lectures I've been talking about the history of domestic building construction. But today I want to begin looking at some contemporary, experimental designs for housing. So, I'm going to start with the house which is constructed more or less under the ground. And one of the interesting things about this project is that the owners – both professionals but not architects – wanted to be closely involved, so they decided to manage the project themselves. Their chief aim was to create somewhere that was as environmentally-friendly as possible. But at the same time they wanted to lie somewhere peaceful – they'd both grown up in a rural area and disliked urban

life. So, the first thing they did was to look for a site. And they found a disused stone quarry in a beautiful area. The price was relatively low, and they liked the idea of recycling the land, as it were. As it was, the quarry was an ugly blot on the landscape, and it wasn't productive any longer, either.

2

Well, most of you probably know Sports World - the branch of a Danish sports goods company that opened a few years ago - it's attracted a lot of custom, and so the company has now decided to open another branch in the area. It's going to be in the shopping centre to the west of Bradcaster, so that will be good news for all of you who've found the original shop in the north of the town hard to get to. I was invited to a special preview and I can promise you, this is the ultimate in sports retailing. The whole place has been given a new minimalist look with the company's signature colors of black and red. The first three floors have a huge range of sports clothing as well as equipment, and on the top floor there's a cafe and a book and DVD section. You'll find all the well-known names as well as some less well-known ones. If they haven't got exactly what you want in stock they promise to get it for you in ten days, unlike the other store, where it can take up to fourteen days. They cover all the major sports, including football, tennis and swimming, but they particularly focus on running and they claim to have the widest range of equipment in the country.

3

Good afternoon everyone. Well, with some of you about to go out on field work it's timely that in this afternoon's session I'll be sharing some ideas about the reasons why groups of whales and dolphins sometimes swim ashore from the sea right onto the beach and, most often, die in what are known as *mass strandings*. Unfortunately, this type of event is a frequent occurrence in some of the locations that you'll be travelling to, where sometimes the tide goes out suddenly, confusing the animals. However, there are many other theories about the causes of mass strandings. The first is that the behavior is linked to parasites. It's often found that stranded animals were infested with large numbers of parasites. For instance, a type of worm is commonly found in the ears of dead whales. Since marine animals rely heavily on their hearing to navigate, this type of infestation has the potential to be very harmful. Another theory is related to toxins, or poisons. These have also been found to contribute to the death of many marine animals. Many toxins, as I'm sure you're aware, originate from plants, or animals. The whale ingests these toxins in its normal feeding behavior but whether these poisons directly or indirectly lead to stranding and death, seems to depend upon the toxin involved.

4

Good morning. Today I'd like to present the findings of our Year 2 project on wildlife found in gardens throughout our city. I'll start by saying something about the background to the project, then talk a little bit about our research techniques, and then indicate some of our interim findings. First of all, how did we choose our topic? Well, there are four of us in the group and one day while we were discussing a possible focus, two of the group mentioned that they had seen yet more sparrow-hawks - one of Britain's most interesting birds of prey - in their own city center gardens and wondered why they were turning up in these gardens in great numbers. We were all very engaged by the idea of why wild animals would choose to inhabit a city garden. Why is it so popular with wildlife when the countryside itself is becoming less so? The first thing we did was to establish what proportion of the urban land is taken up by private gardens. We estimated that it was about one fifth, and this was endorsed by looking at large-scale usage maps in the town land survey office, 24% to be precise. Our own informal discussions with neighbors and friends led us to believe that many garden owners had interesting experiences to relate regarding wild animal sightings so we decided to survey garden owners from different areas of the city.

5

Firstly, it is important to understand that migration patterns are primarily affected by the rules of immigration which determine the conditions of entry. After that, internal changes can affect patterns considerably. To highlight my first point let’s study this diagram of Ellis Island and the process of admitting immigrants in the late nineteenth and early twentieth centuries. Upon arrival at Ellis Island, people underwent a series of examinations and questions before being allowed to enter the US. First of all, there was a medical inspection to ensure the immigrants were not bringing in any contagious diseases. Anyone who did not pass the medical examination was refused entry to New York and sent home on the next available ship. If the examination was passed, immigrants were required to take a further examination; this time a legal examination to establish whether they had any criminal convictions. After this, immigrants were able to change currency and purchase tickets for onward rail travel from New York. Having completed this simple process, immigrants were told to wait - this wait could be as long as five hours - before boarding a ferry to take them to New York City.

6

The concepts of physics can be very difficult for children to understand, but they can also be really exciting. I'm going to describe three different experiments you can use in the classroom to help show children not only how exciting, but also how useful, physics can be. The first one is based on what's known as the Brazil nut effect. Physicists wondered why large Brazil nuts end up at the top of a jar of mixed nuts. To demonstrate this, you need a jar, a marble and some sand. You put the marble and the sand in the jar and get students to predict what will happen to the marble if they shake the jar. As the marble is denser than the sand, they will make the same assumption as the physicists, that the marble will sink to the bottom, In fact, no matter how much they shake it the marble will remain at or near the top of the sand.

7

I think it’s time for us to divide up parts of the business into smaller units. Therefore, over the next five years, I aim to set up two small subsidiary companies in order to focus on international expansion in Europe and Asia. There are many organizations in emerging markets which could benefit from our experience and skills. Which leads me to the next point for future development – that of increasing our workforce. It’s become clear that all our departments are understaffed, so we’ll be taking on more employees over the next year. And the really good news is that to make us a desirable employer, all positions. Current and future will receive a salary increase of ten per cent. Lastly, I know that some people are worried about the financial aspects of having to move to another city as part of the restructure, so Benchmark will be providing a relocation package to all employees thus affected. This is because we would like you all to remain with the company for the foreseeable future.

3. Speech Shadowing[12]

1

This semester, we're going to be looking at the modern aviation industry here in the USA. But today I'd like to take a look at how it all began. When Orville and Wilbur Wright flew history's first airplane in North Carolina in 1903, the significance of their new invention was of course not yet apparent. Twenty years later, by 1923 the first passenger planes did little to change that. The first of these were provided by some of the airmail services flying mail around the country. The US Post Office Department added a few seats for extra revenue, but their planes

[12] The first seven are the transcripts of audio files used in the shadowing task with human participants, while all nine were used to test the automatic speech recognition programs

were noisy, cold and uncomfortable. They couldn't fly over mountains, so passengers took trains for part of their journey. Another problem was that these planes couldn't carry enough seats to make passenger traffic profitable. So the train was still the way to go. In 1927, Charles Lindbergh's transatlantic flight captured America's imagination. Lindbergh flew in a small airplane for 33 hours from New York to Paris.

2

One of the main factors in ensuring a harmonious society is that there are clear established patterns in the way we conduct ourselves. And we expect people to behave according to our accepted standards of behavior. There are those who observe these social mores religiously, and these people are often labeled *conservative*. It's actually through such people that our heritage is preserved, but then, gradually over time, as our society becomes more and more multicultural, there is a blending of these customs and we gradually come to redefine the norm. If we enter a new group, we notice the unwritten rules and social norms of that group. Those who fail to observe these norms are inevitably excluded from that group. Of course, there will always be those who seek to break away from tradition, and to rebel. These people see themselves as unconventional in every sense of the word.

3

The next experiment is called the arm engine and for this one you need a chair that can swivel or rotate and some small hand weights. This is a great experiment for demonstrating an important principle of energy and momentum. Ask one of your students to sit on the chair holding the weights in their hands. Then get another student to spin the chair as fast as they can. Thanks to the weights, the student sitting in the chair will be able to control their own speed. If they hold the weights out, they will slow down and if they hold them close to their body, making themselves narrower, they will accelerate the speed of their rotation. We can observe this principle in the real world in the sport of ice skating, where the skaters manage to spin incredibly fast by tucking their hands in close to their body.

4

The second experiment is always fun as it involves a balloon! You also need a pin and some sticky tape. First you inflate the balloon and then you put the sticky tape on it, but don't tell the students you've done this. Now ask the students what makes a balloon burst. Most people assume balloons make a loud bang when the air is released through the hole. However, if you pierce the balloon through the

sticky tape, instead of bursting it, the air will leak out quietly and slowly. So it can't be the air escaping chat causes the noise. Instead, physics has shown us the loud bang occurs because the hole expands rapidly forming a catastrophic crack. You can also tell your students, when the balloon does burst open, it does so faster than the speed of sound, so the loud bang you hear is actually a sonic boom! In the real world, this principle is used to test different materials to see how elastic they are and how much stress can be put on them.

5

Today I want to run through the fire evacuation procedure now that we're in a new building. First of all, can I just remind you that if you hear the fire alarm, you should always head towards the main stairs in order to leave the building. Please assume that the alarm is real, except if it sounds at 11.00 a.m. on a Tuesday. At this time, it's always a test - we hope. It's vital that you do not spend time collecting your bags or personal belongings because this wastes valuable evacuation time. When you have left the building, please look for the fire marshals, who will be wearing fluorescent orange jackets. They'll show you where the waiting area is, but just so you know, it's the park at the rear of the office block. Your department has a fire safety officer, I believe it's Susan Jenkins, and it's her job to make sure that everyone who signed in has vacated the building.

6

This afternoon I want to go over the new marketing and advertising strategy so that everyone is clear on the streams for each of our product ranges. Let's start with toys for children. Now, last year most of the advertising was done through leaflets posted through people's letterboxes across the city. However, the products are now selling well nationally in department stores, rather than just in our local shop here in Leeds, so we're going to expand the budget and use print media. By this I mean the national newspapers, in order to maximize the exposure to these products. And despite the fact that our competitors advertise baby clothes on TV, we won't be using this method, as our statistics show that it's just not cost-effective. People don't pay much attention to TV ads for baby clothes. But we believe a picture in the newspapers will be much more attractive to potential customers. We're going with this method.

7

In the past, people were seen as either intelligent of unintelligent, and this was measured with an IQ test. However psychologists now recognize that there are many different types of intelligence and these are reflected in your personality.

The multiple intelligence theory first came to light in 1983 in Howard Gardner's book Frames of Mind. In it, Gardner listed seven types of intelligence. The first of these is termed linguistic and this describes people who are more interested in the written word and reading. The next kind of intelligence is logical and this is used to describe people whose strengths are in subjects such as maths and science. Then there is musical intelligence, followed by kinesthetic, which relates to the body and movement. After that there is visual intelligence, which describes people who are attracted by or drawn to images. And then the final two intelligences are interpersonal and intrapersonal.

8

Today I'd like to continue from last week's lecture by looking at what helps people successfully integrate into a new culture. Whereas the reasons for migration are nowadays fairly easy to identify and largely related to employment opportunities or political instability, the factors behind being able to adapt to the new culture and create a new life are considerably more complex. Let's start with an overview of the issues as shown on this diagram. You might think that the list of negative factors would include discrimination, but it doesn't because discrimination comes under the larger category of fear. Now, what you should also notice is that the external factors are not labeled in this way. It's much more difficult to know how to measure the affects of external factors and whether they actually are external or not. The influence of family relationships, climate, beliefs and values, and the ability to communicate in the language of the new culture have wide-ranging effects which are difficult to measure and can distort any research.

9

I think it's time for us to divide up parts of the business into smaller units. Therefore, over the next five years, I aim to set up two small subsidiary companies in order to focus on international expansion in Europe and Asia. There are many organizations in emerging markets which could benefit from our experience and skills. Which leads me to the next point for future development – that of increasing our workforce. It's become clear that all our departments are understaffed, so we'll be taking on more employees over the next year. And the really good news is that to make us a desirable employer, all positions, current and future will receive a salary increase of ten per cent. Lastly, I know that some people are worried about the financial aspects of having to move to another city as part of the restructure, so Benchmark will be providing a relocation package to all employees thus affected. This is because we would like you all to remain with the company for the foreseeable future.

Appendix 2: Examples of Raw Transcripts (Otter and Ava)

Following are the direct transcripts of one speech recording played in the clean condition as well as through all the experimental conditions, as transcribed by Otter. No spelling or punctuation has been altered. The transcripts are ordered based on the percentage of errors found in them (Table 7.1) progressing from the most accurate one (clean) to the most inaccurate one (single babble informational).

Table A.2.1. Full, unaltered transcripts of one speech recording in the clean condition, followed by all the experimental conditions, as transcribed by Otter.

Original
The next experiment is called the arm engine and for this one you need a chair that can swivel or rotate and some small hand weights. This is a great experiment for demonstrating an important principle of energy and momentum. Ask one of your students to sit on the chair holding the weights in their hands. Then get another student to spin the chair as fast as they can. Thanks to the weights, the student sitting in the chair will be able to control their own speed. If they hold the weights out, they will slow down and if they hold them close to their body, making themselves narrower, they will accelerate the speed of their rotation. We can observe this principle in the real world in the sport of ice skating, where the skaters manage to spin incredibly fast by tucking their hands in close to their body.
Otter: Clean
The next experiment is called the arm engine, and for this one you need a chair that can swivel or rotate, and some small hand weights. This is a great experiment for demonstrating an important principle of energy and momentum. Ask one of your students to sit on the chair holding the weights in their hands. Then get another student to spin the chair as fast as they can. Thanks to the weights, the students sitting in the chair will be able to control their own speed. If they hold the weights out they will slow down, and if they hold them close to their body, making themselves narrower, they will accelerate the speed of their rotation. We can observe this principle in the real world, in the sport of ice skating, where the skaters managed to spin incredibly fast by tucking their hands in close to their body.
Otter: Reverb
The next experiment is called the arm engine, and for this one you need a chair that can swivel or rotate, and some small hand weights. This is a great experiment for demonstrating an important principle of energy and momentum. Ask one of your students to sit on the chair holding the weights in their hands. Then get another student to spin the chair as fast as they can. Thanks to the weights, the students sitting in the chair will be able to control their own speed. If they hold the weights out they will slow down, and if they hold them close to their body, making themselves narrower, they will accelerate the speed of their rotation. We can observe this principle in the real world, in the sport of ice skating, where the skaters managed to spin incredibly fast by tucking their hands in close to their body.
Otter: Multi Babble Energetic
The next experiment is called the arm engine, and for this one you need a chair that can swivel or

rotate, and some small hand weights. This is a great experiment for demonstrating an important principle of energy and momentum. Also one of your students to sit on the chair holding the weights in their hands. Then get another student to spin the chair as fast as they can. Thanks to the waste, the students sitting in the chair will be able to control their own speed. If they hold the weights out they will slow down, and if they hold them first to their body, making themselves narrower, will accelerate the speed of their rotation. We can observe this principle in the real world, in the sport of ice skating, where the skaters managed to spin incredibly fast by topping their hands in close to their bodies.
Otter: Phone
Experiment is called the arm engine and and swivel or rotate small hand weights. Experiment, and in important principle and momentum. Ask one of your students to see on the chair holding the weights in that in the chair as fast as they can to the weights. The students sitting in the chair. we'll be able to control their own speed. If they hold the weights out, will slow down, and if they hold them close to their body, making themselves narrower, accelerate the speed of rotation. We can observe this principle in the real world. in the sport of ice skating has managed to spin incredible fast by tucking their close to their body.
Otter: Single Babble Energetic
next experiment is the arm engine voiceover and for this one you need a chair they rotate and some small hand. This is a great experiment for demonstrating an important principle of energy and momentum and in chat. We asked one of your strengths to sit on your chair, holding the weights and then get another student in the chat as fast as they can. Thanks to the way we're sitting in the chair will be able to control their way to slow down and close to their body may find themselves narrower, they will accelerate the speed of their rotation, I mean balancing chat, we can observe this principle in the real world, it was you know where the skaters managed to spin and wait for it to be false by talking their hands in most of their body on
Otter: Construction
The arm engine. For this one you need a chair that can swivel or rotate, and some small hand weights demonstrates the energy and momentum. One of your students to sit on the chair, holding the waiting as fast as they can. Thanks to the weight student sitting in the chair, will be able control the way to out, Slave down. Let me think all them close to their body, making themselves narrower. We can observe this principle in the real world, in the sport of ice skating, where the skaters managed to spin incredibly fast by top. That
Otter: Single Babble Informational
continuing next experiment is Elijah, and for this one you need a Chad and swiveled or rotate and some small handle a forecast track the latest is a great experiment for demonstrating the principle of your energy and momentum early tomorrow. Last one of your students to sit on the chat and holding away. It is extend get another student but just a reason the chat as fast as they can over the gold thanks to the way to the student sitting in the chair will be able to control their own speed going to be conducive, they hold the way vacation, slow down and if they hold them and then They will accelerate the speed rotation real land for, we can observe this principle in the real world, in this music, it takes us where the stage is managed just been incredibly fast away talking their hands and we're close to their body exchange balls

Following are the direct transcripts of one speech recording played in the clean condition as well as through all the experimental conditions, as transcribed by Ava. No spelling or punctuation has been altered. The transcripts are ordered based on the percentage of errors found in them (Table 7.9) progressing from the most accurate one (clean) to the most inaccurate one (construction).

Table A.2.2. Full, unaltered transcripts of one speech recording in the clean condition, followed by all the experimental conditions, as transcribed by Ava.

Original
The next experiment is called the arm engine and for this one you need a chair that can swivel or rotate and some small hand weights. This is a great experiment for demonstrating an important principle of energy and momentum. Ask one of your students to sit on the chair holding the weights in their hands. Then get another student to spin the chair as fast as they can. Thanks to the weights, the student sitting in the chair will be able to control their own speed. If they hold the weights out, they will slow down and if they hold them close to their body, making themselves narrower, they will accelerate the speed of their rotation. We can observe this principle in the real world in the sport of ice skating, where the skaters manage to spin incredibly fast by tucking their hands in close to their body.
Ava: Clean
The next experiment is called the arm engine. And for this one you need a chair that can swivel or rotate and some small hand weights. This is a great experiment for demonstrating an important principle of energy and momentum ask one of your students to sit on the chair holding the weights in their hands. Then get another student to spin the chair as fast as. As they can thanks to the weights, the students sitting in the chair will be able to control their own speed. If they hold the weights out. They will slow down. And if they hold them close to their body making themselves narrower, they will accelerate the speed of their rotation. We can observe this principle in the real world in the sport of ice skating. Where the skaters managed to spin incredibly fast by tucking their hands in close to their body?
Ava: Phone
The next experiment is called the arm. Ian and for this one, you need a chair that can swivel or rotate and some small hand weights. This is a great experiment for demonstrating an important principle of energy and momentum ask one of your students to sit on the chair holding the weights in their hands. Then get another student to spin the chair as fast as they can. Thanks to the weights, the student sitting in the chair, will be able to control their own speed, if they hold the way. Out, they will slow down and if they hold them close to their body making themselves narrower, they will accelerate the speed of their rotation. We can observe this principle in the real world in the sport of ice skating. Where the skaters managed to spin incredibly fast by tucking their hands in close to their body.
Ava: Reverb
The next experiment is called the arm engine. And for this one you need a chair that can swivel or rotate and some small hand weights. This is a great experiment for demonstrating an important principle of energy. Ask one of your students to sit on the chair holding the weights in their hands. Then get another student to spin the chair as fast as they can. Thanks to the weights, the student sitting in the chair, will be able to control their own speed, if they hold the weights out. Will slow down and if they hold them close to their body. Of narrower, they will accelerate the speed of their rotation. We can observe this principle in the real world in the sport of ice skating. Where the skaters managed to spin incredibly fast by tucking their hands in close to their body.
Ava: Multi Babble Energetic
The next experiment is called the arm engine. And for this one, you need a chair that can swivel. Will rotate and some small hand weights. This is a great experiment for demonstrating an important principle of energy and momentum ask one of your students to sit on the chair holding the weights in their hands. Then get another student to spin the chair as fast as they can. Thanks to the waves, the students sitting in the chair will be able to control their own speed. If they hold the weight out. They will slow down and if they hold them close, Body making themselves narrower, they will accelerate the speed of their rotation. We can observe this principle in the real world in the sport of ice skating. Where the skaters managed to spin incredibly fast by talking their hands in close to their body.

Ava: Single Babble Energetic
the next And that is called the arm engine whistle. And for this one, you need a chair that can swivel or rotate. And some small hand weights. This is a great experiment for demonstrating an important principle of energy and momentum ask one of your students to sit on the chair holding the weights in their hands. Then get another student to spin the chair as fast as they can. Thanks to the weights, the students sitting in the chair will be able to control their own speed. If they wait, they will slow down and if they hold them close to their body making themselves narrower, they will accelerate the speed of their rotation. We can observe this principle in the real world in the city water ice skating. Where the skaters managed to spin in with relatively fast by tucking their hands in close to their body.
Ava: Single Babble Informational
The next experiment is called the arm Alger. And for this one you need a chair and swiveled or rotate. And some small hand weights. This is a great experiment for demonstrating an important principle of energy and momentum ask one of your students to sit on the chair, holding the weights in their hands. Then get another student. In the chat as wants to stake a the the gold thanks to the weights in the students sitting in the chair, will be able to control their own speed. Going to be conducive if they hold the way, you know, vacation. They will slow down and if they hold them and their body and making themselves narrower, they will accelerate the speed of their rotation real. And for, we can observe this principle in the real world in the sport of ice and takes the, where the skaters managed to spend incredibly fast by talking their hands in close to their body extreme.
Ava: Construction
The next experiment is called the farm. And for this one you need a chair and swivel will rotate and some small hand weights. This is a great experiment. Where the skaters managed to spin incredibly fast by talking their hands and close to their body.

Appendix 3: Visual Comparison between Original Text and Transcripts

The following two tables (A 3.1 and A 3.2) offer a visual comparison between the original text and transcripts provided by Otter and Ava – the left column presenting the original text, while the right one the transcript. The same speech recording showed in the previous section is presented here. The online tool Text Compare was used to compare the two texts and generate the images. While the webpage does not offer details about what elements their algorithm is programmed to capture, it can be observed from the transcripts that both wording and punctuation, but also spacing were subject of the analysis. The differences in blue shading are marked on both texts that are being compared.

Table A 3.1. Visual comparison by Text Compare tool between the original text and Otter transcripts throughout the conditions.

Otter: Clean **original**	**transcript**
The next experiment is called the arm engine and for this one you need a chair that can swivel or rotate and some small hand weights. This is a great experiment for demonstrating an important principle of energy and momentum. Ask one of your students to sit on the chair holding the weights in their hands. Then get another student to spin the chair as fast as they can. Thanks to the weights, the student sitting in the chair will be able to control their own speed. If they hold the weights out, they will slow down and if they hold them close to their body, making themselves narrower, they will accelerate the speed of their rotation. We can observe this principle in the real world in the sport of ice skating, where the skaters manage to spin incredibly fast by tucking their hands in close to their body.	The next experiment is called the arm engine, and for this one you need a chair that can swivel or rotate, and some small hand weights. This is a great experiment for demonstrating an important principle of energy and momentum. Ask one of your students to sit on the chair holding the weights in their hands. Then get another student to spin the chair as fast as they can. Thanks to the weights, the students sitting in the chair will be able to control their own speed. If they hold the weights out they will slow down, and if they hold them close to their body, making themselves narrower, they will accelerate the speed of their rotation. We can observe this principle in the real world, in the sport of ice skating, where the skaters managed to spin incredibly fast by tucking their hands in close to their body.
Otter: Reverb **original**	**transcript**
The next experiment is called the arm engine and for this one you need a chair that can swivel or rotate and some small hand weights. This is a great experiment for demonstrating an important principle of energy and momentum. Ask one of your students to sit on the chair holding the weights in their hands. Then get another student to spin the chair as fast as they can. Thanks to the weights, the student sitting in the chair will be able to control their own speed. If they hold the weights out, they will slow down and if they hold them close to their body, making themselves narrower, they will accelerate the speed of their rotation. We can observe this principle in the real world in the sport of ice skating, where the skaters manage to spin incredibly fast by tucking their hands in close to their body.	The next experiment is called the arm engine, and for this one you need a chair that can swivel or rotate, and some small hand weights. This is a great experiment for demonstrating an important principle of energy and momentum. Ask one of your students to sit on the chair holding the weights in their hands. Then get another student to spin the chair as fast as they can. Thanks to the weights, the students sitting in the chair will be able to control their own speed. If they hold the weights out they will slow down, and if they hold them close to their body, making themselves narrower, they will accelerate the speed of their rotation. We can observe this principle in the real world, in the sport of ice skating, where the skaters managed to spin incredibly fast by tucking their hands in close to their body.
Otter: Multi Babble Energetic **original**	**transcript**

The next experiment is called the arm engine and for this one you need a chair that can swivel or rotate and some small hand weights. This is a great experiment for demonstrating an important principle of energy and momentum. Ask one of your students to sit on the chair holding the weights in their hands. Then get another student to spin the chair as fast as they can. Thanks to the weights, the student sitting in the chair will be able to control their own speed. If they hold the weights out, they will slow down and if they hold them close to their body, making themselves narrower, they will accelerate the speed of their rotation. We can observe this principle in the real world in the sport of ice skating, where the skaters manage to spin incredibly fast by tucking their hands in close to their body.	The next experiment is called the arm engine, and for this one you need a chair that can swivel or rotate, and some small hand weights. This is a great experiment for demonstrating an important principle of energy and momentum. Also one of your students to sit on the chair holding the weights in their hands. Then get another student to spin the chair as fast as they can. Thanks to the waste, the students sitting in the chair will be able to control their own speed. If they hold the weights out they will slow down, and if they hold them first to their body, making themselves narrower, will accelerate the speed of their rotation. We can observe this principle in the real world, in the sport of ice skating, where the skaters managed to spin incredibly fast by topping their hands in close to their bodies.
Otter: Phone **original**	**transcript**
The next experiment is called the arm engine and for this one you need a chair that can swivel or rotate and some small hand weights. This is a great experiment for demonstrating an important principle of energy and momentum. Ask one of your students to sit on the chair holding the weights in their hands. Then get another student to spin the chair as fast as they can. Thanks to the weights, the student sitting in the chair will be able to control their own speed. If they hold the weights out, they will slow down and if they hold them close to their body, making themselves narrower, they will accelerate the speed of their rotation. We can observe this principle in the real world in the sport of ice skating, where the skaters manage to spin incredibly fast by tucking their hands in close to their body.	Experiment is called the arm engine and and swivel or rotate small hand weights. Experiment, and in important principle and momentum. Ask one of your students to see on the chair holding the weights in that in the chair as fast as they can to the weights. The students sitting in the chair. we'll be able to control their own speed. If they hold the weights out, will slow down, and if they hold them close to their body, making themselves narrower, accelerate the speed of rotation. We can observe this principle in the real world. in the sport of ice skating has managed to spin incredible fast by tucking their close to their body.
Otter: Single Babble Energetic **original**	**transcript**
The next experiment is called the arm engine and for this one you need a chair that can swivel or rotate and some small hand weights. This is a great experiment for demonstrating an important principle of energy and momentum. Ask one of your students to sit on the chair holding the weights in their hands. Then get another student to spin the chair as fast as they can. Thanks to the weights, the student sitting in the chair will be able to control their own speed. If they hold the weights out, they will slow down and if they hold them close to their body, making themselves narrower, they will accelerate the speed of their rotation. We can observe this principle in the real world in the sport of ice skating, where the skaters manage to spin incredibly fast by tucking their hands in close to their body.	next experiment is the arm engine voiceover and for this one you need a chair they rotate and some small hand. This is a great experiment for demonstrating an important principle of energy and momentum and in chat. We asked one of your strengths to sit on your chair, holding the weights and then get another student in the chat as fast as they can. Thanks to the way we're sitting in the chair will be able to control their way to slow down and close to their body may find themselves narrower, they will accelerate the speed of their rotation, I mean balancing chat, we can observe this principle in the real world, it was you know where the skaters managed to spin and wait for it to be false by talking their hands in most of their body on
Otter: Construction **original**	**transcript**
The next experiment is called the arm engine and for this one you need a chair that can swivel or rotate and some small hand weights. This is a great experiment for demonstrating an important principle of energy and momentum. Ask one of your students to sit on the chair holding the weights in their hands. Then get another student to spin the chair as fast as they can. Thanks to the weights, the student sitting in the chair will be able to control their own speed. If they hold the weights out, they will slow down and if they hold them close to their body, making themselves narrower, they will accelerate the speed of their rotation. We can observe this principle in the real world in the sport of ice skating, where the skaters manage to spin incredibly fast by tucking their hands in close to their body.	The arm engine. For this one you need a chair that can swivel or rotate, and some small hand weights demonstrates the energy and momentum. One of your students to sit on the chair, holding the waiting as fast as they can. Thanks to the weight student sitting in the chair, will be able control the way to out, Slave down. Let me think all them close to their body, making themselves narrower. We can observe this principle in the real world, in the sport of ice skating, where the skaters managed to spin incredibly fast by top. That
Otter: Single Babble Informational original	**transcript**
The next experiment is called the arm engine and for this one you need a chair that can swivel or rotate and some small hand weights. This is a great experiment for demonstrating an important principle of energy and momentum. Ask one of your students to sit on the chair holding the weights in their hands. Then get another student to spin the chair as fast as they can. Thanks to the weights, the student sitting in the chair will be able to control their own speed. If they hold the weights out, they will slow down and if they hold them close to their body, making themselves narrower, they will accelerate the speed of their rotation. We can observe this principle in the real world in the sport of ice skating, where the skaters manage to spin incredibly fast by tucking their hands in close to their body.	continuing next experiment is Elijah, and for this one you need a Chad and swiveled or rotate and some small handle a forecast track the latest is a great experiment for demonstrating the principle of your energy and momentum early tomorrow. Last one of your students to sit on the chat and holding away. It is extend get another student but just a reason the chat as fast as they can over the gold thanks to the way to the student sitting in the chair will be able to control their own speed going to be conducive, they hold the way vacation, slow down and if they hold them and then They will accelerate the speed rotation real land for, we can observe this principle in the real world, in this music, it takes us where the stage is managed just been incredibly fast away talking their hands and we're close to their body exchange balls

Table A 3.2. Visual comparison by Text Compare tool between the original text and Ava transcripts throughout the conditions.

Ava: Clean	**original**	**transcript**
	The next experiment is called the arm engine and for this one you need a chair that can swivel or rotate and some small hand weights. This is a great experiment for demonstrating an important principle of energy and momentum. Ask one of your students to sit on the chair holding the weights in their hands. Then get another student to spin the chair as fast as they can. Thanks to the weights, the student sitting in the chair will be able to control their own speed. If they hold the weights out, they will slow down and if they hold them close to their body, making themselves narrower, they will accelerate the speed of their rotation. We can observe this principle in the real world in the sport of ice skating, where the skaters manage to spin incredibly fast by tucking their hands in close to their body.	The next experiment is called the arm engine. And for this one you need a chair that can swivel or rotate and some small hand weights. This is a great experiment for demonstrating an important principle of energy and momentum ask one of your students to sit on the chair holding the weights in their hands. Then get another student to spin the chair as fast as. As they can thanks to the weights, the students sitting in the chair will be able to control their own speed. If they hold the weights out. They will slow down. And if they hold them close to their body making themselves narrower, they will accelerate the speed of their rotation. We can observe this principle in the real world in the sport of ice skating. Where the skaters managed to spin incredibly fast by tucking their hands in close to their body?
Ava: Phone	**original**	**transcript**
	The next experiment is called the arm engine and for this one you need a chair that can swivel or rotate and some small hand weights. This is a great experiment for demonstrating an important principle of energy and momentum. Ask one of your students to sit on the chair holding the weights in their hands. Then get another student to spin the chair as fast as they can. Thanks to the weights, the student sitting in the chair will be able to control their own speed. If they hold the weights out, they will slow down and if they hold them close to their body, making themselves narrower, they will accelerate the speed of their rotation. We can observe this principle in the real world in the sport of ice skating, where the skaters manage to spin incredibly fast by tucking their hands in close to their body.	The next experiment is called the arm. Ian and for this one, you need a chair that can swivel or rotate and some small hand weights. This is a great experiment for demonstrating an important principle of energy and momentum ask one of your students to sit on the chair holding the weights in their hands. Then get another student to spin the chair as fast as they can. Thanks to the weights, the student sitting in the chair, will be able to control their own speed, if they hold the way. Out, they will slow down and if they hold them close to their body making themselves narrower, they will accelerate the speed of their rotation. We can observe this principle in the real world in the sport of ice skating. Where the skaters managed to spin incredibly fast by tucking their hands in close to their body.
Ava: Reverb	**original**	**transcript**
	The next experiment is called the arm engine and for this one you need a chair that can swivel or rotate and some small hand weights. This is a great experiment for demonstrating an important principle of energy and momentum. Ask one of your students to sit on the chair holding the weights in their hands. Then get another student to spin the chair as fast as they can. Thanks to the weights, the student sitting in the chair will be able to control their own speed. If they hold the weights out, they will slow down and if they hold them close to their body, making themselves narrower, they will accelerate the speed of their rotation. We can observe this principle in the real world in the sport of ice skating, where the skaters manage to spin incredibly fast by tucking their hands in close to their body.	The next experiment is called the arm engine. And for this one you need a chair that can swivel or rotate and some small hand weights. This is a great experiment for demonstrating an important principle of energy. Ask one of your students to sit on the chair holding the weights in their hands. Then get another student to spin the chair as fast as they can. Thanks to the weights, the student sitting in the chair, will be able to control their own speed, if they hold the weights out. Will slow down and if they hold them close to their body. Of narrower, they will accelerate the speed of their rotation. We can observe this principle in the real world in the sport of ice skating. Where the skaters managed to spin incredibly fast by tucking their hands in close to their body.
Ava: Multi Babble Energetic	**original**	**transcript**
	The next experiment is called the arm engine and for this one you need a chair that can swivel or rotate and some small hand weights. This is a great experiment for demonstrating an important principle of energy and momentum. Ask one of your students to sit on the chair holding the weights in their hands. Then get another student to spin the chair as fast as they can. Thanks to the weights, the student sitting in the chair will be able to control their own speed. If they hold the weights out, they will slow down and if they hold them close to their body, making themselves narrower, they will accelerate the speed of their rotation. We can observe this principle in the real world in the sport of ice skating, where the skaters manage to spin incredibly fast by tucking their hands in close to their body.	The next experiment is called the arm engine. And for this one, you need a chair that can swivel. Will rotate and some small hand weights. This is a great experiment for demonstrating an important principle of energy and momentum ask one of your students to sit on the chair holding the weights in their hands. Then get another student to spin the chair as fast as they can. Thanks to the waves, the students sitting in the chair will be able to control their own speed. If they hold the weight out. They will slow down and if they hold them close, Body making themselves narrower, they will accelerate the speed of their rotation. We can observe this principle in the real world in the sport of ice skating. Where the skaters managed to spin incredibly fast by talking their hands in close to their body.
Ava: Single Babble Energetic	**original**	**transcript**

The next experiment is called the arm engine and for this one you need a chair that can swivel or rotate and some small hand weights. This is a great experiment for demonstrating an important principle of energy and momentum. Ask one of your students to sit on the chair holding the weights in their hands. Then get another student to spin the chair as fast as they can. Thanks to the weights, the student sitting in the chair will be able to control their own speed. If they hold the weights out, they will slow down and if they hold them close to their body, making themselves narrower, they will accelerate the speed of their rotation. We can observe this principle in the real world in the sport of ice skating, where the skaters manage to spin incredibly fast by tucking their hands in close to their body.	the next And that is called the arm engine whistle. And for this one, you need a chair that can swivel or rotate. And some small hand weights. This is a great experiment for demonstrating an important principle of energy and momentum ask one of your students to sit on the chair holding the weights in their hands. Then get another student to spin the chair as fast as they can. Thanks to the weights, the students sitting in the chair will be able to control their own speed. If they wait, they will slow down and if they hold them close to their body making themselves narrower, they will accelerate the speed of their rotation. We can observe this principle in the real world in the city water ice skating. Where the skaters managed to spin in with relatively fast by tucking their hands in close to their body.
Ava: Single Babble Informational original	**transcript**
The next experiment is called the arm engine and for this one you need a chair that can swivel or rotate and some small hand weights. This is a great experiment for demonstrating an important principle of energy and momentum. Ask one of your students to sit on the chair holding the weights in their hands. Then get another student to spin the chair as fast as they can. Thanks to the weights, the student sitting in the chair will be able to control their own speed. If they hold the weights out, they will slow down and if they hold them close to their body, making themselves narrower, they will accelerate the speed of their rotation. We can observe this principle in the real world in the sport of ice skating, where the skaters manage to spin incredibly fast by tucking their hands in close to their body.	The next experiment is called the arm Alger. And for this one you need a chair and swiveled or rotate. And some small hand weights. This is a great experiment for demonstrating an important principle of energy and momentum ask one of your students to sit on the chair, holding the weights in their hands. Then get another student. In the chat as wants to stake a the the gold thanks to the weights in the students sitting in the chair, will be able to control their own speed. Going to be conducive if they hold the way, you know, vacation. They will slow down and if they hold them and their body and making themselves narrower, they will accelerate the speed of their rotation real. And for, we can observe this principle in the real world in the sport of ice and takes the, where the skaters managed to spend incredibly fast by talking their hands in close to their body extreme.
Ava: Construction original	**transcript**
The next experiment is called the arm engine and for this one you need a chair that can swivel or rotate and some small hand weights. This is a great experiment for demonstrating an important principle of energy and momentum. for demonstrating an important principle of energy and momentum. Ask one of your students to sit on the chair holding the weights in their hands. Then get another student to spin the chair as fast as they can. Thanks to the weights, the student sitting in the chair will be able to control their own speed. If they hold the weights out, they will slow down and if they hold them close to their body, making themselves narrower, they will accelerate the speed of their rotation. We can observe this principle in the real world in the sport of ice skating. where the skaters manage to spin incredibly fast by tucking their hands in close to their body.	The next experiment is called the farm. And for this one you need a chair and swivel will rotate and some small hand weights. This is a great experiment. Where the skaters managed to spin incredibly fast by talking their hands and close to their body.