INVESTIGATING THE ROLE OF VISUAL CUES IN SPEECH PERCEPTION AND SECOND LANGUAGE COMPREHENSION

by

XIAO FENG

(Under the Direction of Linda Harklau)

ABSTRACT

Speech perception is a multimodal process that involves both auditory and visual cues. Visual elements, such as lip and jaw movements, can significantly enhance comprehension, especially for those learning a second language (L2 learners). The McGurk effect, a perceptual illusion where mismatched auditory and visual signals alter speech perception, highlights the important role of visual information in this process. Moreover, visual cues like images have been shown to improve language comprehension in L2 learners, particularly when the visuals are directly related to the content. These findings underscore the significant role of visual cues in both speech perception and language comprehension.

Despite strong evidence supporting the importance of visual cues, individual variations in experiencing the McGurk effect remain unexplained. Factors such as lipreading skills and cognitive functions may contribute to this variability, but further research is needed to fully understand these differences. In addition, a recent study failed to find a positive effect of visual cues on language comprehension. This discrepancy may be attributed to the low effectiveness of the visual cues used in the study or the difficulty of the test materials. A more robust experimental design is necessary to definitively answer this question.

In my dissertation, I designed experiments to investigate the factors that may contribute to the McGurk effect susceptibility. By analyzing the correlation between McGurk effect susceptibility and working memory, as well as between McGurk effect susceptibility and lipreading skills, I found that lipreading skills were weakly correlated with individual differences in McGurk effect susceptibility. In contrast, working memory had no significant impact. Furthermore, by analyzing the effects of visual cues on language comprehension in L2 learners, I found that content-related visuals improved comprehension in language tasks of medium difficulty. This enhancement was observed in L2 learners with intermediate English proficiency, whereas participants with low or high English proficiency levels did not show significant benefits. My study supports the complex role of visual cues in language acquisition for L2 learners.

INDEX WORDS: McGurk effect, speech perception, language comprehension, L2 acquisition, working memory, lipreading, content visual, context visual

INVESTIGATING THE ROLE OF VISUAL CUES IN SPEECH PERCEPTION AND SECOND LANGUAGE COMPREHENSION

by

XIAO FENG

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2024

INVESTIGATING THE ROLE OF VISUAL CUES IN SPEECH PERCEPTION AND SECOND LANGUAGE COMPREHENSION

by

XIAO FENG

Major Professor:	Linda Harklau
Committee:	Liang Chen
	Paula Mellom

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
December 2024

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my advisor, Dr. Linda Harklau. Her unwavering guidance, endless patience, and profound expertise have been the cornerstone of this research. I am truly fortunate to have had her mentorship throughout this journey, and I owe much of this work to her support and encouragement. I am also sincerely thankful to my committee members, Dr. Liang Chen and Dr. Paula Mellom, for their insightful feedback and steadfast support. A special thanks go to the Department of Language and Literacy Education at the University of Georgia for providing me with great support these years and the graduate school for supplying me with the summer research grant that made this study possible.

I am especially grateful to Dr. Strand, Dr. Kamiya, and Karine Dupuis for generously providing their study materials for my dissertation research. I sincerely thank the participants; without their willingness to contribute their time and effort, this research would not have been possible. I also extend my heartfelt appreciation to my friends and colleagues for their constant encouragement and for helping me persevere through the challenges of this academic journey. I am particularly thankful to my loving family, whose unwavering support and encouragement have been my source of strength throughout this journey. Lastly, I sincerely thank my spiritual teachers, whose blessings and spiritual support have carried me through the most difficult times. Thank you for your support and guidance. It is only possible because of you all.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION

Background of the Problem

Speech perception, particularly in second-language (L2) acquisition, is a challenging process that involves the integration of both auditory and visual information. While much of the focus has traditionally been on auditory signals, research increasingly highlights the importance of visual cues, such as lip movements and facial expressions, in aiding speech comprehension. These visual elements help listeners, especially L2 learners, fill in gaps in auditory input and improve their overall understanding of spoken language. However, significant individual variability exists in how people process and benefit from visual cues, particularly in phenomena known as the McGurk effect—a perceptual illusion where conflicting auditory and visual signals lead to altered speech perception.

Understanding this variability is essential for identifying how visual cues can be effectively used to support L2 learners. Previous research has suggested that cognitive functions such as working memory and lipreading capacity may influence how individuals experience the McGurk effect. However, the extent of this relationship remains unclear, raising the first research question: 1) To what extent does individual variability in the McGurk effect depend on working memory and/or lipreading capacity?

Beyond speech perception, visual cues are also believed to aid in language comprehension, especially when learners face more complex listening tasks. Content visuals, such as images directly related to the spoken discourse, may enhance comprehension by providing meaningful

context, while context visuals—visuals that represent the speaker's environment—might be less effective. This leads to the second research question: 2) Do content or context visuals enhance discourse-level listening comprehension in L2 English learners? Lastly, the complexity of the listening task itself could influence the effectiveness of visual cues. While an overly simple listening task might not need the aid from visual cues, a highly difficult task might put a heavy cognitive burden on the participants, forcing them to ignore the visual information and focus on the auditory input. This leads to my final research question: 3) Does the relationship or influence of visual cues, if present, vary depending on the complexity of the listening task?

Purpose of the Study

My dissertation aims to explore the role of visual cues in speech perception and language comprehension among second-language (L2) learners, with a specific focus on the McGurk effect and the impact of content and context visuals on comprehension. This research addresses three key questions: First, it seeks to determine the extent to which individual variability in the McGurk effect is influenced by working memory and/or lipreading capacity. By examining these cognitive factors, this study aims to provide insight into why some individuals experience the McGurk effect more consistently than others.

Second, this study investigates whether content or context visuals enhance discourse-level listening comprehension in L2 English learners. Understanding how different types of visual support affect comprehension is crucial for developing more effective teaching strategies and learning materials for L2 learners.

Lastly, the study examines how the complexity of listening tasks moderates the influence of visual cues. By exploring the interaction between visual cues and task difficulty, this research aims to identify whether visual supports are more effective in certain contexts or for specific

proficiency levels. Ultimately, this study seeks to contribute to a deeper understanding of the cognitive and contextual factors that shape the effectiveness of visual cues in language learning and speech perception.

Significance of the Study

This study holds significance both for the fields of second-language (L2) acquisition and multimodal speech perception, and I expect it to offer novel insights into how visual cues enhance learning and comprehension. First, it addresses a key gap in understanding the factors underlying individual variability in the McGurk effect, a phenomenon central to speech perception. By investigating the role of working memory and lipreading capacity, this research provides a deeper understanding of why some L2 learners are more responsive to audiovisual input than others. These findings could contribute to the development of more personalized approaches in language education, where instruction can be tailored based on cognitive profiles.

Second, this study's focus on the impact of content and context visuals on L2 listening comprehension has practical implications for instructional design and language pedagogy. If content visuals are more effective in enhancing discourse-level comprehension, educators could integrate more meaningful visual elements into L2 learning environments, especially for learners struggling with auditory comprehension. Moreover, understanding the limitations of context visuals will help refine how multimedia content is used in language learning materials.

Finally, this research contributes to broader discussions on how task complexity interacts with visual aids. By identifying how and when visual cues are most beneficial depending on the difficulty of the listening task, this study can inform the design of more effective language assessments and learning resources. The findings can potentially influence both classroom

instruction and the creation of digital learning tools, benefiting L2 learners of varying proficiency levels.

Definition of Terms

1. *Speech Perception*: The process by which the brain interprets and makes sense of spoken language. It involves auditory and visual cues, such as sounds and lip movements, to understand speech in various contexts such as noisy environments.
2. *L2 Language Acquisition*: The process of learning a second language (L2) after the first language (L1) has already been acquired. This process often involves challenges in speech perception and comprehension.
3. *Content Visual*: Visual elements that directly relate to the spoken material and provide meaningful context to support comprehension. For example, images or diagrams accompanying audio that convey critical information to help understand the speech content.
4. *Context Visual*: Visual elements that depict the background or setting of the speaker, without directly contributing to the specific content of the spoken material. Context visuals include environmental or situational cues but may have less impact on comprehension than content visuals.
5. *McGurk Effect*: A perceptual phenomenon in which conflicting auditory and visual stimuli (such as hearing one sound and seeing a different lip movement) alter speech perception. This effect highlights the role of visual information in speech perception and demonstrates how mismatched cues can lead to perceptual illusions.
6. *Lipreading*: The ability to interpret speech by visually observing the lip, tongue, and jaw movement. Lipreading is an important skill for speech perception, especially in noisy

environments or for individuals with hearing impairments, as well as for L2 learners who rely on visual cues to enhance understanding.

7. *Working Memory*: A cognitive function that allows individuals to store and manipulate information temporarily. Working memory plays a role in processing both auditory and visual information during speech perception.
8. *Task Complexity*: Refers to the level of difficulty of a listening or comprehension task, which can affect how well visual cues assist in speech perception and language comprehension. More complex tasks may require more cognitive resources, influencing how learners use visual and auditory information.

Assumptions, Limitations, and Delimitations

Assumptions

This study has several key assumptions. First, it is assumed that the participants in the study, especially the L2 learners, will make an honest effort to complete the tasks to the best of their abilities. It is also assumed that the language proficiency assessments accurately reflect the participants' actual proficiency levels. Additionally, it is assumed that the audiovisual materials used in the study are effective in eliciting the McGurk effect and in providing meaningful content or context visuals for the comprehension tasks. Finally, the study assumes that participants' self-reports regarding working memory capacities, as assessed through standardized tests, are valid indicators of these cognitive characteristics.

Limitations

There are several limitations to this study. One limitation is the sample size, which, while sufficient for statistical analysis, may not fully represent the diversity of L2 learners across different linguistic backgrounds. Another limitation is the reliance on self-reported data for

working memory tests, which may introduce bias or inaccuracies. Additionally, the study focuses on specific types of visual cues (content and context visuals) and may not account for other forms of visual information that could influence speech perception and comprehension. The experimental conditions, including the audiovisual materials and the complexity of the listening tasks, may also limit the generalizability of the findings to real-world language learning environments. Finally, the study only measures the immediate effects of visual cues on speech perception and comprehension, without assessing the long-term impact of these cues on language acquisition.

Delimitations

This study mainly focuses on L2 English learners of the same L1 language (Brazilians) to limit the variation due to different cultural backgrounds. Additionally, the study is limited to the McGurk effect and specific types of visual cues (content and context visuals) to keep the research scope manageable and focused. The study is also restricted to adult learners, as younger learners or children may process audiovisual information differently. Furthermore, the listening tasks are designed with varying levels of complexity but do not encompass the full range of real-world listening scenarios L2 learners might encounter.

Conclusion

This chapter has introduced the foundation of the present study, which investigates the role of visual cues in speech perception and language comprehension among second-language (L2) learners. The study seeks to address the variability in the McGurk effect, the impact of content and context visuals on discourse-level comprehension, and how task complexity influences the effectiveness of these visual cues. By exploring these relationships, the research aims to contribute to a more nuanced understanding of the cognitive and contextual factors that shape speech perception and language comprehension in L2 learners.

The chapter has provided an overview of the background, the problem statement, and the research questions that guide this study. It has also clarified the purpose and significance of the research, emphasizing the potential contributions to the fields of language acquisition, speech perception, and instructional design. Definitions of key terms have been outlined to ensure clarity for the readers, and the study's assumptions, limitations, and delimitations have been discussed to frame the scope of the research.

In the following chapters, I will review relevant literature, outline the methodology for data collection and analysis, and present the findings that emerge from this research. Ultimately, this study seeks to inform educational practices and contribute to the growing body of knowledge on the multimodal nature of language learning, particularly in L2 acquisition.

CHAPTER 2

LITERATURE REVIEW

In this chapter, I will provide an overview of the historical background of listening research and examine the various conceptual frameworks that have contributed to defining listening comprehension and constructing listening models. I will also review the research on visual cues and discuss how different visual cues impact speech perception and listening comprehension.

I will begin by outlining the complexity of listening research, introducing the three primary components involved in listening: cognitive, affective, and behavioral. I will then summarize and synthesize some of the key definitions of listening proposed by scholars across various research fields. In the final section, I will review the role of visual cues in speech perception and language comprehension. Specifically, I will explore the McGurk effect and how visual information contributes to speech perception in L2 learners. Additionally, I will examine different types of visual cues, such as content and context visuals, and their effects on language comprehension. This chapter will provide a background for my dissertation research.

Definitions of Listening

A great deal of listening research over the past half-century has influenced the conceptualization, teaching, and measurement of listening. Scholars in various fields have focused on different aspects of listening research. The following section reviews the most frequently discussed and referenced definitions of listening in the research literature and synthesizes how listening research has shaped its conceptualization and expanded its definitions.

The Complexity of Listening

Listening is a multidisciplinary field and has been defined in various ways across disciplines such as second language acquisition and applied linguistics. However, there is no universally agreed-upon, all-encompassing definition of listening (Wolvin & Coakley, 1996). "Listening" is often used as an umbrella term for both listening and listening comprehension, which are frequently defined and used interchangeably in different fields. The definitions of listening and listening comprehension share similar features, making it difficult to review one without discussing the other. Therefore, in this paper, "listening" refers to both listening and listening comprehension.

The majority of definitions and models of listening were developed in the early 1970s, with researchers drawing on the work of leading attention and memory theorists of the time. While many definitions and models have emerged over the past few decades, no single definition fully encompasses the complexity of listening, due to the field's multidisciplinary and multidimensional nature. As a result, there is significant diversity in theoretical approaches and a wide range of definitions and models of listening (Worthington & Bodie, 2017).

Listening has long been viewed as a passive and self-developing skill. However, this perspective began to change in the late 1960s and early 1970s, when listening was recognized as a fundamental and active skill (Field, 2009; Rubin, 1994). Listening research is fundamentally different from studies on hearing due to its focus on individual attention and cognitive effort. While hearing is often considered effortless, listening is an active process influenced by several factors. These factors include 1) the listener's working memory, anxiety, metacognitive strategies, and second language (L2) proficiency; 2) the speaker's accent and speech rate; 3) the length and complexity of the text, text type and organization, pauses and hesitations, and visual support; and 4) response requirements, time limits, note-taking, and the number and control of hearings

(Bloomfield et al., 2010; Rubin, 1994). These findings reveal that listening is not a simple process of recording a linear sequence of words transmitted to the brain (Brown & Abeywickrama, 2010). Instead, listening is a fundamental and active skill that involves complex, dynamic processes. As I will elaborate later, these processes include cognitive, affective, and behavioral components.

The complexity of listening is further highlighted by the variety of processes proposed by scholars from different fields. According to Wolvin (1989), numerous definitions of listening exist, with various models using different terminology and concepts. These models suggest that listening involves anywhere from three to over twelve distinct processes. Rost (2011) proposed that listening encompasses neurological, linguistic, semantic, and pragmatic processes, among others. Similarly, Nadig (2013) suggested that:

> Listening comprehension encompasses the multiple processes involved in understanding and making sense of spoken language. These include recognizing speech sounds, understanding the meaning of individual words, and/or understanding the syntax of sentences in which they are presented. Listening comprehension can also involve the prosody with which utterances are spoken (which can, e.g., change intended meaning from a statement to a question), and making relevant inferences based on context, real-world knowledge, and speaker-specific attributes (e.g., to what information the speaker has access and about what he/she is likely to be talking). For longer stretches of language or discourse, listening comprehension also involves significant memory demands to keep track of causal relationships expressed within the discourse. It is often viewed as an active process with three main components: attending to the perceptual input (speech), constructing meaning from stretches of speech, and relating what was heard to existing knowledge. (p. 1743)

Regardless of the differences in these proposals, researchers agree that listening is a multidimensional construct consisting of complex processes. In the following section, I will introduce the three primary components of listening: affective, cognitive, and behavioral.

Three Primary Components of Listening: Cognitive, Affective, and Behavioral

While "listening" implies paying attention and being quietly receptive, "hearing" is more associated with detecting sound and the abilities of the ear (Lipari, 2010; Bodie & Crick, 2014). This distinction typically serves to differentiate the focus of research conducted by audiologists, who examine the physiological components of hearing, from that of communication scholars, who investigate the individual and relational elements of listening (Worthington & Bodie, 2017). Halone et al. (1998) proposed that listening is a multidimensional construct that encompasses various complex processes: 1) cognitive processes, such as paying attention, comprehending, receiving, and interpreting both the content and underlying messages; 2) affective processes, including motivation to pay attention to others and the willingness to engage in listening; and 3) behavioral processes, which involve providing verbal and nonverbal responses, such as eye contact and body language. Scholars frequently emphasize one of the three categories, but there are instances where all three categories are combined (Steil et al., 1983).

Worthington and Bodie (2017) explain that cognitive components of listening are the internal processes individuals use to "attend to, comprehend, interpret, evaluate, and make sense of spoken language" and the idea that "listening is an information-processing activity consisting of a stable set of practices that can be trained and improved is the most popular way to conceptualize the term and one that has framed all listening research" since at least the early 1940s (p. 5). The listening research in the early 1920s was primarily focused on the understanding of orally delivered language, particularly the factual recall of large chunks of monolog in a classroom

setting, while the individual motivation or attention was not explicitly stressed (Tucker, 1925; Rankin, 1926). The study by Ralph Nichols in 1948, however, formally proposed that attention during classroom activities and conscientiousness in work habits could help differentiate between good and poor listeners (Nichols, 1948). In Nichols's study, a group of undergraduate students were asked to answer 10 multiple-choice questions after listening to six 10-minute audio-recorded lectures. Student participants were able to recall on average 68% of the lecture material without the assistance of note-taking, and good listeners tended to be more attentive in the classroom than poor listeners (Nichols, 1948). Nichols's study is regarded as a watershed in contemporary listening research because his study was among the first to acknowledge that cognitive ability contributes to listening competence. It is worth noting that memory became a primary cognitive component in listening research around the mid-1960s, and the extensive theorization of the relationship between listening and memory by Bostrom and Waldhart (1980) helped develop new measurements of listening based on short-term and long-term memories.

Affective components of listening include individuals' personal motivation about listening and their enjoyment of the activities. In other words, the affective components of listening refer to the emotional and motivational aspects of listening, such as our attitudes, beliefs, and values that influence our listening behavior. It includes our ability to perceive, identify, and respond to emotional cues conveyed through spoken language, as well as our motivation to engage in the listening process. Based on this idea, individuals' views, attitudes, and interpretations about their listening activities could have a profound impact on their listening experience as well as their personal and professional success (Worthington & Bodie, 2017). We know that early research on listening focused on the retention, recall, and comprehension of spoken information in educational contexts. It was not until 1972, when Weaver included individuals' predispositions and

interpretations into the model of listening, that he established listeners' attitudes and willingness to listen as a separate component of the listening process (Weaver, 1972). Since then, a significant portion of affective research has focused on associations between listening and personality traits that may affect individual motivation.

Behavioral components of listening consist of verbal and nonverbal acts that indicate a listener's motivation, willingness, interest, and attention to listen, such as eye contact and asking questions. Unlike affective or cognitive components, behavioral components were not part of the focus of listening research until the mid-1980s, when developing speaking and listening competencies became an emerging trend in high school and college education (Worthington & Bodie, 2017). The listening competency model that resulted from research conducted in the 1980s closely mirrored the communicative competence model that was popularized by Wiemann, Spitzberg, Rubin, and other researchers (Morreale et al., 1998; Spitzberg & Cupach, 1984; Wiemann & Backlund, 1980; Wilson & Sabee, 2003). The models developed to explain listening proficiency emphasized the observable actions of the listener as the key determinant of whether they were considered a "good" or "poor" listener. This emphasis on behavior was a logical extension of earlier studies that highlighted the importance of retention and recall outcomes (Worthington & Bodie, 2017). At the core of the "listening as competent behavior" perspective is the belief that "an identifiable set of skills, attitudes, and abilities can be formulated and taught to improve individual performance" (Bodie et al., 2008, p. 107). Such views helped promote the shift of the research focus from covert mental processes to overt behavioral ones in the late 1980s throughout the 1990s. Accompanying the conceptual shift, the emergence of new measurement techniques such as third-party and critical-incident techniques (Rubin & Feezel, 1986; Wellmon, 1988) facilitated the assessment of listening behaviors. Furthermore, it was increasingly

recognized that the ability to listen effectively was contextual, which led researchers to investigate listening skills in various fields such as business, education, and healthcare. These researchers have linked listening proficiency (assessed using various methods) to important factors like attentiveness, memory retention, and comprehension, as well as employee drive, career advancement, and academic achievement (Brownell, 1985; Rubin & Feezel, 1986; Sypher, Bostrom, & Seibert, 1989; Wanzer, Booth-Butterfield, & Gruber, 2004; Worthington, 2001).

The Evolving Definition of Listening

As discussed above, the complex nature of listening as well as the expanding focus of listening research contributed to the evolving definition of listening. The table below summarizes sample definitions of listening by research scholars dating back to the 1920's.

Table 1

Sample of Listening Definitions

Author	Year	Definition
Tucker	1925	An analysis of the impression resulting from concentration where an effort of will is required
Rankin	1926	The ability to understand spoken language
Nichols	1948	The comprehension of expository materials presented orally in a classroom situation
Barbe & Meyers	1954	The process of reacting to, interpreting, and relating the spoken language in terms of past experiences and a future course of action
Brown & Carlson	1955	The aural assimilation of spoken symbols in a face-to-face speaker–audience situation, with both oral and visual cues present
Barbara	1957	A definite, usually voluntary, effort to apprehend acoustically
Spearritt	1962	The active process involved in attaching meaning to sounds
Petrie	1964	The whole process through which spoken language is received, critically and consciously attended to, identified, and comprehended in terms of prior experiences and future expectations
Barker	1971	The selective process of attending to, hearing, understanding, and remembering aural symbols

Weaver	1972	A process that takes place when a human organism receives oral data; the selection and retention of aurally received data
Kelly	1975	A rather definite and deliberative ability to hear information, to analyze it, to recall it at a later time, and to draw conclusions from it
Steil *et al.*	1983	The listening act really consists of four connected activities - sensing, interpreting, evaluating and responding
Wolff *et al.*	1983	A unitary-receptive communication process of hearing and selecting, assimilating and organizing, and retaining and covertly responding to aural and nonverbal stimuli
Wolvin & Coakley	1988	The process of receiving, attending to, and assigning meaning to aural stimuli
Brownell	1994	A definition of listening as overt behaviors clustering around five distinct components: hearing (concentration), understanding (comprehension and memory), interpreting (sensitivity to nonverbal cues), evaluating (objectivity), and responding (acting on what was heard)
ILA	1996	The process of receiving, constructing meaning from, and responding to spoken and/or nonverbal messages
Cooper	1997	Listening competency means behavior that is appropriate and effective. *Appropriateness* means that the content is understood, and *effectiveness* deals with the achievement of interactive goals
Bostrom	1997	Listening is the acquisition, processing, and retention of information in the interpersonal context
De Ruyter & Wetzels	2000	A set of interrelated activities, including apparent attentiveness, nonverbal behaviors, verbal behavior, perceived attitudes, memory, and behavioral responses
Brownell	2002	Hearing, comprehending, recalling, interpreting, assessing, and reacting are all parts of listening
Bostrom	2011	The acquisition, processing, and retention of information in the interpersonal context

Note. Adapted from "Defining Listening" (p. 4), by D. L. Worthington & G. D. Bodie, 2017, *The Sourcebook of Listening Research*, pp. 3–17. Copyright 2017 by John Wiley & Sons, Inc.

In the 1950s, psychological research investigated listening as part of attention and

memory research. Early listening studies concentrated on understanding verbally conveyed information in educational settings, a focus that limited listening to a type of information processing without its broader relationship to human communication and relational experiences (Bostrom, 2011). Accordingly, the definition of listening dating back to the 1920s to mid-1940s was focused primarily on the information retention of lecture materials in the classroom setting. Nichols's seminal study in 1948 changed how listening was defined by introducing cognitive components into listening research. After that, the extensive research efforts on attention and memory resulted in the development of new definitions and major cognitive models of listening in the 1970s and 1980s (e.g., Barker, 1971; Weaver, 1972; Kelly, 1975). This is consistent with the content study of fifty definitions of listening by Glenn (1989), which concluded that the cognitive processes of perception, attention, comprehension, memory, and response were the five most used and shared elements in listening definitions. Almost thirty years later, Worthington (2017) showed that it is vital to emphasize that attention, retention (or memory), comprehension, and inference-making are the fundamental cognitive components of listening and these areas gained significance and have continued to be focal points for researchers in the field of listening. I will assess a few of these components in a later section.

The exploration of cognitive components of listening has continued into contemporary listening research. For example, Trenholm and Jensen (2011) define "listening [as] a complex process that takes a great deal of cognitive energy and skill" (p. 108) and "a complex, interactive, multistage process. To listen effectively, it is necessary to attend to, interpret, evaluate, and respond to messages" (p. 109). They also define four interrelated steps in the listening process: "1) attention, the act of selectively focusing on certain communication cues while ignoring others; 2) interpretation, the act of assigning meaning to the stimuli that capture our attention; 3) evaluation,

making decisions about the accuracy and usefulness we have interpreted; 4) responding, offering your partner some overt indication of interest and support" (pp. 109-110). In *Communication in the Real World* (2016), a definition of listening is described as:

> Listening is the learned process of receiving, interpreting, recalling, evaluating, and responding to verbal and nonverbal messages and it has cognitive, behavioral, and relational elements and doesn't unfold in a linear, step-by-step fashion. Models of processes are informative in that they help us visualize specific components, but keep in mind that they do not capture the speed, overlapping nature, or overall complexity of the actual process in action. The stages of the listening process are receiving, interpreting, recalling, evaluating, and responding. (p. 212)

It is interesting to note that although many definitions of listening specify cognitive functions, the authors of the work were primarily communication and listening scholars, who were not directly involved in primary brain research. An examination of the references of their work reveals that their definitions are founded in one of two ways (Janusik, 2002): 1) the researchers referenced psychological scholars, e.g., Broadbent (1958), Treisman (1960, 1964), Deutsch and Deustch (1963), and Kahneman (1973); or 2) the researchers referenced listening theorists whose work is founded on the psychological researchers, e.g., Barker (1971) and Weaver (1972).

Our understanding of listening has deepened in the past decade, as emerging evidence has indicated that listening indeed requires both foundational cognitive skills (working memory and attention), foundational language skills (vocabulary and grammatical knowledge), and higher-order cognitive skills (inferencing and theory of mind) (Florit et al., 2011; Kim, 2016; Kim & Phillips, 2014; Lepola et al., 2012; Tyagi, 2013). Based on these findings, listening involves the

extracting and constructing of meanings from aural signals, and it can be defined as one's ability to understand spoken language at the discourse level (Kim & Pilcher, 2016).

Last but not least, a large amount of research indicates that earlier definitions of listening overlooked the non-verbal, particularly the visual components, in listening processing. Glenn (1989) noticed that only eight out of fifty listening definitions since 1925 acknowledged the visual aspects. Many contemporary researchers define listening by acknowledging the visual roles during the listening process. For example (see Table 1), according to Wolvin and Coakley (1996), listening is "the process of receiving, attending to, and assigning meaning to aural and visual stimuli" (p. 69). This definition implies that listening involves more than just a verbal component. Moreover, according to the International Listening Association (ILA), "listening is the process of receiving, constructing meaning from, and responding to spoken and/or nonverbal messages" (Emmert, 1994, p. 6). Concise yet thorough, this milestone definition explicitly acknowledges the cognitive and behavioral components of listening. This working definition also recognizes both verbal and nonverbal components of the listening process, but it leaves room for readers' interpretation of the word "non-verbal." Despite the growing consensus among academics that visual stimuli play a significant part in listening, there are scenarios in which listening relies only on auditory inputs, such as phone conversations, radio listening, and blind people's listening (Olson, 2003; Wagner, 2007). My research will address this visual aspect of listening specifically.

Reviewing the definitions of listening in the above section acknowledges the complexity of listening, and demonstrates why there is no agreed-upon, all-encompassing definition due to the multidisciplinary and multidimensional nature of the field. In fact, given the multiple processes and components that go into listening, some scholars contend that a universal definition of listening is challenging and possibly even unattainable (Wagner, 2002). Some suggest that it is preferable

to have multiple definitions to acknowledge that listening is multifaceted and multipurpose (Bodie, Janusik & Välikoski, 2008). There are numerous definitions of listening, various models of listening utilize different terms and concepts, and listening encompasses anywhere from three to over a dozen distinct processes. This diversity has both advantages and challenges. It has uncovered the intricate nature of listening, providing researchers and practitioners with additional and improved approaches to enhance this essential life skill. However, it has also led to a fairly fragmented field. (Worthington & Bodie, 2017).

Concluding "The" Definition of Listening

The early listening scholars tried to develop a mutually agreeable definition of listening depicting the inner workings and outward actions that constitute the construct and establishing its connections with other variables. As outlined in the above section, the early listening scholarship was narrowly focused on oral information processing, ignoring its broader links to human communication and social interactions (Bostrom, 2011). Nearly all the early listening measurements highlighted "listening comprehension and recall, a trend that continued for several decades" (Worthington & Bodie, 2017, p. 10). However, it is essential to view listening as a set of complex skills and abilities. The set of skills and abilities that should be included in a definition of listening was debated furiously in listening literature from the 1970s to the 1990s (Worthington & Bodie, 2017). As detailed in the earlier section, some listening literature was grounded in cognitive psychology, such as that of Bostrom—who addressed the role of memory—and others who proposed models based on human information-processing methods (Fitch-Hauser, 1990; Goss, 1982). During the course of the above three decades, "models of listening proliferated, with most stressing the internal, working apparatus thought to be necessary to process spoken language" (Worthington & Bodie, 2017, p. 10). The disagreement among scholars regarding how to precisely

define listening reflects the historical context of the field. It encompasses debates about the nature of listening and varying viewpoints on how it should be approached.

Given the complex multidisciplinary and multidimensional nature of listening and variable appropriate theoretical frameworks for investigating listening, there ought to be myriad definitions that contribute to the development of the field (Bodie, 2010, 2012), and each depends "on the practical purpose pursued by an individual or team of scholars" (Bodie, 2012, p. 114). According to Worthington and Bodie (2017), rather than creating theories based on definitions, definitions are derived from theories, and the meaning of "listening" as a theoretical concept is shaped by the theoretical structure it is viewed. As various theoretical frameworks offer different concepts and procedures, the objective is to create multiple listening definitions and delve into the various intricacies involved in the process of listening rather than trying to establish a common definition.

When an ordinary person talks about listening, they are typically describing a state of social interaction and being attentive to others (Purdy, 2006). These implicit theories of listening form a significant aspect of the cognitive and affective components of listening and may influence how individuals evaluate others who exhibit specific behaviors (Bodie, 2010; Bodie et al., 2015). In other words, implicit theories of listening referring to the unconscious beliefs and assumptions that individuals hold about the nature and process of listening shape people's expectations and perceptions of listening behavior and impact how they evaluate the effectiveness of listeners and their own listening skills. Conversely, scholars who have conceptualized listening have tended to concentrate on "the cognitive processes responsible for understanding, comprehending, evaluating, and responding to spoken messages." Asking which of these views is accurate is akin to inquiring about which of the many definitions of any term provided in a dictionary is correct. Thus, "definitions are functional, not right or wrong, but more or less useful for some particular

purpose" (Worthington & Bodie, 2017, p. 11). They also recommend that researchers prioritize identifying the fundamental characteristics of the specific listening processes or behaviors relevant to their research project. They assert that the primary focus of the investigation should be the research objectives, which will guide the selection of how listening and associated concepts are defined and operationalized.

The definitions and models of listening reviewed in this paper have been expanded and will continue to be developed by researchers in various fields, including second language acquisition, applied linguistics, education, communication, psychology, neuroscience, health, management, and business. Although the listening field has been in existence for more than 50 years, there is still a significant lack of full understanding regarding the factors that contribute to the listening process (Worthington & Bodie, 2017). Reviewing Table 1 again, these conceptualizations of listening are not sufficiently supported by scientific rigor and leave additional room for what contributes to the listening process. Theoretical frameworks should form the basis of conceptual definitions, which should be revised as time goes on. I hope my research focusing on one of the key features of a specific listening process, how visual speech cues influence speech perception and listening comprehension, will contribute to the conceptualization and subsequent operationalization of listening, will further expand the theoretical framework for this multidimensional and fascinating field, and will aid future scholars engaging in more listening research practices.

The McGurk Effect

In a digital multimedia age, in which it has become routine for spoken language messages to be accompanied by visuals, we take for granted the notion that visual cues affect listening and speech perception. Nevertheless, the significance of visual cues once received insufficient

recognition in early speech perception research, which primarily focused on the auditory linguistic input channel. However, the discovery of the McGurk effect by McGurk and MacDonald in 1976 has inspired visual cue research in speech perception. In their pioneering study, McGurk and MacDonald (1976) found that combining the visual velar /kɑ/ and auditory bilabial /pɑ/ resulted in 81 percent of the participants reporting hearing /tɑ/, which is the voiceless stop consonant lying midway between /kɑ/ and /pɑ/. In addition, the visual bilabial /pɑ/ combined with the auditory velar /kɑ/ led to 44 percent of perception of a combined signal /pɑkɑ/. The perception of discordance is influenced by the perceiver's ability to accurately identify visual and auditory cues, which can be challenging for some L2 learners, leading to inconsistent observations.

Consonant-vowel (CV) Syllables Affect the Expression of the McGurk Effect

Subsequent research found that the contribution of information from either the visual or auditory modality to the perceptual outcome depends on factors such as the adjacent vowel, the presence of noise, and the degree of discordance or conflict between the auditory and visual cues. For example, Green et al. (1988) found that the effect of visual cues on the auditory perception of a specific consonant depends on the context of vowels. Using the auditory /b/ and visual /g/ combination, the author investigated the occurrence of illusory /d/ responses in three vowel environments: /a/, /i/, and /u/. The findings showed that the visual /g/ had the strongest impact on auditory /b/ when the /i/ was used as a context vowel, leading to the highest response of illusory /d/. The /a/ vowel had a moderate effect, while /u/ almost had no effect (Green et al., 1988). The results demonstrate that the vowel environment plays a critical role in determining the magnitude of the McGurk effect.

Hardison (1996) later performed experiments to further explore the connection between a variety of consonant-vowel syllables and the McGurk effect. She found that although native

English speakers only showed significant effects in noisy conditions, both native and non-native speakers reported considerable McGurk effects when /p, t, k/ were used. These findings further support the impact of consonant-vowel syllables as well as L1 (which will be discussed below) on the audiovisual integration and expression of the McGurk Effect.

The McGurk Effect Varies in Different L1 Backgrounds

Research has found that L1 impacts the McGurk effect. In the early 1990s, researchers began exploring the McGurk effect in non-English speakers, including Chinese-Dutch bilinguals (De Gelder & Vroomen, 1992), Japanese speakers (Sekiyama & Tohkura, 1991; Sekiyama, 1994), and Chinese speakers (Sekiyama, 1997), and they found differences in the expression of the McGurk effect. In Hardison's (1996) experiment with Japanese, Korean, Spanish, and Malay speakers learning English in the US, the L2 learners were tested with American English consonant-vowel (CV) syllables (including a variety of consonants /p, f, w, r, t, k/ combined with /ɑ/) in both congruent and incongruent audiovisual conditions. Visual cues helped improve Japanese and Korean learners' identification accuracy of /f/ and /r/, which are difficult for these learners, compared to their audio-only identification accuracy. Visual non-labial sounds /t, k/, which were not challenging for these learners, contributed more to perception only when combined with the more confusable auditory labial sounds such as /p, f/.

Further evidence of the influence of L1 on the McGurk effect was provided by Wang et al. (2009). They used audio-only, video-only, and congruent and incongruent audiovisual presentations of English CV syllables for L1 Korean, Mandarin, and English speakers. The stimuli contained fricatives at different points of articulation: labiodental (/f, v/: not found in Korean), interdental (/θ, ð/: not found in Korean and Mandarin), and alveolar (/s, z/: present in all three L1s) combined with the vowels /i, ɑ, u/. The results showed Korean speakers had difficulty identifying

labiodentals in the video-only tests and both Korean and Mandarin speakers had difficulty identifying interdentals in the audio-only tests. Combining auditory and visual cues improved their performance in these tests. These findings highlight the importance of L1 on L2 perception and production.

In another study, Hazan et al. (2010) found that the characteristics of individual speakers affect the relative weight assigned to auditory and visual cues. As a result, the perception varies across listeners. Consonant-vowel syllables /ba/, /da/, and /ga/, produced by Australian English and Mandarin speakers, were presented to Australian English, British English, and Mandarin listeners in several stimulus conditions: audio-only, video-only, congruent and incongruent audiovisual, and in clear or degraded forms (with noise, visual blurring, or combined audiovisual degradations). The results showed that speech produced by a second language (L2) speaker and the listener's native language (L1) background influenced the weighting of auditory and visual cues in perception.

Lastly, research shows that bilinguals are more likely to experience the McGurk effect than their monolingual peers. Marian et al. (2018) compared English-speaking monolinguals with early (highly proficient) and late (less proficient) Korean-English bilinguals, examining their perception of speech sounds accompanied by either congruent or incongruent visual information. The authors found that bilingual participants were more likely to perceive fused McGurk-type sounds (e.g., /da/) when presented with incongruent auditory and visual information (e.g., hearing /ba/ but seeing /ga/) than their monolingual peers. Importantly, both late and early bilinguals experienced similar McGurk effects, indicating that the effect did not simply result from poor language comprehension. Rather, the findings indicate that bilingual individuals were more influenced by

visual input compared to monolinguals, and the challenges of learning and managing multiple languages have a lasting impact on auditory and visual information processing.

The Impact of Visual Cues on Speech Perception

The McGurk effect observed in Japanese and Korean L2 English learners led to further investigations on how visual cues could be enhanced to improve learners' identification accuracy of /r/ and /l/ in American English through targeted perception training. These sounds have been extensively studied in L2 speech research due to their acoustic variability and the perceptual difficulties they pose for Japanese and Korean learners (Hardison, 2003). Being aware that Japanese learners use facial (especially lip) cues to help understand the speaker's utterance, Hardison (2003) conducted high-variability perception training (HVPT) in intermediate-level Japanese and Korean L2 English learners, which "emphasizes the use of multiple speakers and diverse phonetic contents to increase learners' awareness and tolerance of variation" (Pennington and Rogerson-Revell, 2019: 200).

The study by Hardison (2003) contained five different sources of variability that impacted the auditory perception of /r/ and /l/. Three types of variabilities were used previously by Lively et al. (1993), which included the target sound's position in a word, the use of multiple voices instead of a single voice, and natural speech as opposed to synthesized speech. Hardison added two more sources of variability: visual input capturing facial cues from multiple speakers, and the vocalic context of /r/ and /l/. For instance, high vowels /i/ and /u/ typically present greater perceptual challenges compared to lower, unrounded vowels (Hagiwara, 1995). Hardison (2003) found that with audiovisual input, visual cues played a significant role in perception for Japanese speakers, particularly in the challenging phonetic environment of word-initial clusters. Stimuli with relatively open vowels (/ɑ/ or /aɪ/) resulted in improvement earlier in the training process than

those with rounded vowels (/u/ or /o/). In contrast, Korean speakers had higher identification accuracy scores for /r/ and /l/ in word-final positions, especially following /i/. Both language groups experienced improved perceptual accuracy for /r/ and /l/ in intervocalic positions.

Similar to Lively et al. (1993), Hardison (2003) found that the L2 learners showed improved identification accuracy of novel stimuli and those produced by new speakers following the training. These skills also led to significant production improvement without explicit production training, suggesting a link between perception and production. These findings showed that seeing lip gestures significantly facilitated L2 learners' perceptual identification and pronunciation accuracy.

In another study, Li and Somlak (2017) used audio-visual aids (particularly articulatory gestures) to help L2 learners acquire challenging L2 speech sounds in a classroom setting. Participants were exposed to seven audio- or audio-visual poems containing the target contrasts /θ/–/s/ and /ð/–/z/ over the course of seven weeks. Pronunciation tests were conducted using 'read-aloud' tasks. The authors found that the students who received audio-visual recordings and observed the speakers' articulatory gestures showed significant improvement in pronunciation, while the students who received audio-only recordings did not. In the delayed post-test one month after the completion of the teaching program, the audio-visual group still retained the pronunciation improvement. These findings are consistent with Hardison's (2003) study and confirm the effectiveness of audio-visual aids in L2 perception and/or production.

Content vs. Context Visual Cues in L2 Speech Perception

Visual cues can be broadly categorized into content and context visual cues. The distinction between these two is pivotal for understanding their impacts on L2 speech perception.

Content visual cues refer to the visual signals directly related to the linguistic information of speech, such as a photo, graph, or drawing that aligns with the spoken words (Ginther, 2002). For L2 learners, these cues can significantly aid in understanding speech, especially in challenging environments. A study on the audiovisual speech perception of plosive consonants by Cypriot-Greek (CG) learners of English emphasized the importance of these cues for speech perception, suggesting that they play a crucial role in enhancing comprehension (Kkese & Dimitriou, 2023).

By contrast, context visual cues relate to the broader environment or situation in which speech occurs, including the speaker's attire, background setting, and any props or visual aids used during communication (Ginther, 2002). These cues provide a backdrop that can influence the listener's interpretation of the speech. Research on visual prosody and speech intelligibility indicates that head movement, a context visual cue, can improve auditory speech perception, emphasizing the importance of such cues in L2 learning (Munhall et al., 2004). In another study, Spehar et al. (2015) investigated the effects of two types of contextual cues—sentence-based and situation-based—across two modalities: visual-only and auditory-only. They found that participants showed better sentence comprehension with both types of context in the Illustrated Sentence Test and Speech Perception in Noise Test, with greater advantages in the visual-only modality. However, context visual cues may impair listeners' performance in some cases (for example, Suvorov (2009), as discussed later), suggesting how context visual cues impact L2 speech perception depends on exact situations.

The Audiovisual Integration Varies with Different L1 Backgrounds

Similar to the role of L1 in the presentation of the McGurk effect, the native language of the listeners plays an important role in L2 perception. In a study that aimed to evaluate how sensitive second language learners are to phonetic information present in visual cues, Hazan et al.

(2006) tested Spanish and Japanese learners of English on their perception of a labial/labiodental consonant contrast in audio, visual, and audio-visual conditions. The authors found that Spanish students demonstrated better overall performance and significantly higher sensitivity to visual cues than their Japanese peers. Both learner groups achieved higher scores in the AV condition compared to the A condition, indicating a clear benefit of visual cues in L2 perception.

Phonological Short-Term Memory (PSTM) Contributes to Audiovisual Integration

Researchers have also hypothesized that working memory (WM) in general and phonological short-term memory (PSTM) specifically may affect individual ability to integrate audiovisual information. For example, a recent study by Inceoglu (2019) examined the impact of individual differences in lipreading ability, WM, and PSTM on L2 speech perception. The author recruited thirty-two L2 French learners (L1 Australian English) in an L2 French vowel identification task under audiovisual, audio-only, and visual-only conditions. The participants also completed tasks to measure their first language (L1) lipreading ability at phoneme, word, and sentence levels; a listening span task to assess WM capacity; and a non-word repetition task to evaluate PSTM. The findings revealed that audiovisual and audio-only conditions resulted in significantly better speech perception while lipreading, WM, and PSTM tasks showed large individual variability. In addition, PSTM but not WM capacity showed a significant effect on vowel perception scores across all presentation modalities, while word-level lipreading ability was found to predict accurate L2 vowel perception in the visual modality.

Another important cognitive ability for non-native perception is acoustic memory (AM). Though not extensively studied, AM is thought to store acoustic information at a pre-categorical level and contribute to L2 speech perception and learning (Safronova, 2016). To find out how PSTM and AM contribute to L2 perception, Kogan (2022) asked monolingual Spanish speakers

to discriminate between two members of an unfamiliar Russian contrast /i - ɨ/ that does not exist in Spanish. The participants' PSTM and AM were measured using a non-word recognition task and a target sound recognition task, respectively. The results showed that AM capacity influenced the perception of the unfamiliar contrast, while PSTM capacity did not have a significant impact. In a follow-up study with a similar design, the author found that L2 Russian learners (L1 Spanish) relied on both PSTM and AM to discriminate between Russian /i/ and / ɨ /. It appears that naïve listeners utilize auditory-based pre-categorical judgments when discriminating between two members of a novel contrast, while L2 learners employ both acoustic and phonological information in the same task.

Weakness and Gaps of Current Audiovisual Integration Research

Significant Individual Variance in the McGurk Effect

The McGurk effect is a strong illusion that is proposed to support the impact of visual cues on auditory speech perception. However, experimental subjects consistently show considerable individual differences in susceptibility to the illusionary effect. Although some individuals experience the McGurk effect almost every time they encounter an incongruent visual signal, others rarely (or never) show such an effect (Strand et al., 2014). In addition, Basu Mallick et al. (2015) examined the perception of 12 different McGurk syllable pairs by 165 undergraduate students from Rice University tested in the laboratory. There was high variability, both across stimuli (rates ranging from 17% to 58%) and across participants (rates from 0% to 100%). They also found that response type significantly alters the frequency of the McGurk effect, with forced-choice responding increasing the frequency of McGurk perception by an estimated 18 %, as compared with open choice for identical stimuli. Since audiovisual (AV) integration models propose a two-stage speech processing that includes unimodal identification (either auditory or

visual) and AV integration, it is unclear whether the observed variability reflects individual differences in integration skill or unimodal extraction ability, such as lipreading.

The study by Strand et al. (2014) examined the role of lipreading skills in the presentation of the McGurk effect in native English speakers. The authors found that some of the variability in the effect can be attributed to lipreading skills, although the effect was not strong. Lipreading skills have been associated with cognitive abilities such as working memory and processing speed (Feld & Sommers, 2009). To explain the remaining variability in the McGurk effect, it would be helpful to investigate the contribution of cognitive and/or perceptual factors to the effect. In addition, the authors observed significant individual variability in the McGurk effect when measured in both the identification task (reporting fused responses) and the detection task (reporting whether visual and auditory signals are congruent or incongruent). The individual differences in identification and detection are only moderately correlated, suggesting that perceiving McGurk fusions does not entirely depend on failing to detect audiovisual incongruity. In conclusion, the individual differences in the McGurk effect have not been fully explained by lipreading skills, and further work would be needed to understand the mechanisms underlying the variance.

Failure of Visual Cues to Improve L2 Comprehension in Some Studies

Despite all the evidence showing a positive effect of visual cues on L2 perception and/or comprehension, a number of studies failed to find such a connection. For example, the study by Kamiya (2022) investigated the impact of observing gestures and lip movements on English listening comprehension for both high-proficiency and low-proficiency L2 learners (L1 Japanese). Having realized the controversial findings regarding whether visual cues improve L2 perception, Kamiya used a within-participant setup to minimize the individual variance. By using three modality types: body (visible upper half of the body), face (close-up view), and audio, Kamiya

found that listening modality did not affect the performance of either high- or low-proficiency participants, even though they generally preferred watching the whole body, followed by the face, and finally listening only.

Further examination revealed two possible reasons for the failure of the visual cues to improve L2 comprehension. First, the visual cues used in the study may not effectively contribute to listening comprehension. Visual cues can be divided into two types: (1) content visuals, which depict the content of the speech, and (2) context visuals, which refer to the background information in which the speech takes place (Ginther, 2002). Ginther (2002) found that compared to context visuals, content visuals promote better comprehension in listeners. In another study, Suvorov (2009) found that video material with context visuals were linked to worse performance compared to audio and photograph materials, supporting a better role of content visuals in enhancing listening comprehension. The visual materials used in Kamiya's (2022) study contained primarily context visuals, with few content visuals available to the participants, which may have diminished the effectiveness of the visual cues.

Secondly, besides the lack of effectiveness of visual cues, the high difficulty level of the materials used by Kamiya (2022) could also restrict the participants from allocating sufficient cognitive resources to the nonverbal information. The high task demand may force the participants to focus on auditory information and ignore nonverbal cues such as gestures and lip movements (Cross, 2011). Together with the reduced effectiveness of visual cues (context versus content), the limited cognitive processing capacity might explain the lack of effects of visual cues on listening comprehension.

In all, further work is needed to understand the mechanisms underlying individual variance in McGurk effects. In particular, more work is needed to elucidate the effects of WM, the effects

of content versus context visual cues, and the mutual effects of listening task complexity and visual cues in L2 listening comprehension.

Research Questions

In this dissertation study, I investigated the role of visual cues in speech perception and language comprehension in L2 English learners. In addition, I controlled the difficulty level of the test materials to ensure they place a proper level of cognitive burdens. Such control is to avoid too little burden, such that participants understand well without relying on visual cues, or too much burden, such that participants cannot pay sufficient attention to nonverbal cues. Specifically, I addressed the following questions:

(1) To what extent does the individual variability in the McGurk effect depend on working memory, and/or lipreading capacity?

(2) Do content or context visuals enhance discourse-level listening comprehension in L2 English learners?

(3) Does the relationship or the influence of visual cues, if present, vary depending on the complexity of the listening task?

CHAPTER 3

RESEARCH METHOD

Participants

The study recruited a total of 98 Brazilian L1 speakers of Portuguese and L2 speakers of English, of which 93 completed the research tasks[1]. All participants were given informed consent and underwent pre-test screenings before the experiments.

The participants were recruited through two channels: posters and solicited emails. For posters, I designed and distributed informative and engaging posters across the UGA campus, particularly in high-traffic areas such as student centers, libraries, dining halls, and departmental bulletin boards. The poster included a brief study overview, the expected time commitment, and the compensation details. Interested students contacted me through the email on the poster. For solicited emails, I collaborated with my friends to send out solicited emails. Interested participants contacted me through email, and I provided more in-depth information about the study.

All recruitment materials disclosed the purpose of the study and how the data would be collected and used. To value the time and effort of the participants and encourage their participation, I offered $30 compensation, which was paid to the participants through Zelle or PayPal once their test completion was confirmed. The test was designed to ensure participants could easily navigate and the whole test took no more than two hours to finish.

Data Collection Procedures

[1] I also recruited 10 Chinese, 1 Russian, 1 Nigerian, and 2 Vietnamese participants whose data is not included in this analysis.

All experiments were completed online to facilitate data collection. After reading the Informed Consent letter, participants were provided with the link for a course on the Thinkific online platform and received instructions on how to sign up for the course (https://xiao-s-site-6167.thinkific.com/courses/take/test-1/texts/56120840-introduction). They signed up for a Thinkific account for free using their Google account or preferred email address. Participants were then instructed to enroll in the course to finish the survey and the tests. They were then instructed to complete several initial screenings including 1) a background questionnaire (Biographical information and Linguistic information), 2) a vision screening, 3) a hearing screening, and 4) a working memory task (detailed below). The screening took about 25 minutes to finish, and all participants except one passed. The participants were then instructed to perform two tasks: 1) a speech perception task (McGurk effect) and 2) a language comprehension task (detailed below). The whole test took about 2 hours to complete. Table 2 lists the details of each test.

Table 2

The Sequence of Tasks for Online Data Collection

Sequence	**Task**	**Approximate Time to Complete (mins)**
1	Language Background Questionnaire	5
2	Vision Screening Test	2-3
3	Hearing Screening Test	5
4	Working Memory Test	5
5	McGurk Effect (Speech Perception)	30

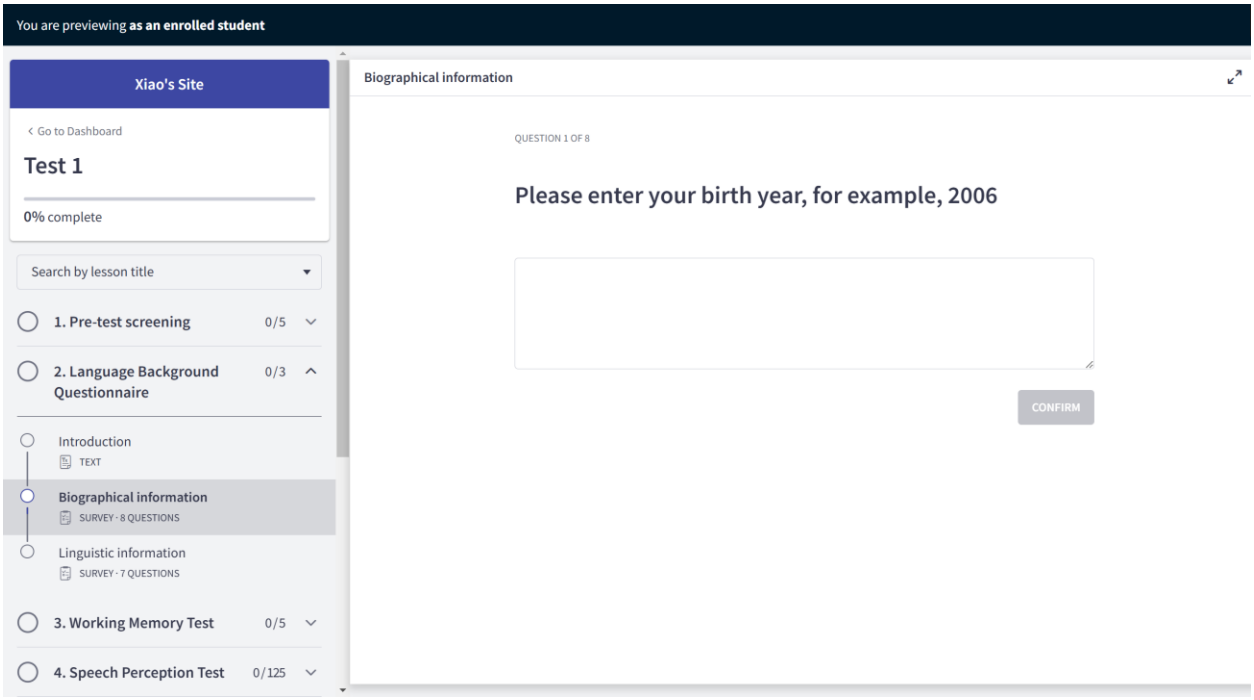
6	Lipreading Test	20
7	Language Comprehension (content visual, context visual, and audio tests)	50
Total		120

Background Questionnaire

The language background questionnaire followed the design of Sabourin et al. (2016). The questionnaire identified participants' native language, secondary language, time of language exposure, level of proficiency, parental language information, participant occupation, and standardized English test scores. The questionnaire was completed in the online course I created at (https://xiao-s-site-6167.thinkific.com/courses/take/test-1/texts/56153453-introduction), which took 5-10 minutes to finish. (See Appendix A for background questionnaire questions.) Figure 1 shows the screenshot of the Language Background Questionnaire.

Figure 1

Screenshot of the Language Background Questionnaire Section



Pre-test Screenings

Vision Screening

Figure 2

Screenshot of the Vision Screening Instructions

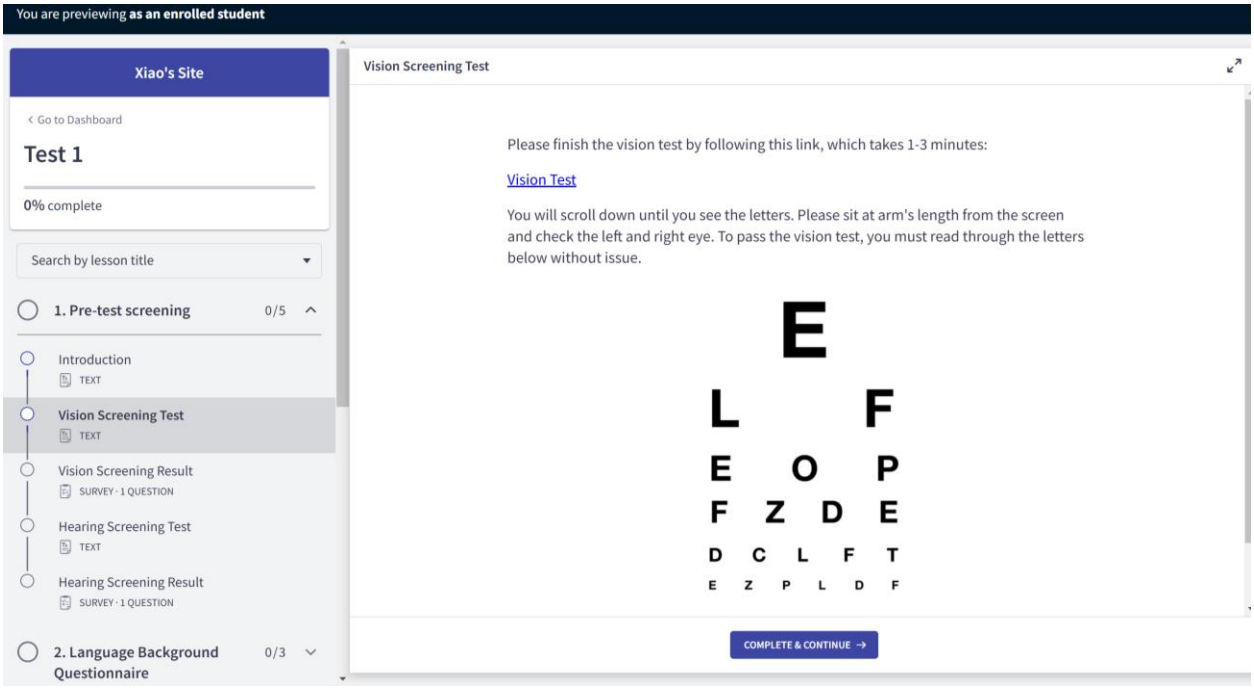
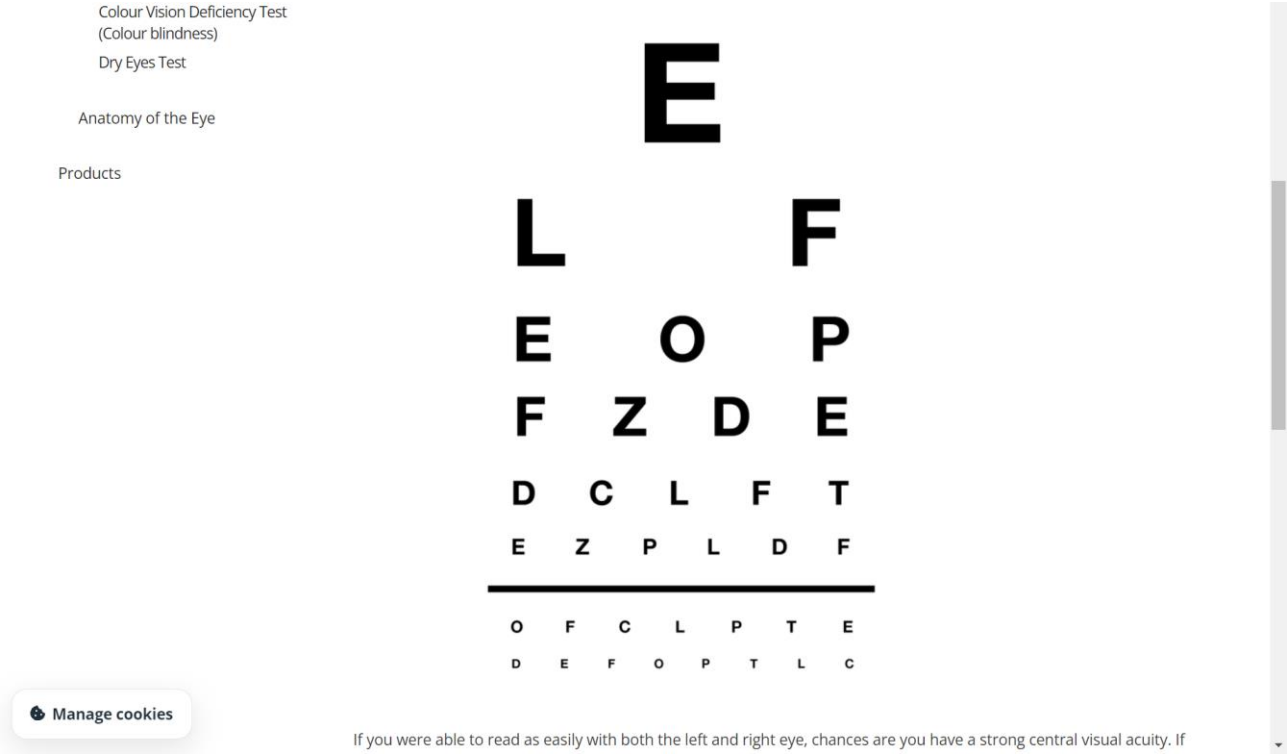


Figure 3

Screenshot of the Online Vision Screening

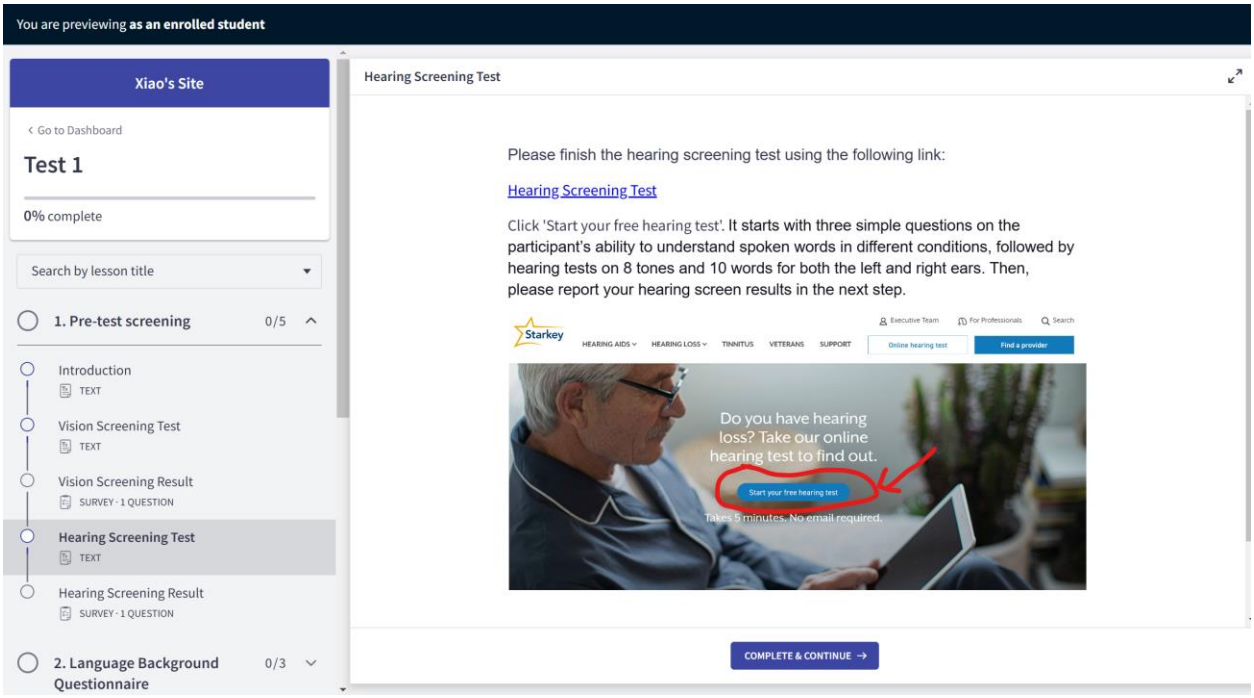


Participants' vision was screened using the online test at https://www.optoplus.com/en/eye-health/virtual-eye-test/online-eye-test/ (Figures 2 & 3). The link and instructions were included in the online course, and participants followed the instructions to finish the vision screening. Those with corrected vision were instructed to wear their glasses during the test. Participants needed to be able to read through lines 1 to 6 without any issues to pass the screening. The vision screen took three minutes to complete. All participants passed the screening.

Hearing Screening

Figure 4

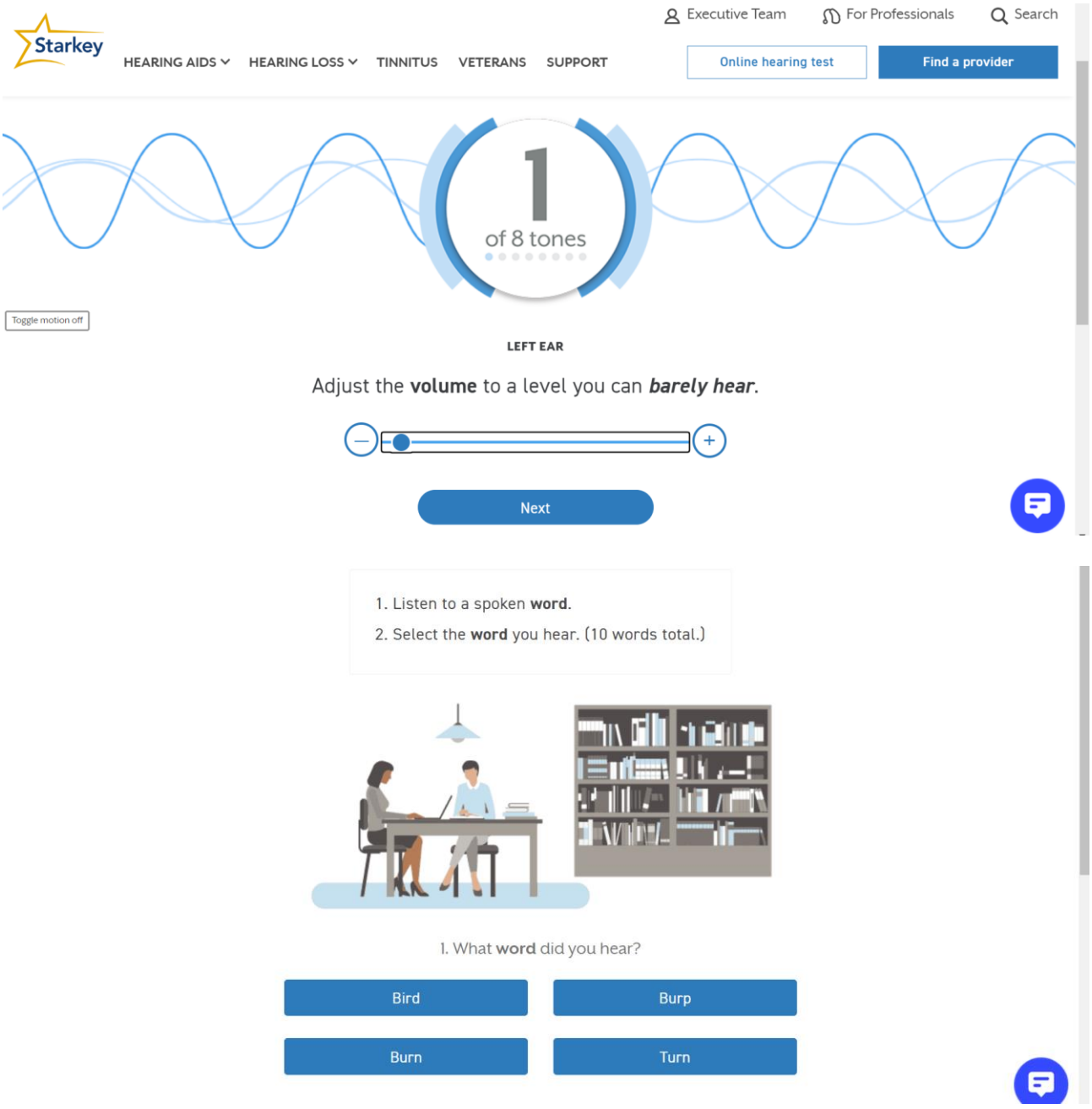
Screenshot of the Hearing Screening Instructions



The hearing test was done using the free online test: https://www.starkey.com/online-hearing-test#!/HearingTestQuestions (Figures 4 & 5). It started with three simple questions on the participant's ability to understand spoken words in different conditions, followed by hearing tests on 8 tones and 10 words for both the left and right ears. The test was done with headphones or earbuds, and it was designed to screen patients with potential hearing loss. The test could be done within five minutes. Participants were asked to not wear hearing aids for the study, and they reported their hearing screen results using the online survey (https://xiao-s-site-6167.thinkific.com/courses/take/test-1/surveys/56028817-hearing-screening-result). Participants were not allowed to continue if they failed the hearing test. Out of 93 effective participants, 90 passed the hearing test and continued with the following speech perception and language comprehension tasks.

Figure 5

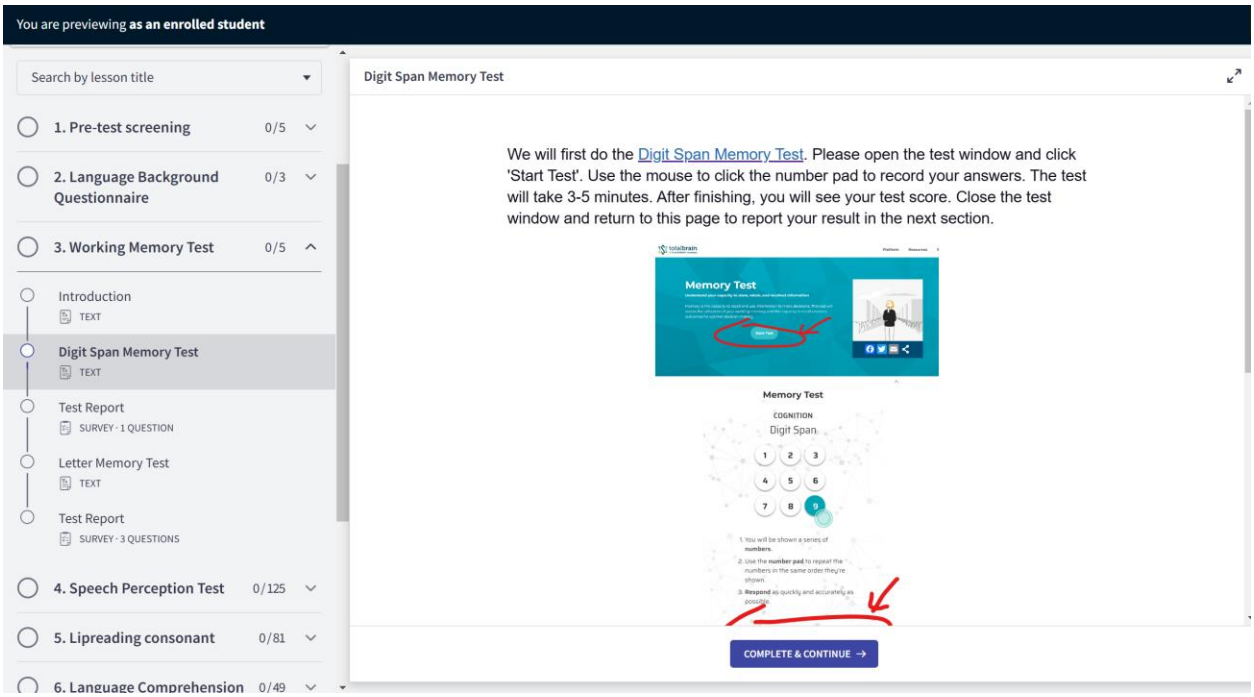
Screenshot of the Online Hearing Screening



Working Memory Testing

Figure 6

Screenshot of the Instructions for Digit Span Memory Test



The study used two types of working memory tests, the digit span memory test, and the letter memory test, to evaluate participants' working memory capacity. These tests are commonly used to assess subjects' cognitive function involved in temporarily holding and manipulating information. The digit span memory test presents participants with a sequence of digits that they must recall in the order they were presented (forward digit span). The length of the sequence gradually increases, challenging the participant's ability to retain and process more information. The primary purpose of this test is to measure numerical working memory and attentional control, with higher scores indicating better capacity for holding and processing numerical information. In contrast, the letter memory test presents participants with a sequence of letters, and participants are instructed to recall the letter strings in the sequence. It primarily assesses the ability to update and maintain verbal information. These tests provide a comprehensive measure of working memory critical for language comprehension and problem-solving.

Participants were instructed to visit the following websites for Digit Span Memory Test (https://www.totalbrain.com/mental-health-assessment/memory-test/) (Figures 6 & 7) and Letter

Memory Test (https://metodorf.com/tests/memory/memorizing_rows_letters.php?method=lettersrows&expos=4000&mod=start#main) (Figures 8 & 9), which assessed their cognitive abilities. Both tests were straightforward and took about five minutes to complete. In the Digit Span Memory Test, participants reported how many numbers they could memorize without making mistakes using the online survey (https://xiao-s-site-6167.thinkific.com/courses/take/test-1/surveys/55307234-test-report). Similarly, participants finished Letter Memory Tests and reported how many letters they could remember correctly. The working memory reports were later analyzed with their speech perception test data. Regardless of their scores, all results were included in the analysis to determine the correlation between working memory and the expression of the McGurk effect.

Figure 7

Screenshot of the Digit Span Memory Test

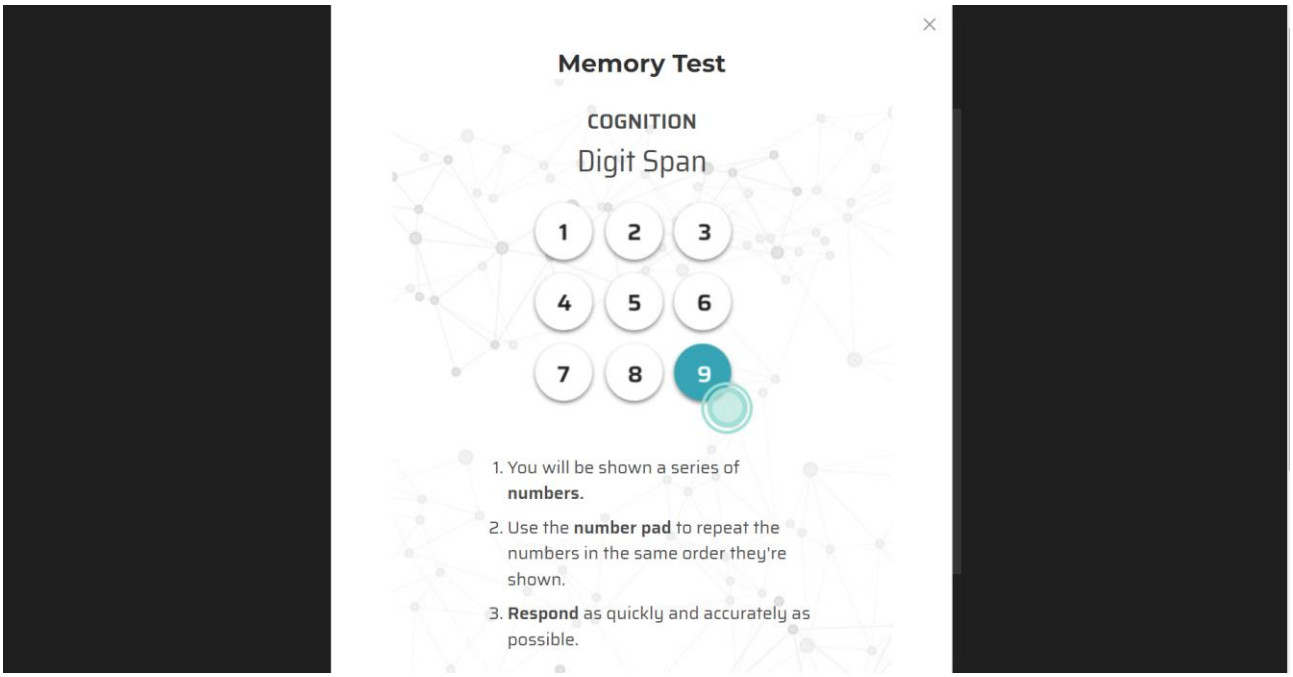


Figure 8

Screenshot of the Instructions for Letter Memory Test

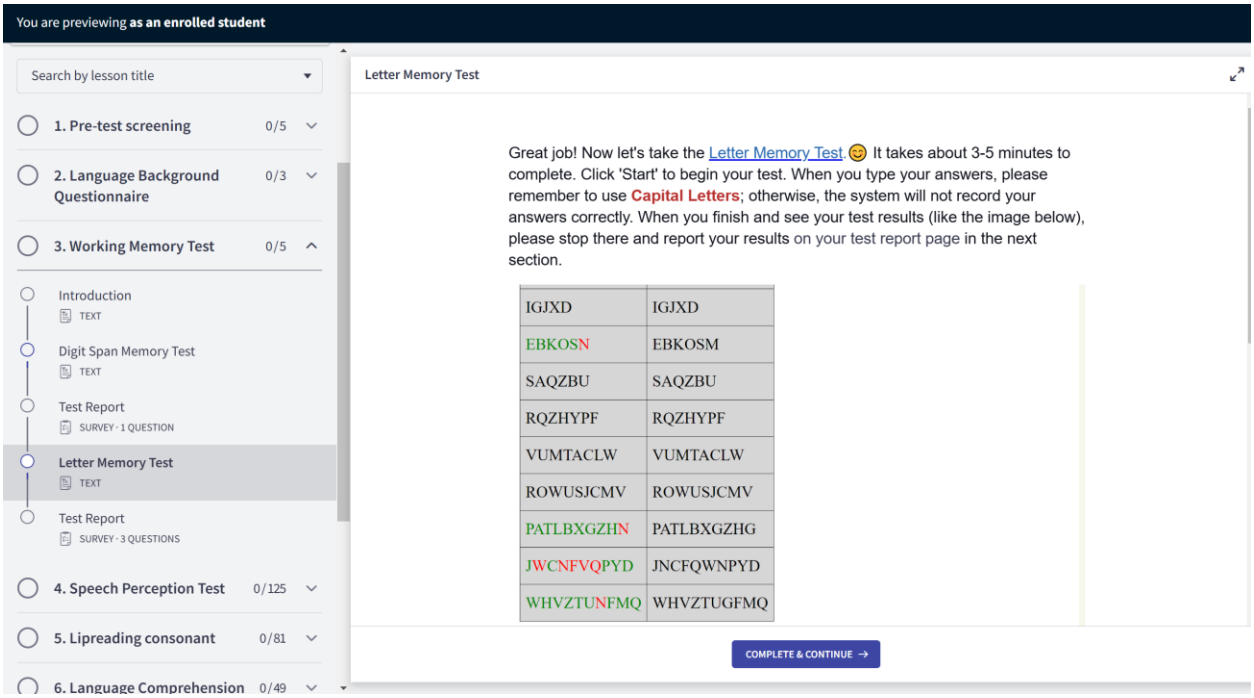
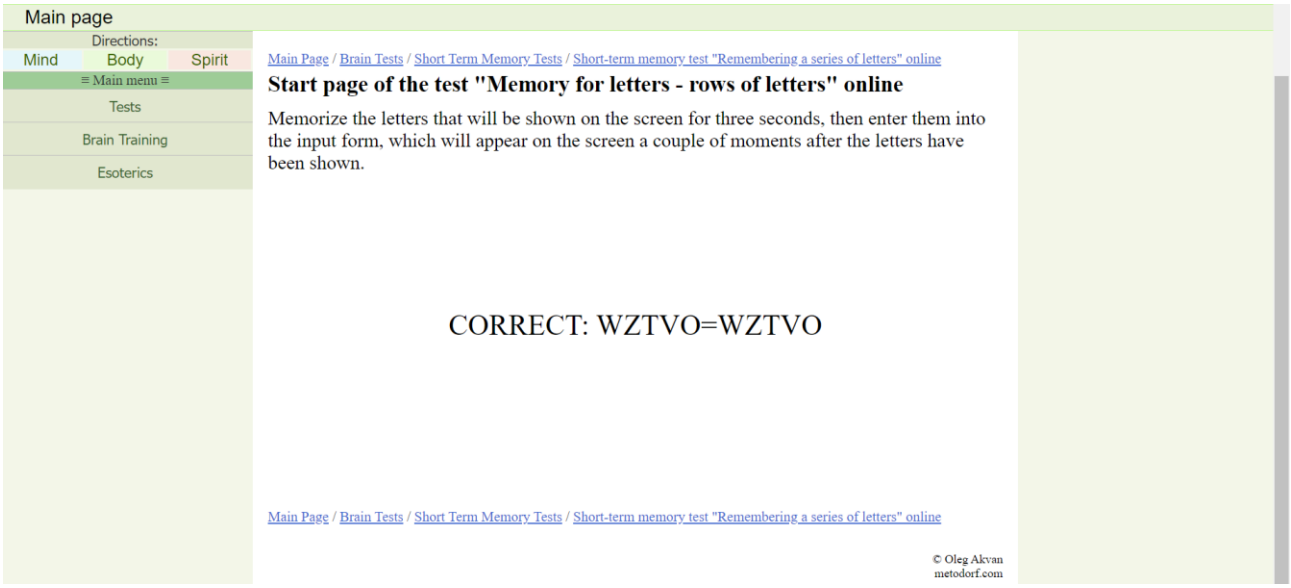


Figure 9

Screenshot of the Letter Memory Test



After successfully completing the screening tasks and working memory tests, participants continued on to three tasks: a McGurk Effect speech perception task, one lipreading tests, and a language comprehension test incorporating items with content visuals, items with context visuals, and audio-only items.

The McGurk Effect (Speech Perception) Tasks

The speech perception task was adapted from Strand et al. (2014) and tested whether working memory and/or lipreading capacity contribute to individual variability of the McGurk effect. Participants completed the following tasks in a fixed order: McGurk identification, and lipreading consonants. Figures 10 and 11 show the screenshots of a speech perception test and its questions.

Production of Speech Perception Tasks

Figure 10

Screenshot of the Speech Perception Task

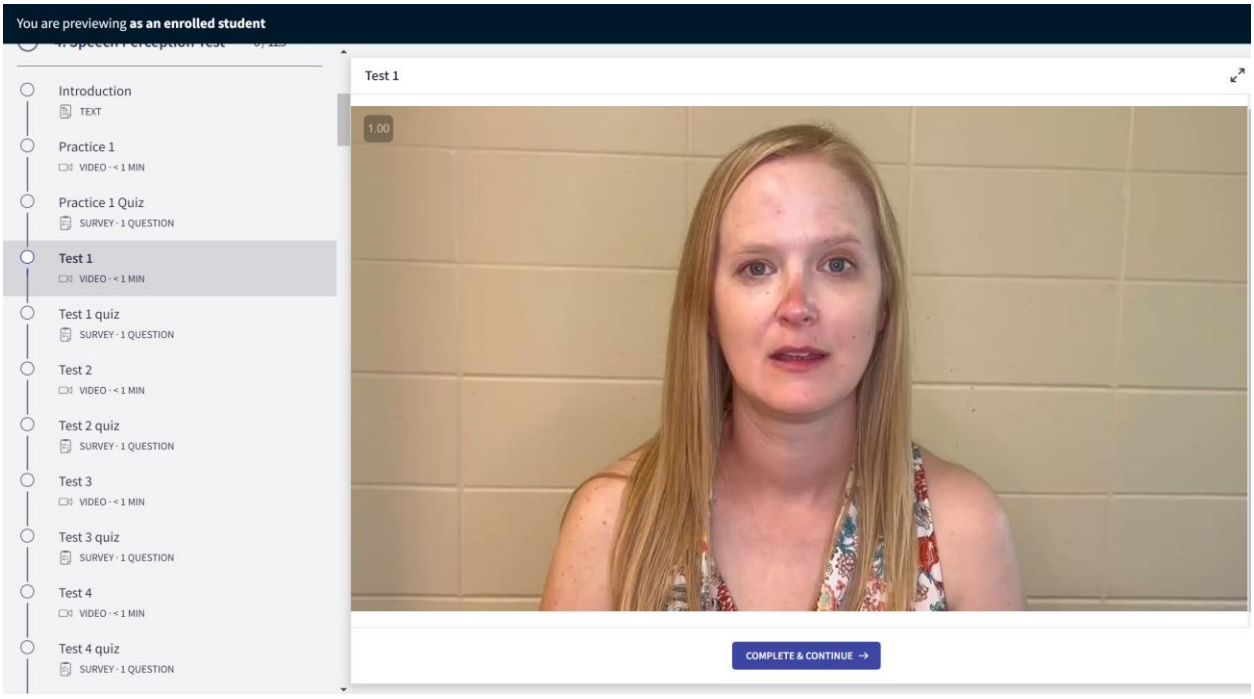
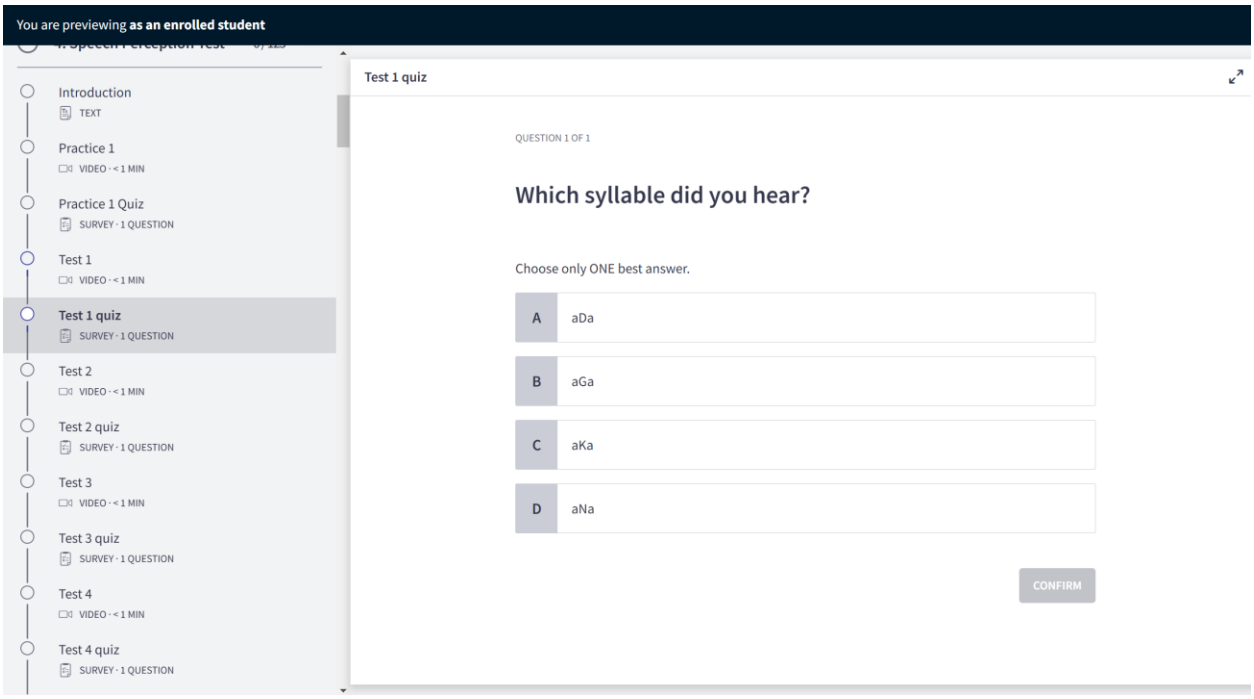


Figure 11

Screenshot of the Quiz for the Speech Perception Task



The stimuli used for the speech perception task were generated following Strand's (2014) method. Briefly, a white female speaker with a standard Southern accent read aloud the test syllables, and their readings were recorded using an iPhone 13 (Apple) and an illuminating iPhone tripod (Amazon). The videos included the speaker's face and shoulders with lip movements clearly visible to participants. The videos with matched lip movement and audio output are congruent stimuli. The free video software Openshot (https://www.openshot.org/) was used to edit congruent videos to generate incongruent videos, where mismatched audio output was aligned with the incorrect lip movement.

Speech Perception Task Administration

The congruent and incongruent videos were mixed and played randomly, and the playback order remained the same for all participants to focus on individual differences and avoid experimental biases. Participants were able to watch the speaker's face for all tasks.

For the McGurk identification task, participants listened to 60 audiovisual (AV) consonant stimuli in an "aCa" context (e.g., /aGa/), including congruent and incongruent AV tokens. The congruent AV tokens were /aBa/, /aDa/, /aFa/, /aGa/, /aKa/, /aMa/, /aNa/, /aPa/, /aTa/, and /aVa/. Forty congruent AV tokens were played, with each of the 10 tokens presented four times by the speaker. The incongruent AV tokens were studied previously and were expected to produce the McGurk effect: AbVf = AVv (audio /aBa/ with visual /aFa/ results in the perception of /aVa/), AbVg = AVd, AmVg = AVn, AmVt = AVn, ApVg = AVk, ApVk = AVt, and AtVb = AVp. Twenty-one incongruent AV tokens were played, with each of the 7 tokens presented three times by the speaker. In total, each participant completed 61 identifications, with 40 congruent AV tokens and 21 McGurk tokens. Participants were then prompted to choose the syllable they thought they heard from 10 options (b, d, f, g, k, m, n, p, t, v). After the completion of data collection, the results were downloaded from the online course website, and the McGurk scores were calculated following the Strand et al. (2014) procedure.

Lipreading Tasks

In the lipreading consonants task, participants watched the same 40 congruent videos (with four repeats for each of the 10 consonants) without audio. Participants were prompted to identify which consonant they heard based on the visual input (Strand et al., 2014). The lipreading capacity was calculated by dividing the correct responses by the total number of questions they answered (should be 40). As described by Strand *et al.* (2014), a correct response includes all possible

consonants that share the same Place of Articulation (POA). The study followed the POA grouping described by Strand *et al.* (2014): {b, m, p} {f, v} {d, n, t} {k, g}.

Discourse-level Language Comprehension Tasks

The discourse-level language comprehension tasks included simple, intermediate, and complex passages (see Appendix B). The simple and intermediate talks were sourced from Dupuis (2011), and the complex passages were kindly provided by Kamiya (2022). The primary objective of the language comprehension task was to assess whether visual cues, such as images related to the content of the testing materials, enhance language comprehension in L2 English learners. The study focused on two independent variables: the modality of presentation (audio-only vs. content/context visual) and the syntactic complexity of the content (simple vs. intermediate vs. complex). This resulted in six passage types:

- Audio Simple (10 passages)
- Audio Intermediate (10 passages)
- Audio Complex (2 passages)
- Content/Context Simple (10 passages)
- Content/Context Intermediate (10 passages)
- Content/Context Complex (2 passages)

Each participant was tested under all six conditions, including 22 audio-only passages (10 simple, 10 intermediate, and 2 complex) and 22 audiovisual passages (10 simple, 10 intermediate, and 2 complex). To ensure the participants only heard each passage once, I divided the test into 4 groups, as shown in the table below (Table 3). The audiovisual passages contained half context and half content visuals, which were assigned randomly among the test groups.

Table 3

Assignment of Simple, Intermediate, and Complex Listening Tasks to Individual Participant Groups

Group	1	2	3	4	AV
simple	1-10	1-10	11-20	11-20	Audio
medium	1-10	1-10	11-20	11-20	
complex	2,3	2,3	4,5	4,5	
simple	11-20	-	1-10	-	Visual Context
medium	-	11-20	-	1-10	
complex	4,5	-	2,3	-	
simple	-	11-20	-	1-10	Visual Content
medium	11-20	-	1-10	-	
complex	-	4,5	-	2,3	

Note. The listening materials are in the format of audio-only, audios with context visuals, and audio with content visuals.

In total, each participant finished 44 tasks. After each task, they were prompted to answer two questions for simple and intermediate tasks, and six questions for complex tasks. The ratio of correct responses was calculated to evaluate their level of understanding of the task materials.

Production of Content and Context Visuals

The task materials are attached in Appendix B. The simple and medium passages were adapted from the master's study of Dupuis (2011) at the University of Calgary, and the complex passages were kindly provided by Dr. Kamiya based on his study in 2022. These passages were used to create 44 discourse-level language comprehension tasks. Three variations were created for each passage: content visual, context visual, and audio only. The content visuals were generated by selecting images that match the content of the text (Figure 13). The context visual, by contrast, used an image of a classroom to reflect the teacher's background (Figure 12). The images for the content visuals were selected from free image libraries, which contain thousands of images (e.g., https://pixabay.com/; https://www.freepik.com/). These images cover various topics such as

sports, economy, education, and medication. I searched for appropriate images using keywords and used them as content visual cues. The videos were generated by artificial intelligence (AI)-assisted platform Vidnoz (https://www.vidnoz.com/) using the text-to-speech function. The AI platform features life-like voice generation, and the sound quality is indistinguishable from authentic human voice. I used Vidnoz to generate content visuals and context visuals for online test. The audio was extracted from the videos using the free software Audacity (https://www.audacityteam.org/), and the audio files were used as the audio-only test materials.

Visual Task Administration

After the presentation of each passage, participants were immediately prompted to finish two multiple-choice questions (Figure 14). The test results were downloaded from the course website for subsequent analysis.

Figure 12

Screenshot of the Listening Task with Context Visual Cues

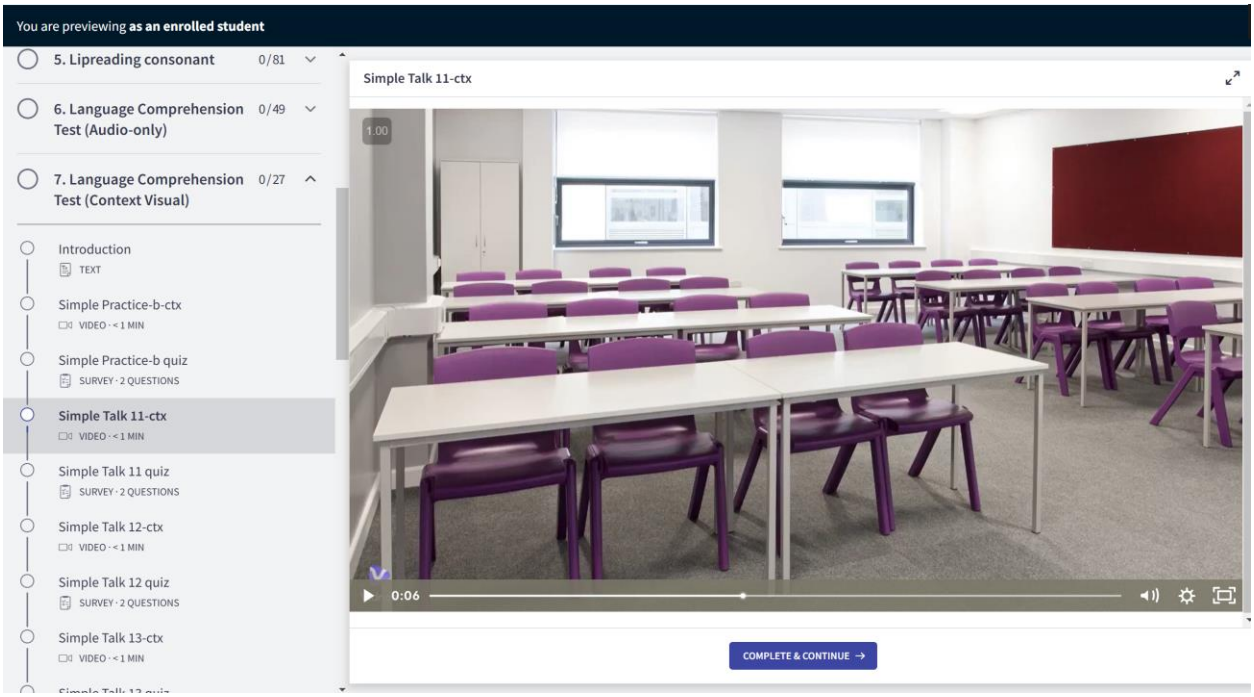


Figure 13

Screenshot of the Listening Task with Content Visual Cues

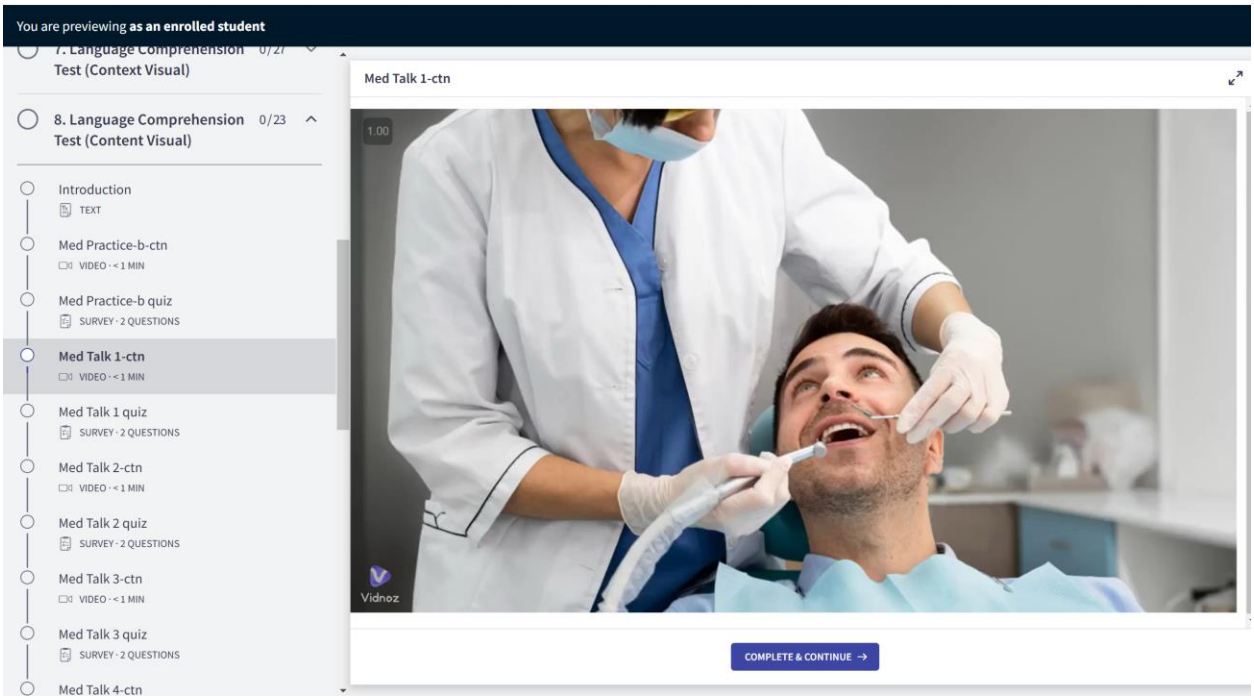
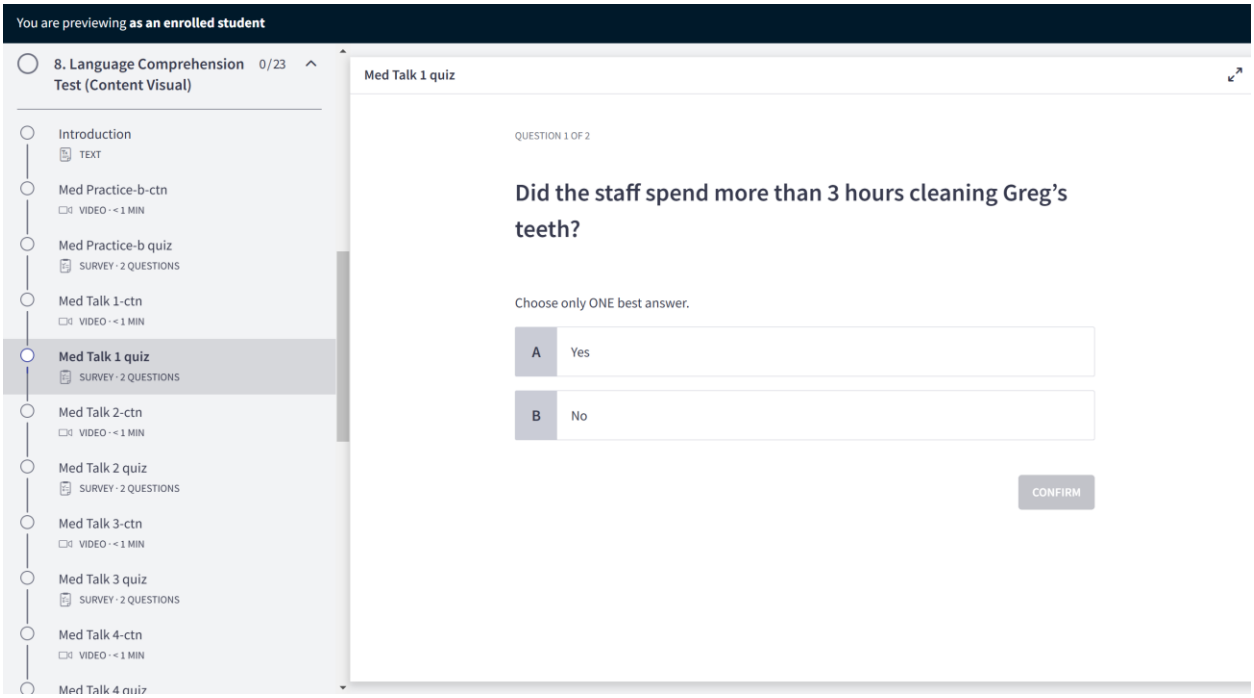


Figure 14

Screenshot of the Quiz for the Audiovisual Listening Task



Statistical Analysis

The statistical analysis for this study consisted of two primary tasks: speech perception and language comprehension. For each task, appropriate statistical tests were used to determine the relationship between visual cues and other relevant factors. Dependent variables include participants' performance on the two primary tasks: speech perception and language comprehension. Independent variables include working memory, lipreading capacity, and language task complexity.

Speech Perception Task

The speech perception task focused on the individual variability in the McGurk effect and its potential correlation with cognitive factors, namely working memory and lipreading capacity.

McGurk Effect and Working Memory. Pearson's correlation coefficient (r) was used to determine whether variability in the McGurk effect was significantly correlated with individual differences in working memory capacity. Participants' working memory was measured using the digit span and letter memory test, and McGurk effect susceptibility was quantified as the percentage of trials in which participants reported illusory perception during mismatched audiovisual stimuli.

McGurk Effect and Lipreading Capacity. Similarly, Pearson's correlation coefficient (r) was used to analyze whether there was a significant correlation between McGurk effect variability and lipreading skills, as assessed by a standardized lipreading test. The goal was to examine whether individuals who demonstrated greater lipreading proficiency were more or less susceptible to the McGurk effect.

Language Comprehension Task

For the language comprehension task, the non-parametric Kruskal-Wallis test was used to analyze whether content and context visual cues influenced participants' listening comprehension across different levels of task difficulty. This test was selected due to its ability to compare the distribution of scores across multiple independent groups (audio, context visual, content visual), especially when the data do not meet the normality assumption required for parametric tests like ANOVA.

Task Difficulty Levels. The comprehension tasks were divided into three levels—simple, intermediate, and complex—based on sentence length, syntactic structure, and vocabulary complexity. Participants were exposed to content (e.g., images providing salient information related to the audio) and context (images showing a classroom) visual cues during these tasks, and their comprehension scores were recorded.

Kruskal-Wallis Test. The significance level for the analysis was set at $p < 0.05$, and outliers were included in the analysis. The effect size was measured using an offset of 0.3 to gauge the practical significance of the findings.

Multiple Comparisons. Once the Kruskal-Wallis test indicated significant differences between the groups, post-hoc multiple comparisons were conducted to identify which specific groups (audio, context visual, and content visual) were different from one another. Dunn's test was chosen for this purpose, as it considers the total number of groups (k), even when comparing only two groups. This method allowed for a more accurate comparison between audio-only and audiovisual groups while adjusting for the number of pairwise comparisons made.

Correction Method. A Bonferroni correction was applied to account for the increased risk of Type I error associated with multiple comparisons. This adjustment ensured that the significance level remained at 0.05 after accounting for the number of tests conducted.

The results from Dunn's test were further verified using the Mann-Whitney U test, which supported the findings from the Kruskal-Wallis test when comparing two groups at a time.

All statistical analyses were conducted using R, ensuring robustness and rigor in determining the impact of visual cues on L2 learners' listening comprehension. The graphs were generated using the online statistical tool Statistics Kingdom (https://www.statskingdom.com/index.html).

Hypotheses

This study posits three key hypotheses based on the research questions:

1. Individual Variability in the McGurk Effect: It was hypothesized that individual differences in the McGurk effect are positively correlated with lipreading capacity. This suggests that participants with stronger lipreading abilities are more likely to experience the McGurk effect. However, the role of working memory in this variability remains uncertain; it may or may not significantly influence the individual expression of the effect.
2. Impact of Content and Context Visuals on Language Comprehension: It was anticipated that content visual cues would enhance language comprehension in L2 learners compared to audio-only input. These visuals provide salient information related to the spoken content, thus aiding understanding. In contrast, context visuals, which merely display the speaker's background environment, are not expected to improve comprehension significantly.
3. Effectiveness of Content Visuals Based on Task Complexity: The hypothesis further predicted that the beneficial impact of content visuals will be most pronounced in tasks of

intermediate difficulty. For simple tasks, where minimal cognitive load is required, visual aids may not be necessary. Conversely, in highly complex tasks, the cognitive demands may exceed participants' ability to process both auditory and visual information, making the content visuals less effective.

Conclusion

In this chapter, I have outlined the research methods used to investigate the role of visual cues in speech perception and language comprehension among L2 English learners. The study comprises three key tasks: a speech perception task that examines the individual variability in the McGurk effect, working memory tests that assess participants' working memory capacity, a lipreading task that examines the participants' lipreading capacity, and a language comprehension task that assesses the impact of content and context visual cues across varying levels of task complexity.

The statistical analyses, including Pearson's correlation for the speech perception task and the Kruskal-Wallis Test with Dunn's multiple comparisons for the language comprehension task, provide the methodological framework to address the research questions. By using these well-established quantitative methods, the study aims to offer robust insights into the influence of visual cues in L2 speech perception and listening comprehension.

The subsequent chapter will focus on presenting the results of these analyses, offering a detailed examination of the data concerning the hypotheses outlined earlier. Through these results, I seek to further our understanding of how multimodal inputs, such as visual cues, can assist L2 learners in overcoming challenges in speech perception and language comprehension.

CHAPTER 4

RESULTS

This chapter presents the findings of the study, which examined the impact of visual cues on speech perception and language comprehension among L2 English learners. The results are organized around the research questions introduced in Chapter 1 and the hypotheses articulated in the previous chapter. First, the speech perception task data are analyzed, focusing on the McGurk effect and its correlation with working memory and lipreading capacity. Then, the findings from the language comprehension task are discussed, highlighting the effects of content and context visuals across simple, intermediate, and complex language tasks. Each section will provide a detailed presentation of the results, followed by interpretations that align with the research objectives. Through this chapter, I aim to provide a comprehensive overview of the data, addressing the key research questions and offering insights into the role of visual cues in L2 learning.

Participant Characteristics

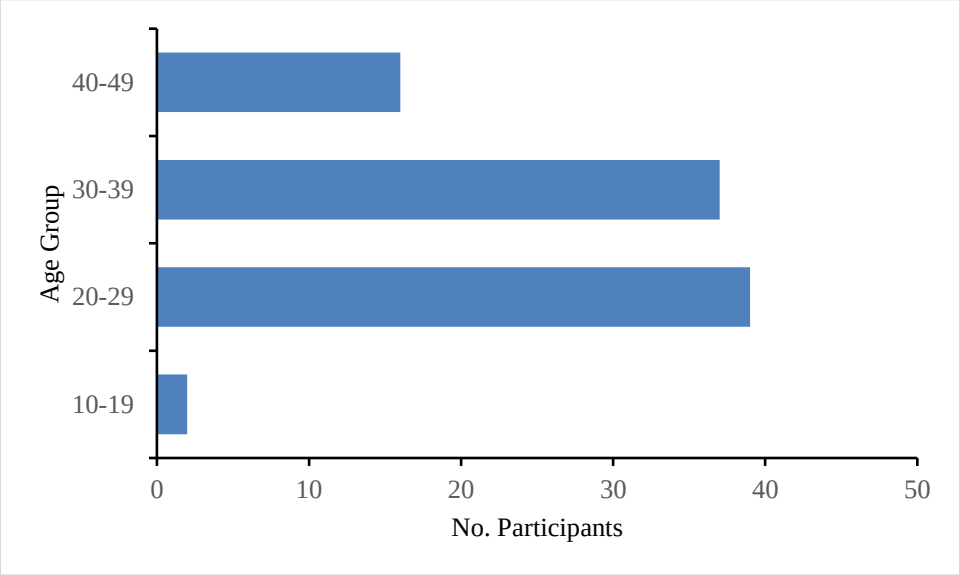
Age Distribution

As stated previously, most (*n*=90) of the recruited participants are Brazilians, and I decided to focus on this single group for all data analyses to minimize inconsistency introduced by cultural differences. Among all Brazilian participants, 63 (70%) are females, and 27 (30%) are males. The age distribution of these participants spans across four distinct age groups (Figure 15). Most participants fall within the 20-29 age group, accounting for 37 individuals, which makes up 41% of the participants. This is followed closely by the 30-39 age group, with 36 participants (40%).

The 40-49 age group includes 15 individuals (17%), while the youngest cohort, aged 10-19, has 2 participants (2%). Overall, the participant pool is predominantly younger adults.

Figure 15

Age Distribution of Participants



Occupations

The initial aim of the study was to recruit on-campus students. However, when I had trouble recruiting enough participants, I expanded my recruitment to any L2 English learners who have standardized English test scores. To better understand their backgrounds, I decided to ask for their occupations. However, because the background questionnaire did not include occupations in the early phase of data collection, I could only collect information on occupations for 67 out of the 90 participants. Briefly, the participants represent a diverse range of professional and academic backgrounds. Most (*n*=36, 53.7%) participants are engaged in higher education, either as graduate or undergraduate students, across various fields such as law, biological sciences, engineering, and the humanities. In addition, 12 (17.9%) of the participants are teachers and professors working in educational institutes. Besides students and educators, 19 (28.4%) participants are professionals

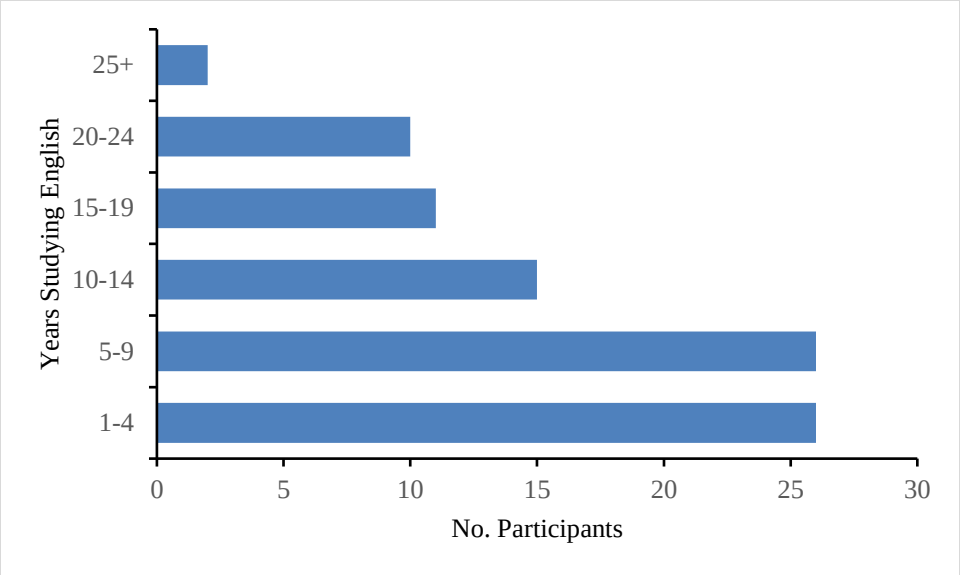
working as psychologists, nurses, teachers, civil engineers, and data scientists, many of whom hold advanced degrees such as master's or doctoral qualifications. Overall, the occupational distribution highlights a strong academic presence, with over half of the participants being students, followed by teaching, healthcare, and engineering professionals.

English Proficiency

The self-reported English proficiency levels of the Brazilian participants vary wildly in terms of years spent studying the language, and participants reported that the study duration was as short as 1 year to as long as 30+ years (Figure 16). It is striking that 38 (42.2%) participants reported studying English for over 10 years. Among those seasoned English learners, 15 (16.7%) have studied for 10-14 years, 11 (12.2%) have studied for 15-19 years, 10 (11.1%) have studied for 20-24 years, and 2 (2.2%) have studied over 25 years. In contrast, 26 (28.9%) participants have English learning experience between 5 and 9 years, and the remaining 26 (28.9%) participants have been studying English for shorter durations of 1-4 years. This distribution reflects varying levels of English language experience, suggesting the presence of both novice learners and more advanced, long-term English learners within the group. Overall, this diversity in proficiency could influence their performance in tasks requiring language comprehension.

Figure 16

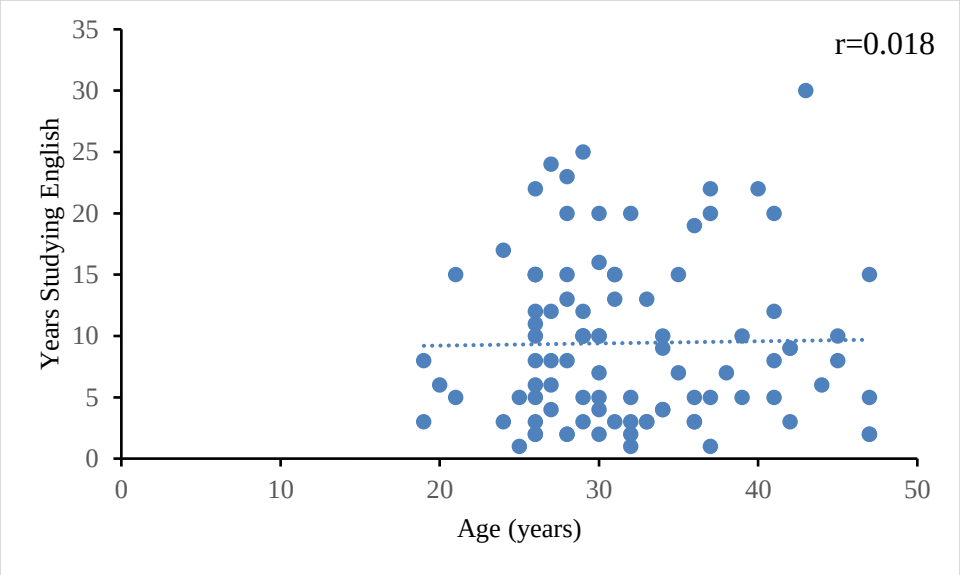
Distribution of Participants Based on Years Spent Studying English



I further plotted the age of the participants against their years spent in English learning and asked if there was a correlation. As shown in the scatter plot, there is no clear correlation between the two (Figure 17). The calculation of Pearson's correlation coefficient resulted in r = 0.018 (P = 0.864), indicating a very weak negative correlation between the two variables that is not significant. I concluded that the participants' ages and the number of years they have studied English do not show a meaningful linear relationship. The negative sign suggests that the years spent studying English slightly decrease as age increases, but this trend is not statistically significant.

Figure 17

Scatter Plot of Participants' English Study Experience Versus Their Age

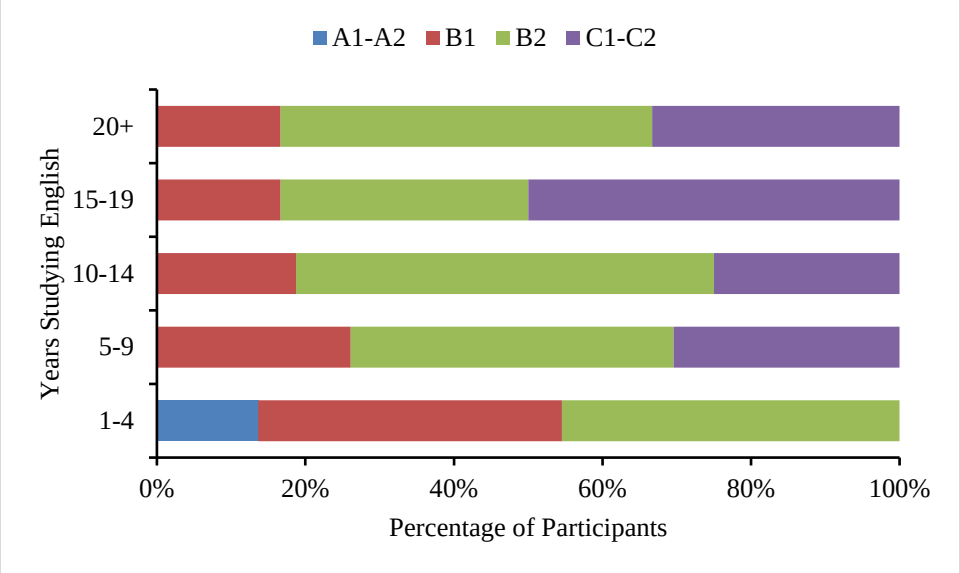


Next, I examined the distribution of English learners' proficiency levels as a function of years spent studying English. Fifty-three out of ninety participants had standardized English test scores in TOEFL, IELTS, or Duolingo, which were converted into Common European Framework of Reference for Languages (CEFR) levels using the conversion charts (https://theedge.com.hk/conversion-table-for-toefl-ibt-pbt-cbt-tests/, https://englishtest.duolingo.com/scores). The remaining participants reported their proficiency level based on their Brazilian college English exams or other non-standardized English tests. For those with 1 to 4 years of English study, over half (54.5%) are in the low proficiency tiers, being either A1-A2 (beginner) or B1 (intermediate) on the Common European Framework scale. The remaining is at B2 (upper-intermediate). None of these learners have achieved C1-C2 (advanced) proficiency yet. As the years of study increase to 5 to 9 years, the distribution shifts to a more balanced spread across proficiency levels, with 26.1% at B1, 43.5% at B2, and 30.4% at C1-C2, while none at A1-A2. The participant groups with over 10 years of English studying experience show a further decrease in the lower English proficiency level B1, from 18.8% in 10-14 years to 16.7% in 20+ years, and an increase of the combined B2 and C1-C2 learners from 81.2% to 83.3%

(Figure 18). This indicates that longer study duration tends to result in higher English proficiency. Nevertheless, a few learners still report lower levels even after many years of study.

Figure 18

Distribution of Participants Across Varying Levels of English Proficiency Based on Years Spent Studying English



Working Memory Tests

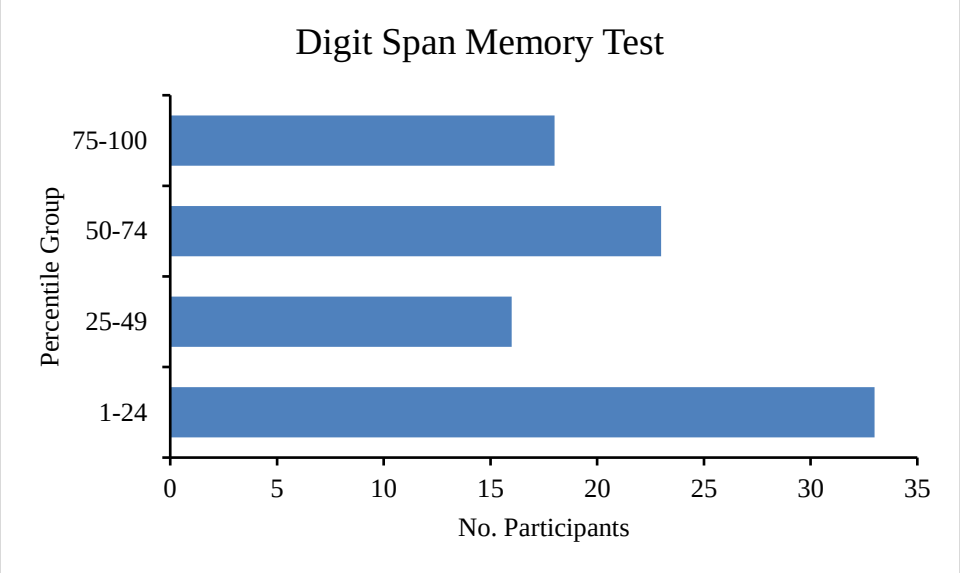
Digit Span Memory Test

The results of the digit span memory test show a diverse distribution of working memory abilities across participants. Thirty-three (36.7%) participants scored between the 1st and 24th percentiles, indicating low working memory capacity. Sixteen (17.8%) scored between the 25th and 49th percentiles, representing a range of below-average to moderately low working memory performance. Twenty-three (25.6%) participants fall between the 50th and 74th percentiles, suggesting above-average working memory performance. The remaining 18 (20%) participants scored between the 75th and 100th percentiles, highlighting individuals with good to superior working memory abilities (Figure 19). These results indicate significant variability in working

memory. Scores did not fit a normal distribution, with a roughly even distribution across the lower, middle, and upper percentiles, except for the 1st to 24th percentiles, which contain a bigger proportion of participants.

Figure 19

Distribution of Participants Based on Digit Span Memory Percentile



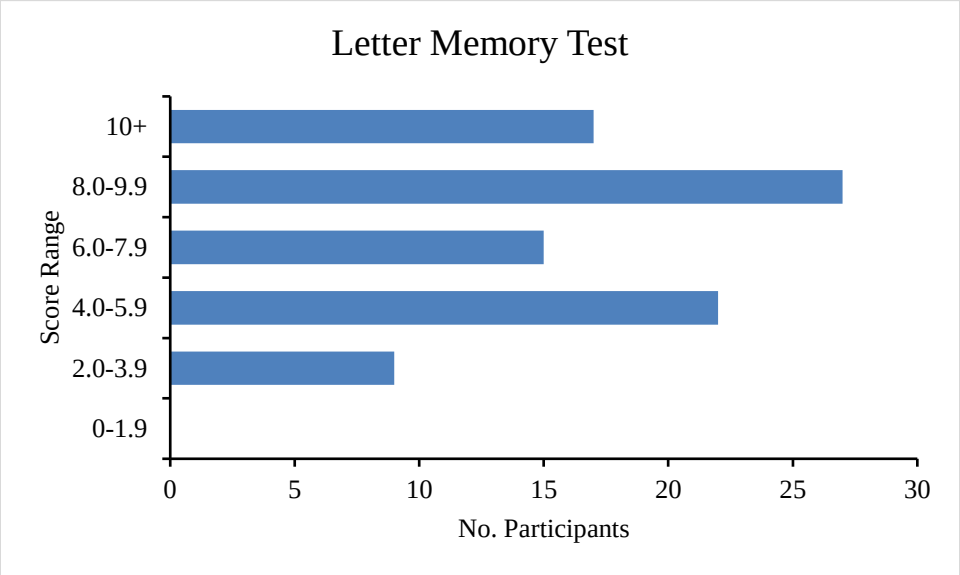
Letter Memory Test

The letter memory test results were calculated based on three scores: 1) the longest letter strings remembered, which counts the number of letters in the longest letter string recalled correctly; 2) No. letter strings remembered, which counts how many letter strings were recalled correctly; and 3) No. mistakes made, which counts how many letter strings were recalled incorrectly. The overall letter memory score was derived using the following equation: Score = Longest letter strings remembered + 0.2 × No. letter strings remembered correctly – 0.1 × No. mistakes made. The letter memory scores are grouped into six ranges (Figure 20). No participants scored below 2.0, and 9 participants (10%) scored between 2.0 and 3.9. Twenty-two participants (24.4%) scored between 4.0 and 5.9, which is below average, and fifteen participants (16.7%)

scored between 6.0 and 7.9. Most participants, specifically 27 out of 90 (30%), scored between 8.0 and 9.9, showing strong performance on the test. The remaining 17 participants (18.9%) scored 10 or higher, and they are among the strongest performers. Overall, the distribution was not normal. A substantial portion of the participants performed well, with 47% scoring 8.0 or higher, indicating a strong ability to update and maintain verbal information in this test.

Figure 20

Distribution of Participants Based on Letter Memory Score



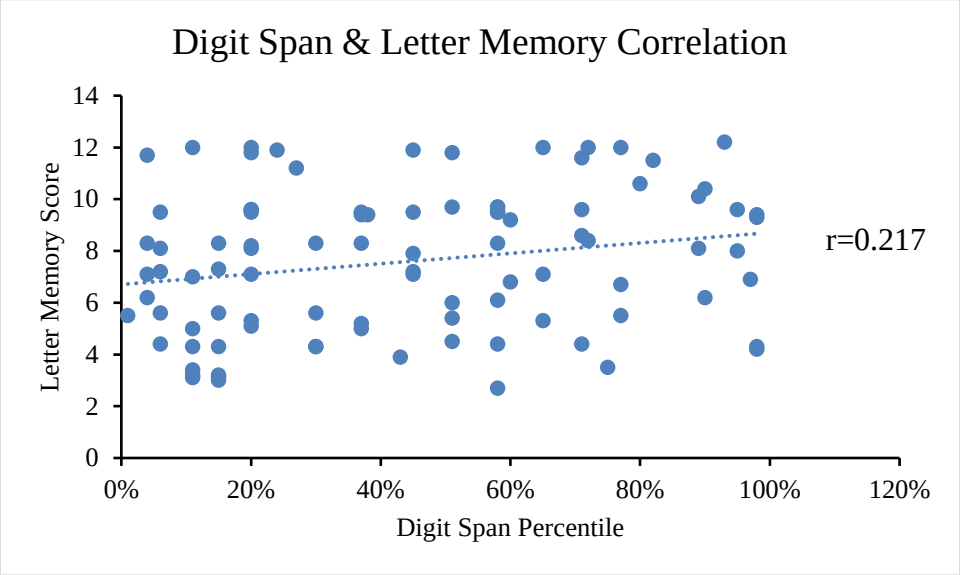
Correlation Between the Two Memories

Next, I asked if the digit span memory shows a correlation with the letter memory. Figure 21 shows the scatter plot between the two memory test scores. Pearson's correlation coefficient between the digit span memory percentile and the letter memory score is r = 0.217 (P = 0.036), indicating a weak yet significant positive correlation. This suggests that participants with higher digit span memory percentiles tend to have slightly higher letter memory scores. Because the correlation is not strong, it implies that digit span and letter memories likely tap into slightly different cognitive processes. Alternatively, they may reflect distinct aspects of working memory.

Therefore, despite a modest correlation, the results suggest that performance in one test does not strongly predict performance in the other.

Figure 21

Pearson's Correlation Analysis of Digit Span and Letter Memory



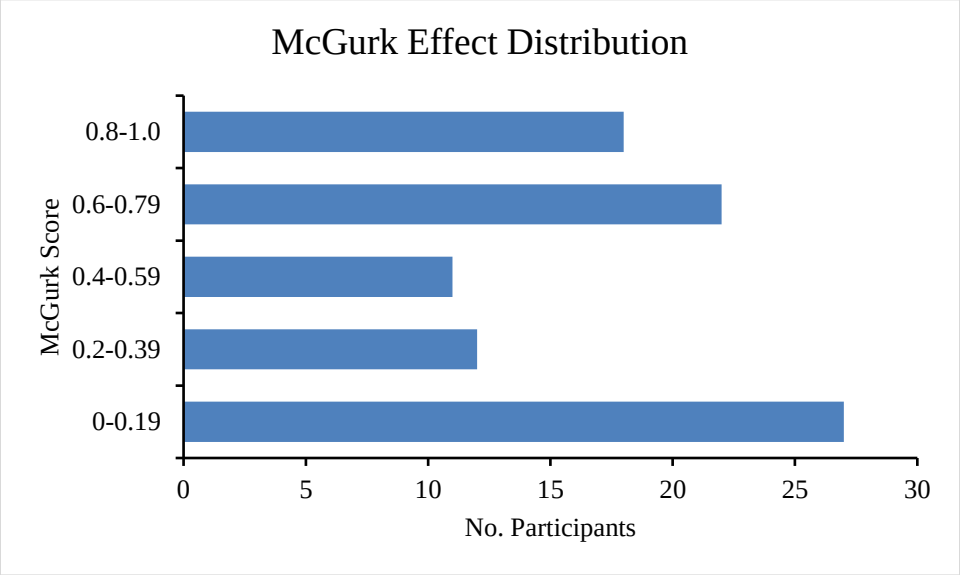
Speech Perception and McGurk Effects

Next, I evaluated how the participants experienced the McGurk effects differently. To calculate the McGurk score, the number of McGurk responses (in which the mismatched auditory and visual signals resulted in an illusionary perception of a consonant-vowel syllable) was divided by the total number of responses. The participants showed a wide range of individual variability in how audiovisual information is integrated (Figure 22). Of the 90 participants, 27 (30%) exhibited a low McGurk effect score between 0 and 0.19, and another 12 (13%) scored between 0.2 and 0.39, indicating low susceptibility to the illusion. Eleven participants (12%) scored between 0.4 and 0.59 and fell within the mid-range. Twenty-two participants (24%) scored between 0.6 and 0.79, while the remaining 18 participants (20%) scored between 0.8 and 1.0,

showing a strong McGurk effect. This distribution suggests a significant variation among participants in how they process and integrate auditory and visual speech information.

Figure 22

Distribution of Participants Based on the McGurk Effect Scores

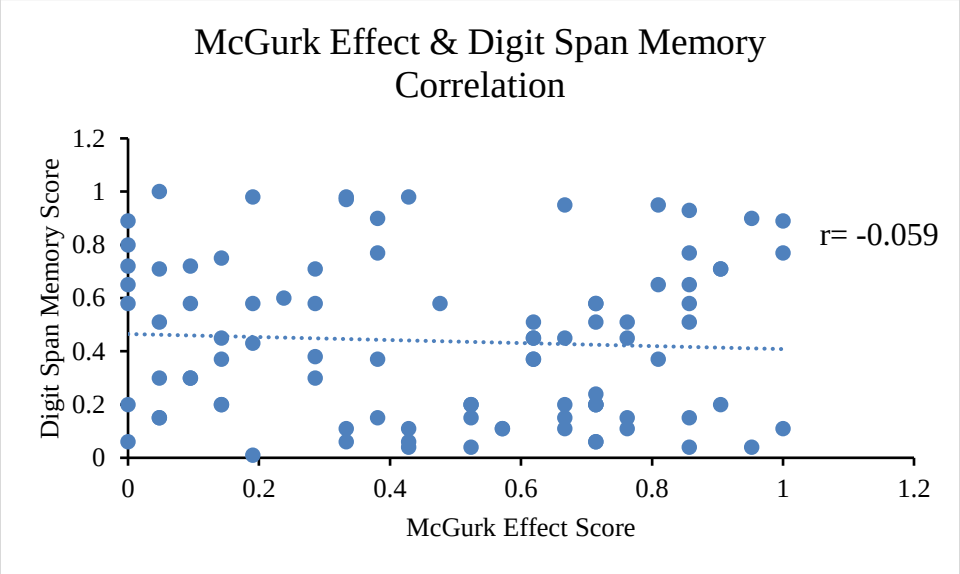


Correlation Between Working Memory and the McGurk Effect

Next, I asked if the participants' susceptibility to the McGurk effect is correlated to their working memory. The correlation analysis between participants' susceptibility to the McGurk effect and their digit span memory percentile yielded a Pearson's coefficient of r = -0.059 (P = 0.577) (Figure 23). This result indicates a very weak negative correlation that is not significant, suggesting that there is little to no meaningful relationship between individuals' digit span memory capacity and their susceptibility to the McGurk effect. In other words, participants' ability to recall sequences of digits (a measure of working memory) does not appear to influence how they integrate audiovisual information in speech perception significantly. Rather, the weak negative correlation suggests other cognitive or sensory factors might contribute to variability in audiovisual speech integration.

Figure 23

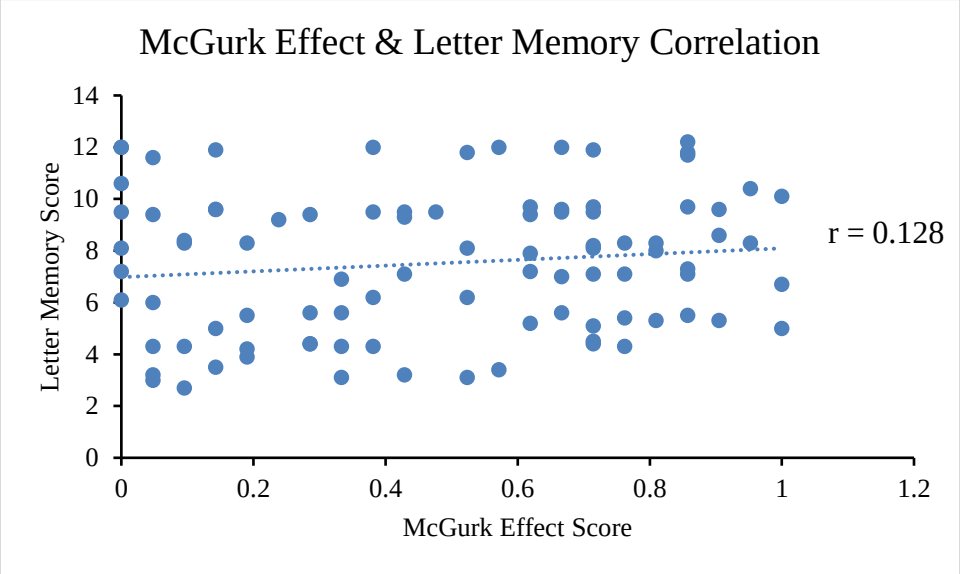
Pearson's Correlation Analysis of the McGurk Effect and Digit Span Memory



I then examined the correlation between participants' susceptibility to the McGurk effect and their letter memory scores and derived a Pearson's coefficient of r = 0.128 (P = 0.228) (Figure 24). This indicates a weak positive correlation that is not significant, suggesting a slight relationship between participants' letter memory capacity and their susceptibility to the McGurk effect. While the correlation is not significant, it hints that individuals with better letter memory performance may be slightly more likely to experience the McGurk effect. This finding suggests that the cognitive processes involved in updating and manipulating information in working memory, as assessed by the letter memory task, might play a minor role in audiovisual speech integration. However, the weak correlation indicates that other factors, such as left/right brain preferences, sensory processing, attention, or specific linguistic skills, might have a greater influence on the variability in susceptibility to the McGurk effect. Further research could explore these additional cognitive dimensions to gain deeper insight into how different memory systems interact with audiovisual speech perception.

Figure 24

Pearson's Correlation Analysis of the McGurk Effect and Letter Memory



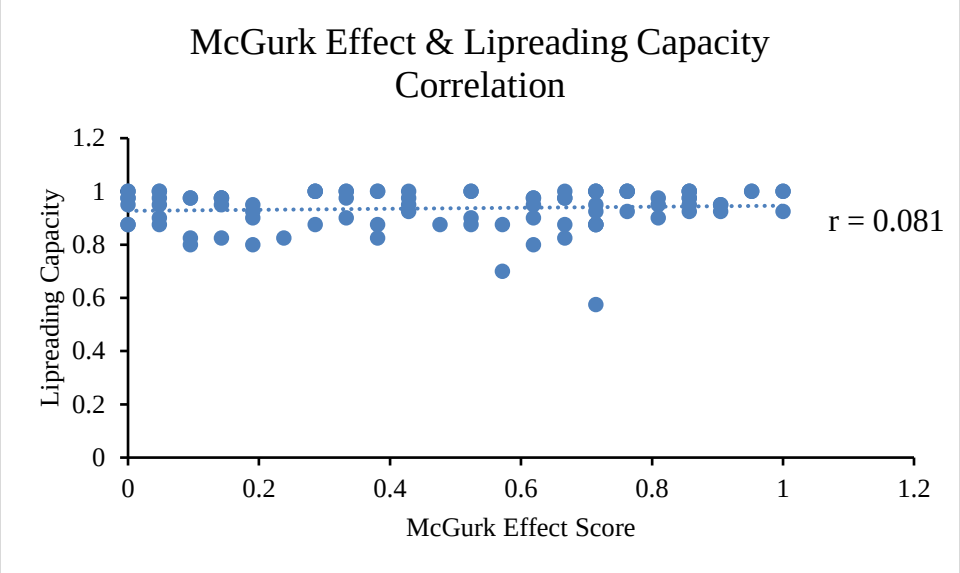
Correlation Between Lipreading Capacity and the McGurk Effect

I then examined whether the participants' McGurk effect is correlated with their lipreading capacity. The two have a Pearson's coefficient of r = 0.081 (P = 0.447), indicating a very weak positive relationship that is not significant (Figure 25). Such a weak correlation suggests a minimal connection between an individual's ability to lipread and their susceptibility to the McGurk effect. In other words, those with better lipreading capacity might show slightly higher McGurk effect scores, but the relationship is not significant. The weak correlation implies that while lipreading and audiovisual speech perception involve integrating visual information with auditory cues, the ability to accurately read lips may not significantly predict how susceptible someone is to the McGurk effect. This makes sense, as good lipreading is required, but may not be sufficient, for the expression of the McGurk effect. This could also suggest that the McGurk effect taps into broader cognitive or perceptual processes beyond just the skill of lipreading, such as how visual and auditory inputs are weighted in speech perception. It may also indicate that other factors, like

individual differences in multisensory integration or attention, play a larger role in explaining variability in experiencing the McGurk effect. Further investigation could explore these possibilities and how different visual and auditory skills contribute to speech perception.

Figure 25

Pearson's Correlation Analysis of the McGurk Effect and Lipreading Capacity



Next, I examined whether working memories show a correlation with lipreading capacity. For digit span memory, Pearson's correlation analysis shows no significant correlation with the lipreading capacity, with a Pearson coefficient of r=0.058 and a p-value of 0.5820 (Figure 26). This indicates that digit span memory has little to no meaningful relationship with lipreading performance. In contrast, letter memory, which involves the recall of letters, shows a moderate and statistically significant positive correlation with lipreading capacity, with a Pearson coefficient of r=0.309 and a p-value of 0.0028 (Figure 27). This suggests that individuals with stronger letter memory tend to perform better in lipreading. Overall, these findings imply that while lipreading capacity does not correlate with the McGurk effect score, it is moderately correlated with letter

memory. The findings support that lipreading and the McGurk effect are distinct processes involving different cognitive functions.

Figure 26

Pearson's Correlation Analysis of Digit Span Memory and Lipreading Capacity

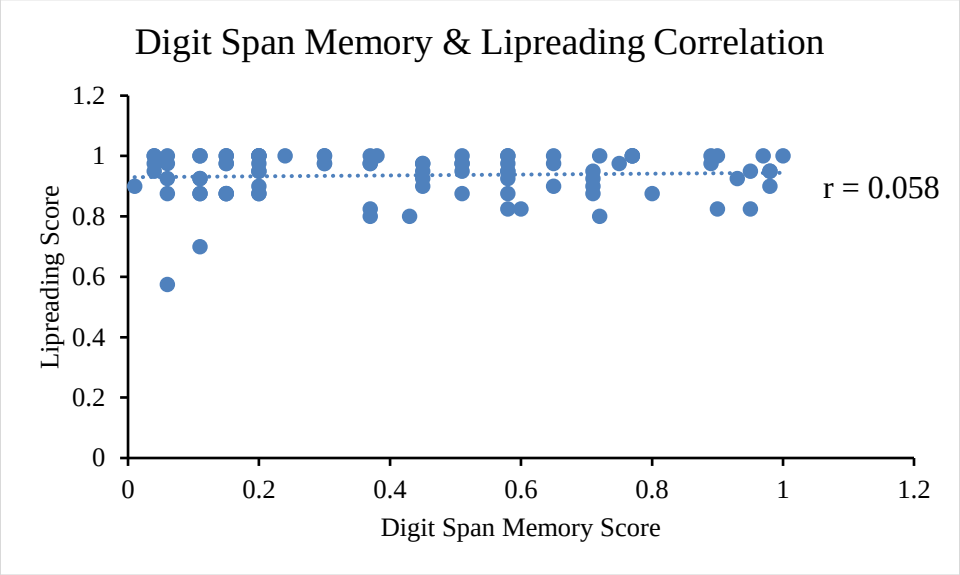
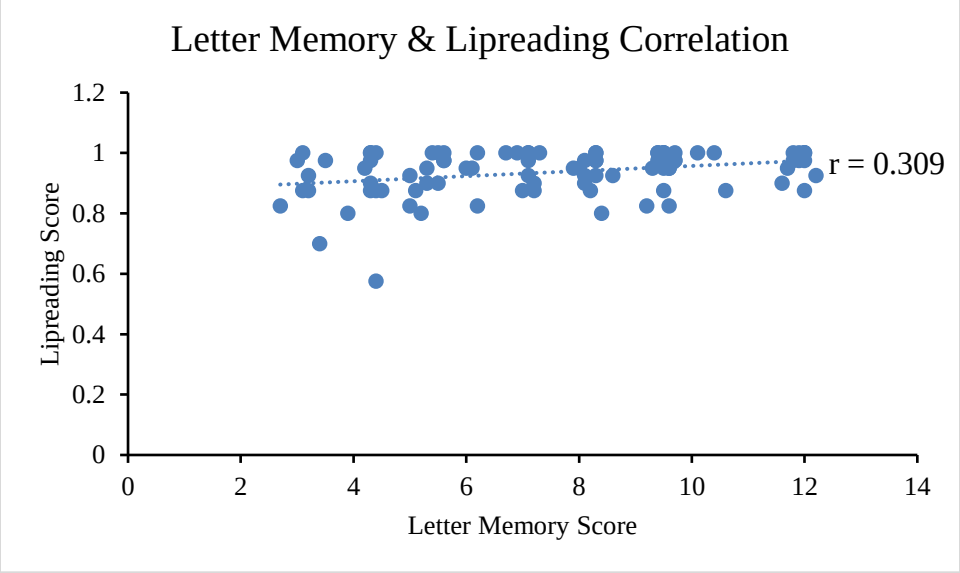


Figure 27

Pearson's Correlation Analysis of Letter Memory and Lipreading Capacity



Visual Cues and Language Comprehension

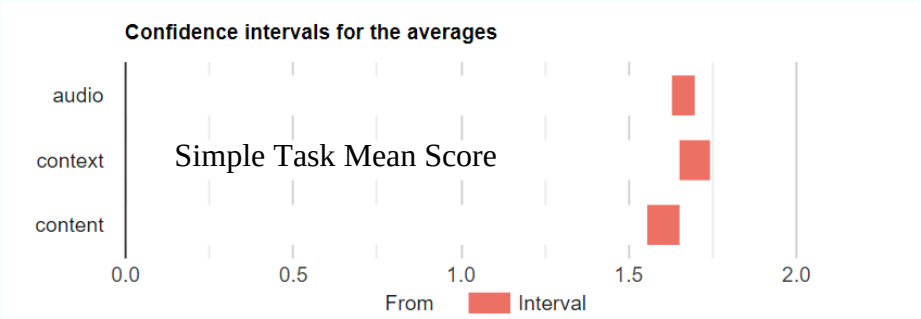
Next, I examined how context and content visual cues affect listening comprehension of discourse-level conversations in L2 English learners. The listening tasks were divided into three categories: simple, intermediate, and complex, each containing talks in audio-only or audiovisual formats. We begin by analyzing the effect of these visual aids on simple listening tasks, followed by a more detailed examination of the intermediate and complex levels.

Simple Talks

The simple listening tests contain short conversations with 2-3 sentences followed by one direct and one indirect question. A correct answer equals one point, and the mean points under each audiovisual condition (audio-only, audio with context visual, and audio with content visual) were calculated. Figure 28 shows that the mean number of questions answered correctly is 1.661 for the audio-only condition, 1.697 for audio with context visual cues, and 1.602 for audio with content visual cues. Pairwise comparisons revealed no statistically significant difference between the audio-only and context-visual conditions ($P = 0.3266$), suggesting that the addition of context visuals did not significantly improve comprehension in the simple listening task. However, there was a statistically significant difference between the audio-only and content visual conditions ($P = 0.03174$), as well as between the context visual and content visual conditions ($P = 0.007073$), indicating that content visuals may have negatively impacted comprehension compared to both audio-only and context visuals. These results imply that, for simple listening tasks, content visuals might introduce extraneous information or distractions that hinder comprehension, while context visuals have little to no effect. This finding suggests that the type of visual aid used can have varying impacts depending on the nature of the listening task and that more straightforward tasks may not always benefit from added visual stimuli, especially regarding content-related imagery.

Figure 28

Participants' Mean Score on Simple Listening Tasks Under Different Audiovisual Conditions



Because there were 20 simple talks and each talk had different visual cues, I then asked if all the talks showed the same effect of the visual cues on listening comprehension. To do this, I compared the mean for audio-only, context-visual, and content-visual tests for each talk. As shown in Figures 29-32, only simple talks 3, 9, 14, and 16 showed statistical differences in the test scores under different audiovisual conditions. Both talks 3 and 14 displayed lower mean scores for the content-visual condition compared to audio-only. Specifically, in Talk 3, the mean score for content-visual (1.483) was significantly lower than for audio-only (1.836, P = 0.003991), and in Talk 14, content-visual (1.25) was significantly lower than audio-only (1.618, P = 0.007072) and context-visual (P = 0.03291). This suggests that for both talks, content visuals negatively impacted comprehension, indicating that certain types of visual information can distract or hinder the listener's ability to process the spoken material. In contrast, talks 9 and 16 exhibited different patterns for context visuals. For talk 9, the context-visual condition (1.68) resulted in a significantly lower mean score compared to the audio-only condition (1.918, P = 0.02035), suggesting that context visuals may have been distracting in this case. However, for talk 16, the context-visual condition (1.6) produced a significantly higher mean score than audio-only (1.163, P = 0.01439), indicating that context visuals enhanced comprehension for this particular talk.

These results show that while content visuals tend to reduce comprehension in certain talks (talks 3 and 14), the effect of context visuals is less predictable, sometimes helping (talk 16) and sometimes hindering (talk 9) understanding. Overall, because no other talks showed statistical differences in the test scores under the three audiovisual conditions, I concluded that the impact of visuals on language comprehension in simple listening tasks is limited.

Figure 29

Participants' Mean Score on Simple Talk 3 Under Different Audiovisual Conditions

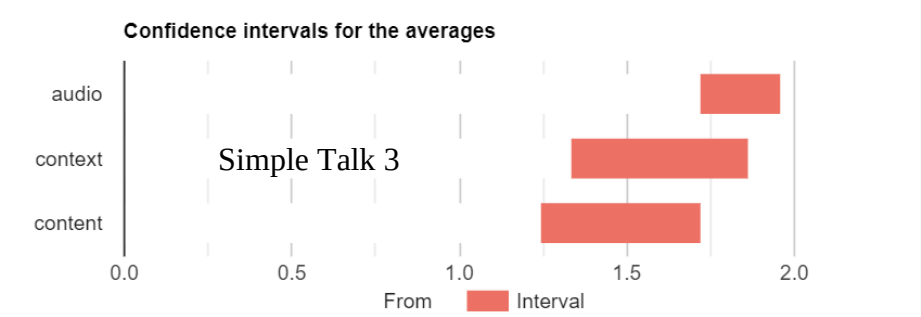


Figure 30

Participants' Mean Score on Simple Talk 9 Under Different Audiovisual Conditions

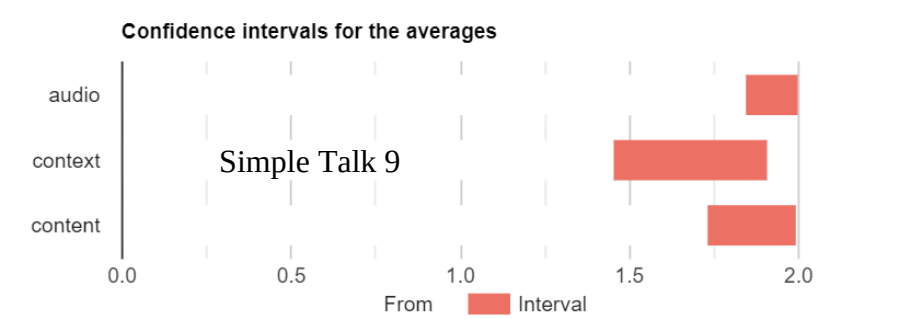


Figure 31

Participants' Mean Score on Simple Talk 14 Under Different Audiovisual Conditions

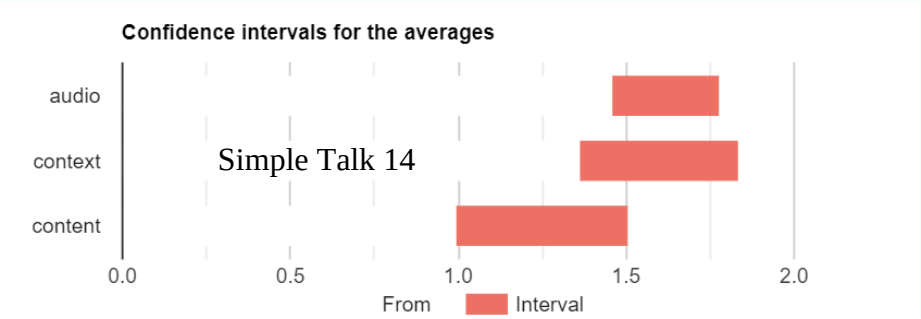
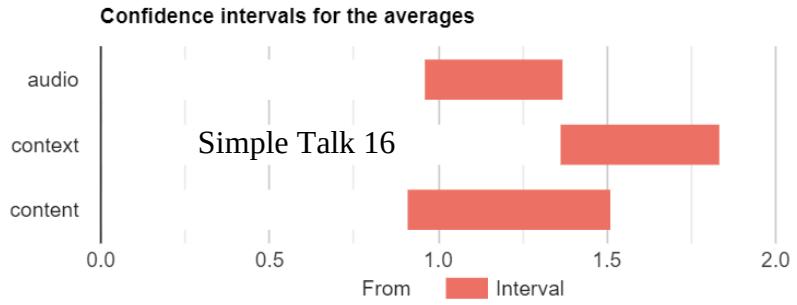


Figure 32

Participants' Mean Score on Simple Talk 16 Under Different Audiovisual Conditions



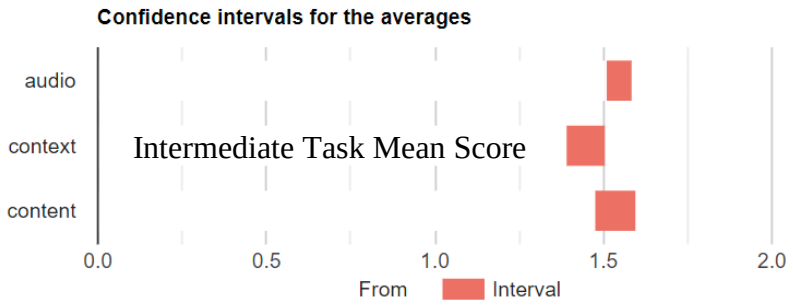
Intermediate Talks

The analysis of listening comprehension for intermediate tasks revealed interesting patterns in how different visual cues influenced performance. The intermediate tests contain slightly more complex talks followed by one direct and one indirect question. The mean scores were as follows: audio-only yielded the highest mean score (1.547), followed closely by content visual (1.536), while context visual resulted in a lower mean score (1.448). Statistical comparisons showed that the audio-only condition significantly outperformed the context-visual condition, with a P value of 0.00736. This suggests that, for intermediate tasks, the presence of context visuals may have distracted or hindered comprehension compared to relying solely on audio. However, there was no significant difference between the audio and content visual conditions (P = 0.9712), indicating that content visuals neither helped nor hurt comprehension. Additionally, a significant difference was observed between the context and content visual conditions (P = 0.02618), with content

visuals performing better than context visuals (Figure 33). These results suggest that for intermediate listening tasks, context visuals tend to impair comprehension compared to both audio-only and content visuals. Content visuals, in contrast, seem to be overall neutral in their effect, neither enhancing nor diminishing comprehension. The findings highlight the potential for context visuals to be distracting in more challenging listening tasks.

Figure 33

Participants' Mean Score on Intermediate Listening Tasks Under Different Audiovisual Conditions



Because there were 20 intermediate talks and each talk had different visual cues, I then asked if all the talks showed the same effect of the visual cues on listening comprehension. To do this, I compared the mean for audio-only, context-visual, and content-visual tests for each talk. The analyses reveal a nuanced effect of visual cues on listening comprehension. For 4 out of 20 intermediate tasks, content visuals improved performance compared to audio-only conditions, as seen in talks 12, 13, 14, and 20, where the content visuals consistently yielded higher mean scores (Figures 34-37). For talk 12, the mean for content visuals was 1.681, significantly higher than the audio-only condition (1.290), with a P value of 0.02515. Similarly, talk 13 showed a significant improvement in the content visual condition (1.681) compared to audio-only (1.309), with a P value of 0.03939. Talk 14 followed this trend, where the content visual condition (1.727)

performed significantly better than the audio-only condition (1.418), with a P value of 0.0476. Finally, talk 20 also showed improved performance with content visuals (1.636) over audio-only (1.327), with a P value of 0.03308. These results suggest that, for these intermediate tasks, content visuals enhance comprehension by providing learners with additional cues that help them process the spoken information more effectively. The lack of significant differences between the audio and context visual conditions for these tasks further suggests that context visuals neither help nor hinder comprehension.

Figure 34

Participants' Mean Score on Intermediate Talk 12 Under Different Audiovisual Conditions

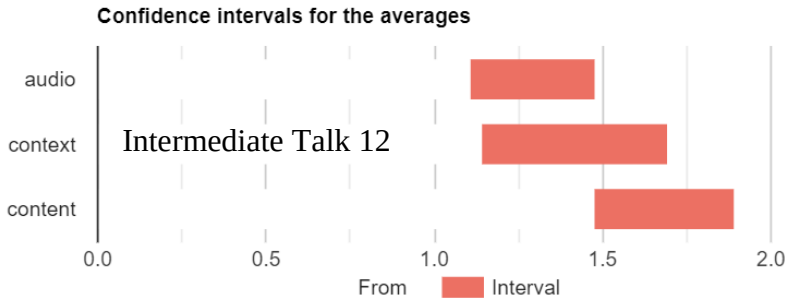


Figure 35

Participants' Mean Score on Intermediate Talk 13 Under Different Audiovisual Conditions

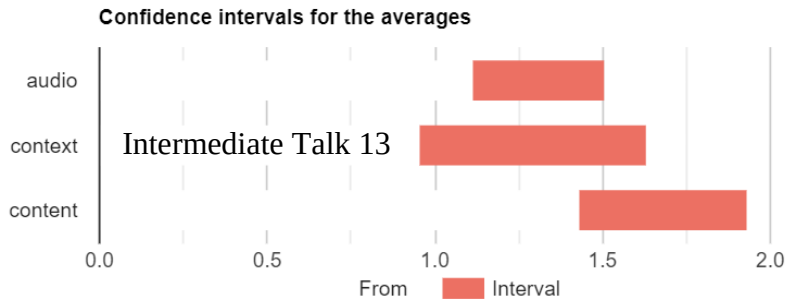


Figure 36

Participants' Mean Score on Intermediate Talk 14 Under Different Audiovisual Conditions

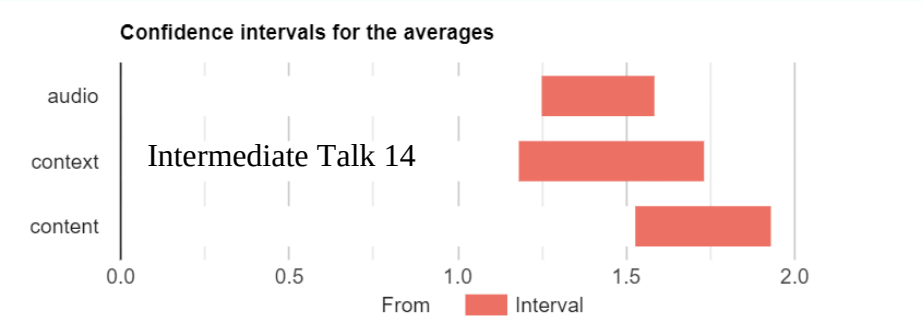
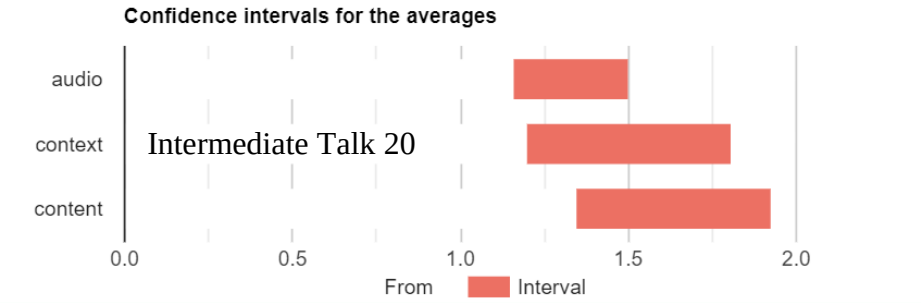


Figure 37

Participants' Mean Score on Intermediate Talk 20 Under Different Audiovisual Conditions



However, talk 9 presents a notable exception. In this case, the content visuals resulted in significantly worse performance (1.260) compared to the audio-only condition (1.765), with a P value of 0.001138. This suggests that for this particular task, the content visuals may have introduced confusion or distracted participants from the audio, leading to decreased comprehension. A further examination found that talk 9 describes how a young woman, Kathy, decided to learn photography using her newly bought camera. Kathy was practicing taking photos in a park; however, she was not happy with her photographs until she tried the Auto mode. The content visual for this talk is an image of a smiling young woman holding a camera in a room. Although relevant, the image does not provide correct information about Kathy's location (in a park), and one cannot determine if Kathy is good at the Manual or Auto mode from the image. The lower score of content visuals in this talk highlights that content visuals, while generally helpful,

can occasionally interfere with comprehension, possibly due to the quality and relevance of content visual cues.

On the other hand, 5 out of 20 intermediate tasks, talks 2, 3, 6, 7, and 15, showed worse performance of the context visuals compared to audio-only conditions (Figures 38-42). This suggests that, for these particular tasks, the additional information provided by context visuals may have been distracting or misaligned with the auditory content, leading to decreased performance. In talk 2, participants performed significantly better in the audio-only condition (1.744) compared to the context visual condition (1.433), with a P value of 0.02477, while the content visual condition (1.652) did not significantly differ from the audio-only condition. Similarly, talk 3 showed significantly worse performance in the context visual condition (1.533) compared to the audio-only condition (1.829), with a P value of 0.02689, while content visuals (1.652) again showed no significant difference. Talk 6 demonstrated a similar result, with the context visual condition (1.033) leading to worse performance compared to the audio-only condition (1.361), with a P value of 0.04766. In this case, neither context nor content visuals seemed to provide substantial aid, as the content visual condition also resulted in a lower mean score (1.086), though this difference was not statistically significant. Talk 7 also followed this pattern, with the context visual condition (1.5) resulting in significantly lower performance than the audio-only condition (1.787), with a P value of 0.02844. Lastly, in talk 15, context visuals once again hindered performance (1.375) compared to the audio-only condition (1.672), with a P value of 0.03043. Interestingly, this task also showed a significant difference between the context and content visual conditions (P value 0.02785), where content visuals resulted in better performance (1.727) than context visuals, suggesting that the type of visual information provided may have a crucial impact on comprehension.

Figure 38

Participants' Mean Score on Intermediate Talk 2 Under Different Audiovisual Conditions

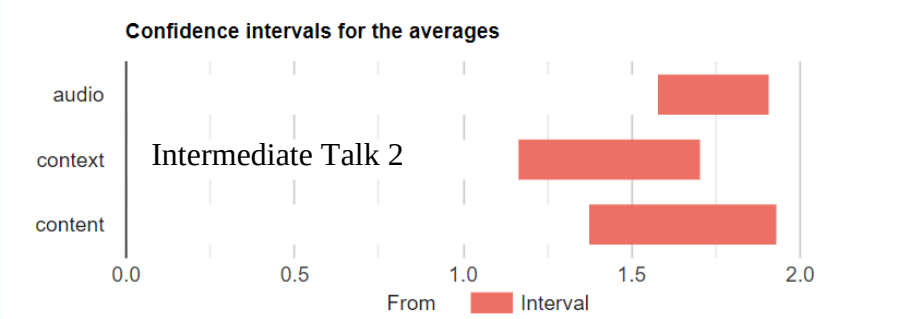


Figure 39

Participants' Mean Score on Intermediate Talk 3 Under Different Audiovisual Conditions

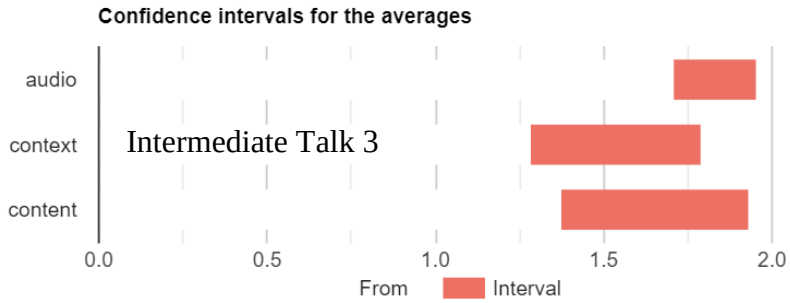


Figure 40

Participants' Mean Score on Intermediate Talk 6 Under Different Audiovisual Conditions

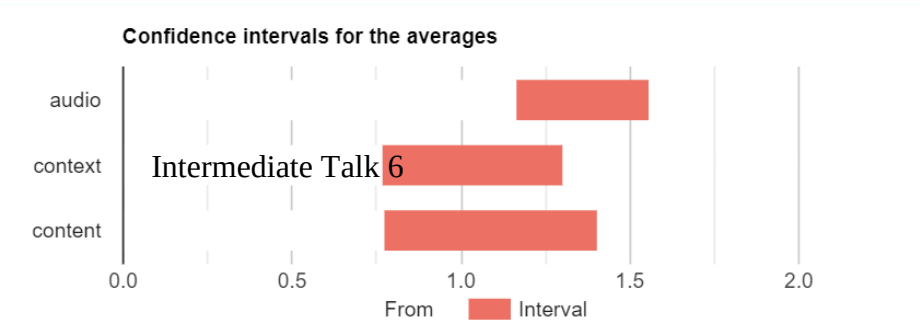


Figure 41

Participants' Mean Score on Intermediate Talk 7 Under Different Audiovisual Conditions

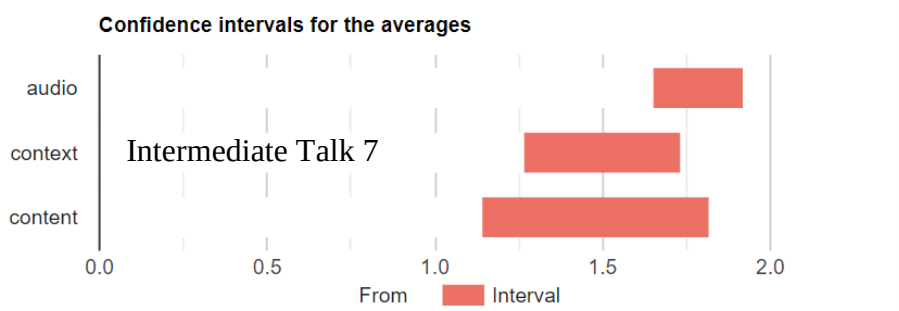
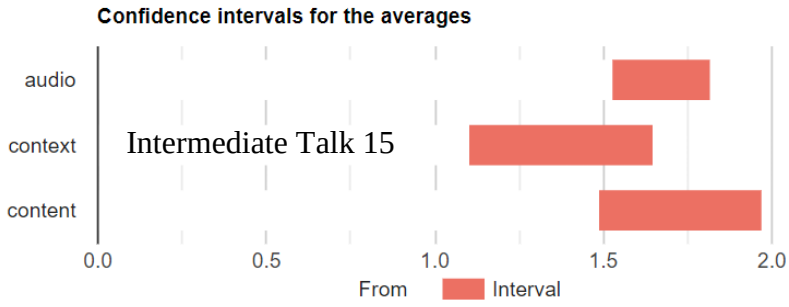


Figure 42

Participants' Mean Score on Intermediate Talk 15 Under Different Audiovisual Conditions



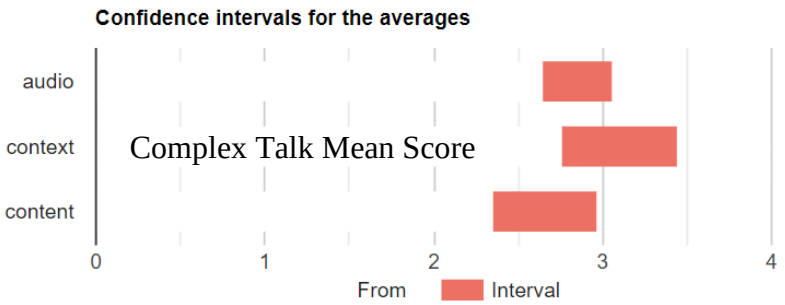
These results imply that context visuals, which provide situational information about the speaker (like a classroom), may sometimes distract participants or compete with the auditory information, leading to confusion or cognitive overload. The performance drop with context visuals across multiple tasks highlights the importance of careful visual aid selection. Not all visuals will enhance listening comprehension—some may inadvertently impair it, especially in tasks with intermediate complexity. In contrast, content visuals, which directly relate to the speech content, seem to better align with the auditory input and can improve or maintain comprehension levels. Occasionally, content visuals may introduce difficulties depending on the specific context or content of the talk. Such variability in performance suggests that careful consideration is needed when integrating content visuals.

Complex Talks

Next, I examined whether visual cues help improve comprehension of complex listening materials. As shown in Figure 43, the results show distinct performance trends between the three conditions: audio-only, context visual, and content visual. The mean score for the audio-only condition was 2.849 (with 6 questions for each complex talk and a maximum score of 6), compared to a higher mean score of 3.095 for the context visual condition, and a lower mean score of 2.66 for the content visual condition. Although the context visual condition yielded the highest mean score, indicating that participants performed better with context visuals than audio-only or content visuals, the pairwise comparison between audio and context visual conditions did not reach statistical significance (P = 0.1917). This suggests that while context visuals might have offered some benefit in aiding comprehension of complex tasks, the improvement over the audio-only condition was not strong enough to be conclusive. Similarly, the comparison between audio and content visual conditions yielded a P value of 0.3889, showing no significant difference between these two conditions. On the other hand, the comparison between context and content visual conditions was close to significance (P = 0.06156), indicating a trend where context visuals may be less distractive than content visuals for these complex tasks. In other words, the situational background information provided by context visuals might demand less cognitive power than information-rich content visuals, which help listeners focus on the auditory signals to comprehend difficult listening materials.

Figure 43

Participants' Mean Score on Complex Listening Tasks Under Different Audiovisual Conditions



I then asked if all 4 complex talks showed the same effect of the visual cues on listening comprehension. As shown in Figure 44, complex talk 2 showed significantly better performance in audio-only than context or content-visual conditions. The mean score for the audio condition was 2.857, compared to 2.174 for the context visual and 2.069 for the content visual condition. The pairwise comparison showed a P value of 0.04939 for the audio-context comparison and 0.01889 for the audio-content comparison, indicating that participants comprehended the material significantly better when relying solely on audio. The context-content comparison showed no significant difference (P = 0.8485), suggesting that neither type of visual cue was notably beneficial over the other. The fact that audio-only conditions outperformed visual conditions implies that the visual cues—context or content—might have introduced distractions or cognitive overload when participants were already handling challenging content. A further examination reveals that talk 2 describes the increasing social phenomenon of overconsumption, and how the Buy-Nothing movement helps save us from economic and environmental problems. The questions ask many details about overconsumption, including the money spent purchasing clothes and the percentage of clothes sent for recycling. The content visuals, though relevant to cloth shopping and recycling, do not contain information that can accurately answer the questions. It is possible that for this specific talk, the auditory information was more straightforward to process without additional, possibly irrelevant, visual input, leading to better performance.

Figure 44

Participants' Mean Score on Complex Talk 2 Under Different Audiovisual Conditions

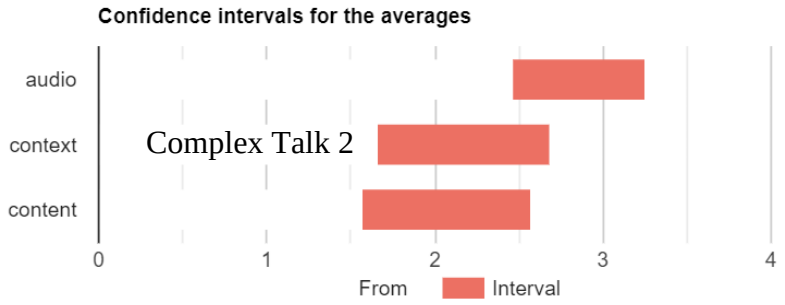
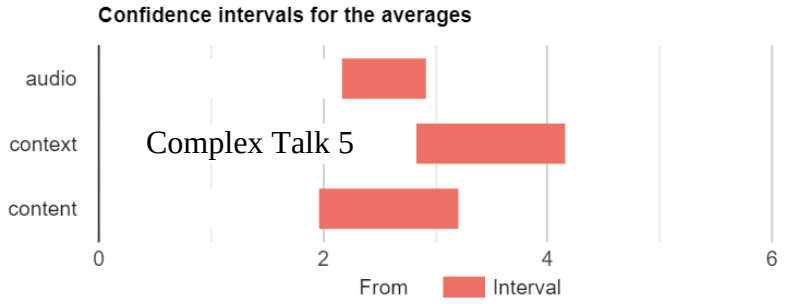


Figure 45

Participants' Mean Score on Complex Talk 5 Under Different Audiovisual Conditions



In contrast, talk 5 demonstrated the opposite trend: participants performed significantly better in the context-visual than in the audio-only condition, with a mean score of 3.5 for context visuals compared to 2.545 for audio and 2.583 for content visuals (Figure 45). The pairwise comparison showed a P value of 0.01202 for the audio-context comparison, confirming a positive effect of context visuals. In contrast, the audio-content (P = 0.7504) and context-content (P = 0.06304) comparisons did not reach significance. The near-significant result between context and content visuals suggests that the situational cues provided by context visuals might be less distractive than the information-rich, content-specific visuals. The findings for talks 2 and 5 highlight how the complexity of the listening material interacts with the type of visual support

provided. In some cases, as with talk 2, participants may rely more effectively on auditory information alone, while in others, like talk 5, context visuals provide beneficial results, even though its mechanism remains unclear. These findings suggest that for complex and difficult listening materials, information-rich content visuals might not help improve listening comprehension compared with audio-only inputs. Nevertheless, due to the limited sample size, further investigation is needed to determine visual cues' precise role in understanding complex tasks.

English Proficiency and Language Comprehension

In this section, I explored the potential relationship between participants' English proficiency levels and the impact of visual cues on their listening comprehension. Given participants' diverse range of English proficiency, it is important to assess whether higher language skills correlate with greater (or lesser) benefits from context and content visual aids. By analyzing the interaction between proficiency and visual cue effectiveness, I expected to determine if visual support enhances comprehension differently for individuals based on their language capabilities, especially when dealing with varying levels of task complexity.

Dividing Participants into Different Proficiency Levels

I divided participants into three different proficiency levels. I employed two approaches to ensure a comprehensive assessment of their language abilities.

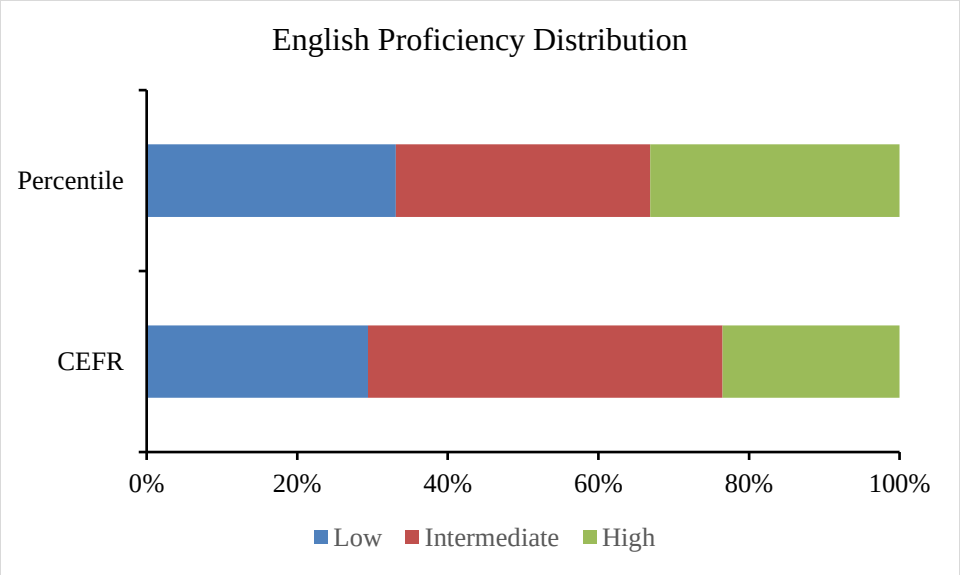
In the first approach, I converted all participants' standardized test scores into the Common European Framework of Reference for Languages (CEFR) scale using conversion tables (https://theedge.com.hk/conversion-table-for-toefl-ibt-pbt-cbt-tests/, https://englishtest.duolingo.com/scores) as discussed previously. For the participants who had only non-standardized English scores, such as the Brazilian Local College English Test and Test of

English for Academic Purposes (TEAP), they reported the equivalent CEFR level based on their test scores. Based on the CEFR scales, I classified the learners into three proficiency groups: A2-B1 (advanced beginner to lower intermediate), B2 (upper intermediate), and C1-C2 (advanced). This method allowed me to place participants into well-established language proficiency bands based on their previous test results. However, one limitation of this approach was that some standardized test scores were obtained several years ago. As a result, the actual language proficiency of some participants may have changed since their original assessment. This led to some discrepancies, where participants assigned to higher CEFR levels based on their past scores performed poorly on the current tests. Conversely, those with lower CEFR levels sometimes performed better than expected.

To address the potential inconsistencies from outdated standardized scores, I used a second approach to classify participants according to their overall performance on the listening comprehension tests (which included simple, intermediate, and complex tasks). This allowed me to categorize participants based on their most recent listening skills. For this approach, I divided participants into three proficiency groups: low proficiency (the bottom 50% of scores), intermediate proficiency (25-75% of scores), and high proficiency (the top 50%). This ranking system reflected their current listening abilities, providing a more accurate reflection of their language comprehension skills at the time of testing while ensuring a sufficient sample size in each group for meaningful analysis. Figure 46 shows the distribution of participants using either a self-reported CEFR scale or performance percentiles.

Figure 46

Distribution of Participants Based on Performance Percentiles or CEFR Scales



By utilizing both of these methods, I was able to cross-validate participants' proficiency levels and provide a robust framework for analyzing the effects of visual cues on listening comprehension across different language skill levels.

Simple Talks

I first examined how low, intermediate, and high English proficiency participants performed in simple talks. As shown in Table 4, the results for simple listening tasks reveal interesting trends regarding how participants of different proficiency levels respond to visual cues. While no significant enhancement was found for content visuals across all participants, there is a noticeable trend where content visuals seem to improve comprehension for participants in the B2 intermediate (n=40), Lower 50% low-proficiency (n=46), and 25%-75% intermediate-proficiency (n=47) groups over the audio-only condition. For example, in the B2 intermediate group, talk 15 shows a higher mean score for content visuals (2.0) compared to audio-only (1.765), with a near-significant P value of 0.09266. Similarly, in the lower 50% group, talk 15 shows a slight improvement with content visuals (mean 1.909 vs. audio mean 1.69, P = 0.07662). The 25%-75% group also exhibits a positive trend, with talk 12 showing a content visual mean of 2.0

compared to an audio-only mean of 1.833 (P = 0.07891). These trends suggest that content visuals may provide some benefit for participants with low to intermediate proficiency, aiding their comprehension of simple tasks. A larger sample size is required to draw more definitive conclusions.

Table 4

Simple Talks Showing Context- or Content-Visual-Based Interference or Enhancement, Grouped by English Proficiency Level

Simple Talks	Context interference	Content interference	Context enhancement	Content enhancement
All participants	9	3, 14	16	-
A2-B1	-	-	-	-
Lower 50%	-	3	2, 6	-
B2	3	-	17	-
25%-75%	1, 9	14	-	-
C1-C2	-	15	-	-
Top 50%	-	14	-	-

However, similar to the results with all participants, visual cue interference is also evident in low- and intermediate-proficiency groups, with some context or content visuals disrupting comprehension. For example, in the B2 intermediate group, talk 3 shows a significant negative impact of context visuals (audio mean = 1.913, context mean = 1.333, P = 0.0225) compared to the audio-only condition. Similarly, in the 25%-75% intermediate group, both talk 9 (audio mean = 1.957, context mean = 1.538, P = 0.005761) and talk 14 (audio mean = 1.708, content mean = 1.286, P = 0.01228) demonstrate a detrimental effect of visual cues. The lower 50% group also shows a negative impact of content visuals in talk 3, with a significant drop in performance (audio mean = 1.706, content mean = 1.176, P = 0.01544). Interestingly, the advanced groups (Top 50% and C1-C2 advanced) showed minimal visual cue interference, with significant effects seen only in talk 14 (Top 50%, content visual interference) and talk 15 (C1-C2, content visual interference).

This suggests that participants with lower and intermediate proficiency levels are more susceptible to the impact of visual cues—both positively, with content visuals enhancing comprehension in some cases, and negatively, where visual overload or distraction may hinder comprehension.

These findings indicate that the effect of visual cues on listening comprehension varies depending on participants' proficiency levels. For lower and intermediate proficiency learners, content visuals may provide additional support in some cases, particularly in simple tasks. However, these learners are also more prone to visual interference, suggesting a potential cognitive overload or distraction. Advanced participants seem less affected by visual cues overall, which could imply that their stronger linguistic skills enable them to rely more on auditory input, with less reliance on or interference from visual aids. This insight could guide the design of instructional materials, suggesting that more careful consideration is needed when incorporating visual elements for lower-proficiency learners.

Intermediate Talks

I then examined how low, intermediate, and high English proficiency participants performed in talks of intermediate difficulty. As shown in Table 5, content visuals enhanced comprehension over the audio-only condition for intermediate-level participants (B2 and 25%-75% groups) but also showed some benefits for low-proficiency participants (A2-B1). For example, content visuals significantly improved comprehension in talks 12 and 20 for intermediate-proficiency participants. The B2 group showed better performance with content visuals in talk 12 (audio mean: 1.294, content mean: 1.818, $P = 0.02586$), and the 25%-75% group showed a similar enhancement in talk 20 (audio mean: 1.32, content mean: 1.8, $P = 0.03857$). Additionally, there were beneficial trends of content visual enhancement in the low-proficiency A2-B1 group for talk 13 ($P = 0.05012$) and in the 25%-75% group for talk 12 ($P = 0.07299$), though

these effects were not as statistically robust in some cases. In contrast, advanced groups did not show a consistent benefit from content visuals, except for a trend in talk 12 for the C1-C2 group (P = 0.06336). This suggests that more advanced participants might rely less on content visuals to aid comprehension, possibly due to their higher language proficiency and better ability to process audio-only information.

Table 5

Intermediate Talks Showing Context- or Content-Visual-Based Interference or Enhancement, Grouped by English Proficiency Level

Intermediate Talks	Context interference	Content interference	Context enhancement	Content enhancement
All participants	2, 3, 6, 7, 15	9	-	12, 13, 14, 20
A2-B1	3	-	1	13 (P = 0.05012)
Lower 50%	3, 15	-	-	-
B2	-	-	-	12
25%-75%	5, 15	-	-	12 (P = 0.07299), 20
C1-C2	6, 8	9	-	12 (P = 0.06336)
Top 50%	-	9	-	-

On the other hand, context visuals seemed to interfere with comprehension in several cases. Talks 2, 3, 6, 7, and 15 showed significant context interference across all participants, with talk 9 showing content interference. This pattern was also evident in specific proficiency groups: for instance, in the A2-B1 low-proficiency group, talk 3 showed significant context interference (audio mean: 2, context mean: 1.4, P = 0.02195). Similar interference was found in lower-proficiency groups, such as the lower 50% group, where talks 3 (audio mean 1.867; context mean 1.263; P=0.008928) and 15 (audio mean 1.548; context mean 1; P=0.01023) exhibited significant context interference, and the 25%-75% group, which showed context interference in talks 5 (audio mean 1.591; context mean 1.071; P=0.02641) and 15 (audio mean 1.8; context mean 1.273; P=0.007175). In contrast to the intermediate and low-proficiency groups, the advanced groups

(C1-C2, Top 50%) showed context interference with talks 6 and 8, and content interference with talk 9. This suggests that for more advanced listeners, visual cues—whether context or content—may not offer additional comprehension benefits and may, in some cases, introduce distractions or interfere with their ability to process auditory information efficiently. This emphasizes that for proficient learners, visuals need to be carefully curated to avoid distracting from the core auditory information, especially in complex listening tasks. Together, these findings suggest that context visuals might sometimes hinder comprehension for participants of all proficiency levels (low, intermediate, advanced), while content visuals can provide some benefit, especially for participants at the intermediate proficiency level. Further research with larger samples could help clarify the conditions under which visual cues assist or interfere with comprehension.

Complex Talks

Finally, I examined the effects of visual cues on listening comprehension of complex materials in participants of different proficiency levels. As discussed previously, talk 2 showed the best performance with audio-only clips (no visual inputs), while talk 5 seemed to have the best results from context visuals. Interestingly, when examining participants of specific proficiency levels, content visuals appeared to impair comprehension in low-proficiency learners (lower 50%, n=44) for talk 2, with a significant drop in performance (audio mean 2.059, content visual mean 1.333, P=0.03649) (Table 6). This suggests that lower-proficiency learners may struggle to integrate visual and auditory information effectively, particularly when the audio content is too complex or dense. In contrast, advanced learners (C1-C2, n=20) experienced mixed results with context visuals. For talk 3, context visuals interfered with comprehension (audio mean 5.714, context mean 4.6, P=0.01264), but for talk 5, context visuals enhanced their performance (audio mean 2.615, context mean 4.25, P=0.03419). Given the potential inaccuracy of the self-reported

CEFR ranking and the small sample size (n=20), I re-examined the advanced learners who ranked among the top 50% of all participants (n=46). I found no significant differences in their performance across audio-only, content-visual, and context-visual conditions (Table 6), indicating top performers are less susceptible to interference from visual cues and can process complex auditory information without relying on additional visual input. Likewise, participants with intermediate proficiency performed equally well across audio-only, content-visual, and context-visual conditions, suggesting visual cues are dispensable for these learners to comprehend complex materials.

Table 6

Complex Talks Showing Context- or Content-Visual-Based Interference or Enhancement, Grouped by English Proficiency Level

Complex Talks	Context interference	Content interference	Context enhancement	Content enhancement
All participants	2	-	5	-
A2-B1	-	-	-	-
Lower 50%	-	2	-	-
B2	-	-	-	-
25%-75%	-	-	-	-
C1-C2	3	-	5	-
Top 50%	-	-	-	-

The Role of Visual Cues in Listening Comprehension Across Proficiency Levels and Task Complexity

In summary, the analysis of visual cues—both content and context—in listening comprehension tasks reveals a complex relationship that varies with participants' English proficiency and the difficulty of the listening material. Across all proficiency levels, the data highlight that visual aids can both enhance and impair listening comprehension depending on the nature of the task and the learner's ability to integrate multimodal information.

For low-proficiency learners (A2-B1 and Lower 50% Groups), the results consistently show that visual cues, especially content visuals, often introduce more challenges than benefits, particularly in complex listening tasks. For instance, in simple and intermediate listening tasks, there were trends suggesting that content visuals could enhance comprehension, as seen in simple talk 15 and intermediate talk 13, although the difference was not statistically significant ($0.05<P<0.1$). However, for simple talk 3 and complex talk 2, content visuals significantly impaired comprehension. The limited language processing ability of low-proficiency participants likely leads to cognitive overload, where managing both visual and auditory inputs proves too challenging.

For intermediate learners (B2 and 25%-75% Groups), there is evidence of both benefits and drawbacks of visual cues. Content visuals showed a significant positive effect on comprehension in several tasks, particularly in intermediate talks 12 and 20. This suggests that intermediate learners may benefit from visuals that provide specific, relevant information, helping them to better understand and retain the audio content. However, there were also instances of visual interference in this proficiency range, particularly from context visuals. For example, simple talks 1, 3, 9, and intermediate talks 5 and 15 showed a strong interference effect, where context visuals lowered comprehension scores compared with the audio-only conditions. This indicates that while intermediate learners can benefit from well-matched visuals, poorly aligned or irrelevant visuals can distract from their auditory processing, resulting in decreased performance.

For advanced learners (C1-C2 and Top 50% Groups), the role of visual cues appears less significant. The data show that for most of the tasks, advanced learners (particularly the Top 50% group) performed consistently across audio-only, context-visual, and content-visual conditions. This suggests that at higher proficiency levels, participants can process auditory information more

efficiently and may not rely on visual support to the same extent as lower- or intermediate-level learners.

On the other hand, the complexity of the listening material plays a significant role in how visual cues affect comprehension. For simpler tasks, visual aids seem to result in more potential interference, while content visuals showed the most beneficial effects for intermediate talks, helping bridge the gap between comprehension and understanding more detailed information. However, as task complexity increases, the effectiveness of visual cues appears to diminish. For complex listening tasks, visual aids are often more likely to interfere with comprehension than enhance it. This was observed in complex talk 2, where both content and context visuals impaired comprehension across all proficiency levels. It is likely that for complex tasks, learners are already cognitively overloaded by processing the dense auditory information, and additional visual cues introduce more elements that compete for attention rather than support comprehension.

Conclusion

In summary, the results of this chapter offer valuable insights into how visual cues affect speech perception and language comprehension across different proficiency levels. The findings highlight that while working memory and lipreading capacity play a limited role in the expression of the McGurk effect, the impact of visual cues on listening comprehension is more complex and proficiency-dependent. For intermediate-proficiency learners, visual aids—especially content visuals—can enhance comprehension when well-aligned with the listening material, but they can also interfere when the task complexity overwhelms the learner's cognitive capacity. In contrast, advanced learners exhibit less reliance on visual cues, and in some cases, even experience negative effects from additional visual information, particularly during complex tasks. These results underscore the importance of carefully integrating multimodal elements in language learning,

providing guidance for educators to tailor their instructional strategies based on learner proficiency and task demands. Overall, this chapter presents a comprehensive analysis of the interplay between visual and auditory information in second-language learning, paving the way for more effective teaching methods that meet diverse learner needs.

CHAPTER 5

SUMMARY, IMPLICATIONS, and DISCUSSION

This chapter provides a summary of the key findings from this study, discusses their broader implications, and explores potential future directions for research and practical application. By investigating the roles of working memory, lipreading capacity in the McGurk effect's variation and visual cues, and language proficiency in listening comprehension, this research contributes to a deeper understanding of second language acquisition. In particular, the study sheds light on the complexities of how learners at varying proficiency levels process auditory and visual information in different task settings. The chapter will synthesize these findings and explore their significance for language educators, curriculum designers, and researchers. Additionally, the study's limitations will be addressed, along with recommendations for future research to deepen our understanding of the multifaceted relationship between cognitive factors and language comprehension.

Summary of Findings

This study investigated the relationship between working memory, lipreading capacity, and the McGurk effect's expression. It also examined how visual cues and proficiency levels impact listening comprehension among second-language learners. The key findings can be summarized as follows:

Working Memory, Lipreading Capacity, and the McGurk Effect

Participants showed variability in experiencing the McGurk effect, with working memory and lipreading capacity playing a limited role. Pearson's correlation analyses indicate a weak

positive relationship between letter working memory, lipreading capacity, and the McGurk score, suggesting that other cognitive processes are likely involved in the individual variance of the McGurk effect.

Impact of Visual Cues on Listening Comprehension

The role of visual cues—both context and content visuals—varied across task complexity and proficiency levels.

Simple Tasks. Across all proficiency levels, no significant benefit was observed from content visual cues in simple listening tasks. However, a trend of content visuals improving comprehension was noted in intermediate (B2) and lower-proficiency participants, while advanced participants showed no such enhancement.

Intermediate Tasks. Content visuals enhanced listening comprehension scores, particularly in intermediate-level learners (B2) and participants ranked in the middle proficiency range (25%-75%). However, context visuals interfered with performance in several tasks, especially for low- (A2-B1, Lower 50%) and intermediate-proficiency (25%-75%) participants.

Complex Tasks. In complex tasks, content and context visuals were often found to impair comprehension scores. While context visuals enhanced comprehension for one task, overall performance was not significantly improved for higher-proficiency participants. Low-proficiency learners, however, showed occasional negative impacts from content visuals.

Proficiency Level and Visual Cue Interaction

The impact of visual cues on comprehension was closely linked to proficiency levels. Lower- and intermediate-proficiency participants were more likely to benefit from content visuals in less complex tasks but were also more susceptible to interference from context visuals.

Advanced learners showed less benefit from visuals overall, particularly in complex tasks, where their reliance on audio-only cues appeared to be more effective.

Together, these findings reveal that while visual cues can support comprehension in certain contexts, their effectiveness is highly contingent on task difficulty and learner proficiency. Moreover, the potential for interference from irrelevant or conflicting visuals highlights the importance of task design in educational settings.

Discussion

Research Question 1: To What Extent Does Individual Variability in the McGurk Effect Depend on Working Memory and/or Lipreading Capacity?

The findings of this study indicate a subtle relationship between working memory and the McGurk effect. Specifically, the correlation between the McGurk effect and letter memory was low ($r = 0.128$), and even weaker for digit span memory ($r = -0.059$). Additionally, the McGurk effect score showed only a slight correlation with participants' lipreading capacity ($r = 0.081$), which is not significant, suggesting that these cognitive factors alone do not significantly explain variability in how individuals experience the McGurk effect. The modest correlation with letter memory may imply that working memory plays a minor role in processing conflicting auditory and visual information, but it is not a dominant factor. Furthermore, the lack of a strong link between lipreading ability and the McGurk effect suggests that the ability to rely on visual cues may not fully determine susceptibility to the illusion.

Given the competition between auditory and visual inputs during the McGurk tasks, I propose that the left-right brain preference may play a role in the individual’s variation of the McGurk effect. Indeed, studies in the right-handers discovered a 150% higher possibility of the left-hemisphere controlling language processing compared to left-handers (Knecht et al., 2000). In

addition, a recent meta-analysis revealed an association between the genetic influences on handedness and asymmetries in cortical thickness and surface areas of language-related regions, suggesting a link between handedness and language development (Sha et al., 2021). Interestingly, two studies on the brain laterality and language processing found that although left hemisphere is dominant for the auditory task (right-ear advantage, REA), right hemisphere is dominant for the visual task (left visual half-field (VHF) advantage) (Voyer and Boudreau, 2003; Oltedal and Hugdahl, 2017). Given the evidence of the left hemisphere dominant in language and auditory processing and the right hemisphere in visual processing, individual differences in the left-right hemisphere preference could impact how auditory and visual inputs are integrated during speech perception. This is a promising avenue for future research, as investigating brain lateralization might provide deeper insights into why some individuals are more susceptible to the McGurk effect than others.

In conclusion, while the current study suggests only weak connections between working memory and lipreading ability with the McGurk effect, the role of brain lateralization remains a compelling area for further exploration. This could help clarify how cognitive and neural mechanisms interact to shape multisensory integration in speech perception.

Research Question 2: Do Content or Context Visuals Enhance Discourse-level Listening Comprehension in L2 English Learners?

The results demonstrated a complex relationship between visual cues and listening comprehension, with both content and context visuals exerting different effects depending on task complexity and participant proficiency levels.

For simple listening tasks, content visuals did not significantly enhance comprehension across all proficiency levels, although there was a slight trend of improvement with content visuals

in lower- and intermediate-proficiency learners. This suggests that when the listening task is less cognitively demanding, learners might not rely heavily on external visual information to understand the material, possibly because the auditory input alone suffices for comprehension.

For intermediate listening tasks, content visuals generally aided comprehension, particularly in intermediate-level learners (B2) and those in the middle proficiency range (25%-75%). This indicates that when the listening task becomes more challenging, content-rich visuals may serve as helpful supplementary information to aid learners in grasping the material. However, context visuals often interfered with performance, particularly among lower-proficiency participants, suggesting that too much or irrelevant visual information can overwhelm learners and hinder their ability to focus on the key auditory content.

For complex listening tasks, visual cues—both content and context—tended to impair comprehension. This was true for lower-proficiency learners, who demonstrated better comprehension with audio-only input. The added information from visuals may have created cognitive overload, detracting from their ability to focus on the speech itself. Intermediate and advanced learners performed equally well across audio, context visual, and content visual conditions, suggesting that visual cues may be less effective in highly demanding listening contexts.

Research Question 3: Does the Influence of Visual Cues Vary Depending on the Complexity of the Listening Task?

As discussed previously, the data indicate that the effect of visual cues is dependent on the complexity of the listening task. For simple tasks, visual cues had limited influence, suggesting that learners can manage the auditory input without additional support. For intermediate tasks, content visuals were most beneficial, likely because they provided additional context and

reinforcement for understanding more complex ideas. However, context visuals often introduce extraneous information that could confuse or distract learners, particularly those with lower proficiency. This pattern suggests that the usefulness of visual cues lies in their relevance to the content rather than their general presence. In complex tasks, the results demonstrated that both content and context visuals could detract from comprehension. The cognitive load associated with processing complex auditory information, coupled with visual distractions, seemed to overwhelm learners, leading to lower comprehension scores. Advanced learners, in particular, may have developed strong listening strategies that rely more on auditory input, reducing the need for visual support.

Pedagogical Implications

The findings underscore the importance of tailoring visual cues to match both the learners' proficiency level and the complexity of the listening material. For lower- and intermediate-proficiency learners, visual cues can offer valuable support when carefully designed and relevant to the task, but they can also introduce interference when not well-matched to the learner's cognitive capacity. Advanced learners, on the other hand, demonstrate less reliance on visual aids and may experience negative effects from unnecessary visual information, particularly for complex tasks. Educators and instructional designers should be mindful of these dynamics when integrating multimedia content into listening tasks. Visual aids should be thoughtfully selected to ensure they complement, rather than compete with, the auditory material, especially for lower-proficiency learners. For advanced learners, the emphasis might shift more toward enhancing auditory processing skills, with less use of visuals unless they provide essential support for particularly challenging material.

Suggestions for Future Research

Building on the findings of this study, several avenues for future research could further illuminate the role of cognitive and visual factors in speech perception and listening comprehension. First, the weak correlations between working memory and the McGurk effect suggest that other cognitive functions, such as attention control or processing speed, might play a more significant role and warrant investigation. Additionally, since brain lateralization could be key in understanding how auditory and visual inputs are integrated, future studies should incorporate left-right brain preference tests to explore how hemispheric dominance affects susceptibility to the McGurk effect. Expanding the research to include neuroimaging techniques, such as fMRI, could offer a more direct look at the neural networks involved in speech perception. Lastly, larger sample sizes and more diverse participant groups would improve the generalizability of results, particularly when examining the effects of visual cues across different language proficiency levels and listening task difficulties. These directions will provide more comprehensive insights into the intricate relationship between cognition, visual cues, and auditory processing in language learners.

Conclusion

In conclusion, this study provides valuable insights into the complex interplay between working memory, the McGurk effect, visual cues, and listening comprehension in L2 English learners. While the McGurk effect showed only minor correlations with working memory and lipreading capacity, the role of visual input in language processing remains multifaceted and task-dependent. The findings reveal that content and context visuals can either enhance or hinder listening comprehension, depending on the proficiency level of the learners and the complexity of

the listening tasks. Intermediate-level learners (B2) tended to benefit the most from content visuals, particularly in moderately difficult tasks, while advanced learners showed minimal gains from visual cues in complex tasks. These results suggest that the effectiveness of visual cues in supporting listening comprehension varies widely and is influenced by both language proficiency and task difficulty. The study's findings underscore the importance of tailoring visual support in language learning materials to meet learners' needs. By integrating cognitive and perceptual elements into language instruction, educators can better facilitate listening comprehension in L2 learners. However, given the limitations of sample size and the absence of brain lateralization measures, further research is needed to deepen our understanding of the cognitive mechanisms underlying the McGurk effect and visual cue processing in L2 learners.

On the other hand, the findings of this study hold practical implications in our increasingly digitalized world, where virtual reality (VR) and artificial intelligence (AI) are playing larger roles in language teaching and intercultural communication. As VR and AI-driven platforms become more prevalent, understanding how visual and auditory information can best support comprehension is essential. This research suggests that while visual cues may support understanding in certain contexts, they may also introduce interference depending on task complexity and learner proficiency. By expanding our knowledge in this area, future technology-enhanced learning environments can better adapt to learners' needs, enhancing L2 comprehension and fostering more effective cross-linguistic communication. Ultimately, this research contributes to a broader understanding of how multimodal inputs shape language learning and comprehension, offering practical implications for language education and cognitive science.

REFERENCES

Barker, L. (1971). *Listening behavior*. Englewood Cliffs, NJ: Prentice Hall

Basu Mallick, D., Magnotti, J. S., & Beauchamp, M. (2015). Variability and stability in the McGurk effect: Contributions of participants, stimuli, time, and response type. *Psychonomic Bulletin & Review, 22*(5), 1299e1307. https://doi.org/10.3758/ s13423-015-0817-4

Bloomfield, A., Wayland, S.C., Rhoades, E.A., Blodgett, A., Linck, J., & Ross, S. (2010). *What makes listening difficult? Factors affecting second language listening comprehension*. University of Maryland Center for Advanced Study of Language. https://apps.dtic.mil/sti/pdfs/ADA550176.pdf

Bodie, G. D. (2010). Treating listening ethically. *International Journal of Listening, 24*, 185–188. doi:10.1080/10904018.2010.513666

Bodie, G. D. (2012). Listening as positive communication. In T. Socha & M. Pitts (Eds.), *The positive side of interpersonal communication* (pp. 109-125). New York: Peter Lang.

Bodie, G. D., & Crick, N. (2014). Listening, hearing, sensing: Three modes of being and the phenomenology of Charles Sanders Peirce. *Communication Theory*, *24*, 105-123. doi:10.1111/comt.12032

Bodie, G. D., Janusik, L. A., & Välikoski, T-R. (2008). *Priorities of listening research: Four interrelated initiatives. A white paper sponsored by the research committee of the International Listening Association*. International Listening Association. http://www.listen.org/Resources/Documents/White_Paper_PrioritiesResearch.pdf

Bodie, G. D., Pence, M. E., Rold, M., Chapman, M. D., Lejune, J., & Anzalone, L. (2015). Listening competency in initial interactions, II: Applying trait centrality to discover the relative placement of listening competence among implicit competency theories. *Communication Studies, 66*, 528–548. doi:10.1080/10510974.2015.1039657

Bostrom, R. N. (2011). Rethinking conceptual approaches to the study of "listening." *International Journal of Listening*, *25*, 10–26. doi:10.1080/10904018.2011.536467

Bostrom, R.N. & Waldhart, E. (1980). Components in listening behavior: The role of short-term memory. *Human Communication Research*, *6*(3), 221-227. https://doi.org/10.1111/j.1468-2958.1980.tb00142.x

Broadbent, D.E. (1958). *Perception and communication*. Pergamon Press. https://doi.org/10.1037/10037-000

Brown, H. D., & Abeywickrama, P. (2010). Language assessment: Principles and classroom practices (2nd ed.). White Plains, NY: Pearson Education

Brownell, J. (1985). A model for listening instruction: Management applications. *Business Communication Quarterly, 48,* 39–44. doi:10.1177/108056998504800312

Communication in the Real World: An Introduction to Communication Studies (2nd ed.). (2016). University of Minnesota Libraries Publishing.

Cross, J. (2011). Comprehending news videotexts: The influence of the visual content. *Language Learning & Technology, 15*(2), 44–68.

De Gelder, B., & Vroomen, J. (1992). Auditory and visual speech perception in alphabetic and non-alphabetic Chinese-Dutch bilinguals. In *Advances in psychology* (pp. 413–426). https://doi.org/10.1016/s0166-4115(08)61508-3

Deutsch, J.A. & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review*, *70*, 80-90. https://doi.org/10.1037/h0039515

Dupuis, K. (2011). *Bimodal Cueing in Aphasia: The Influence of Lipreading on Speech Discrimination And Language Comprehension*. [Master Dissertation, The University of British Columbia].

Emmert, P. (1994). A definition of listening. *Listening Post, 51,* 6.

Feld, J., & Sommers, M. S. (2009). Lipreading, processing speed, and working memory in younger and older adults. *Journal of Speech, Language, and Hearing Research, 52,* 1555–1565. doi:10.1044/1092-4388(2009/08-0137)

Field, J. (2009). *Listening in the language classroom*. Cambridge University Press. https://doi.org/10.1017/CBO9780511575945

Fitch Hauser, M. (1990). Making sense of data: Constructs, schemas, and concepts. In R. N. Bostrom (Ed.), *Listening behavior: Measurement and application* (pp. 76–90). New York: Guilford.

Florit, E., Roch, M., & Levorato, M. C. (2011). The relationship between listening comprehension of text and sentences in preschoolers: Specific or mediated by lower and higher level components? *Applied Psycholinguistics*, *34*(2), 395–415. https://doi.org/10.1017/s0142716411000749

Ginther, A. (2002). Context and content visuals and performance on listening comprehension stimuli. *Language Testing*, *19*(2), 133–167. https://doi.org/10.1191/0265532202lt225oa

Glenn, E.C. (1989). A Content Analysis of Fifty Definitions of Listening. *International Journal of Listening, 3*(1), 21-31. https://doi.org/10.1207/s1932586xijl0301_3

Goss, B. (1982, April). Listening to language: An information processing perspective.

Paper presented at the Southern Communication Association convention, Hot Springs, AR. Retrieved from http://files.eric.ed.gov/fulltext/ED215379.pdf

Green, K. M., Kuhl, P. K., & Meltzoff, A. N. (1988). Factors affecting the integration of auditory and visual information in speech: The effect of vowel environment. *Journal of the Acoustical Society of America*, *84*(S1), S155. https://doi.org/10.1121/1.2025888

Hagiwara, R. E. (1995). *Acoustic realizations of American /r/ as produced by women and men*. University of California, Los Angeles.

Halone, K., Cunconan, T. M., Coakley, C. G., & Wolvin, A. D. (1998). Toward the establishment of general dimensions underlying the listening process. *International Journal of Listening*, *12*, 12–28. doi:10.1080/10904018.1998.10499016

Hardison, D. M. (1996). Bimodal Speech Perception by Native and Nonnative Speakers of English: Factors Influencing the McGurk Effect. *Language Learning*, *46*(1), 3–73. https://doi.org/10.1111/j.1467-1770.1996.tb00640.x

Hardison, D. M. (2003). Acquisition of second-language speech: Effects of visual cues, context, and talker variability. *Applied Psycholinguistics*, *24*(4), 495–522. https://doi.org/10.1017/s0142716403000250

Hazan, V., Kim, J., & Chen, Y. (2010). Audiovisual perception in adverse conditions: Language, speaker and listener effects. Speech Communication, 52(11–12), 996–1009. https://doi.org/10.1016/j.specom.2010.05.003

Hazan, V., Sennema, A., Faulkner, A., Ortega-Llebaria, M., Iba, M., & Chung, H. (2006). The use of visual cues in the perception of non-native consonant contrasts. *The Journal of the Acoustical Society of America*, *119*(3), 1740–1751. https://doi.org/10.1121/1.2166611

Inceoglu, S. (2019). Individual Differences in L2 Speech Perception: The Role of Phonological Memory and Lipreading Ability. *The Modern Language Journal*, *103*(4), 782–799. https://doi.org/10.1111/modl.12591

Janusik, L.A. (2002). *Listening and Cognitive Processing: Is There a Difference?* Panel presentation at the annual conference of the National Communication Association, New Orleans, LA.

Kahneman, D. (1973). *Attention and effort*. *1063*, 218-226. Englewood Cliffs, NJ: Prentice Hall. DOI:10.2307/1421603

Kamiya, N. (2022). The limited effects of visual and audio modalities on second language listening comprehension. *Language Teaching Research*, 136216882210962. https://doi.org/10.1177/13621688221096213

Kelly, C. M. (1975). Empathic listening. In R. L. Applbaum, O. O. Jenson, & R. Carroll (Eds.), *Speech communication: A basic anthology* (pp. 115–123). New York: Macmillan.

Kim, Y. S., & Phillips, B. (2014). Cognitive Correlates of Listening Comprehension. *Reading Research Quarterly*, *49*(3), 269–281. https://doi.org/10.1002/rrq.74

Kim, Y. S. G. (2016). Direct and mediated effects of language and cognitive skills on comprehension of oral narrative texts (listening comprehension) for children. *Journal of Experimental Child Psychology, 141*, 101–120. https://doi.org/10.1016/j.jecp.2015.08.003

Kim, Y. S. G., & Pilcher, H. (2016). What is listening comprehension and what does it take to improve listening comprehension? In R. Schiff & M. Joshi (Eds.), *Handbook of interventions in learning disabilities* (pp. 159-174). New York: Springer.

Kkese, E., & Dimitriou, D. (2023). Audio-visual speech perception of plosive consonants by CG learners of English. Journal of Monolingual and Bilingual Speech, 5(1), 1–28. https://doi.org/10.1558/jmbs.23017

Knecht, S., Deppe, M., Dräger, B., Bobe, L., Lohmann, H., Ringelstein, E., & Henningsen, H. (2000). Language lateralization in healthy right-handers. Brain, 123(1), 74–81. https://doi.org/10.1093/brain/123.1.74

Kogan, V. V. (2022). The Role of Acoustic Memory and Phonological Short-Term Memory. In E. Kkese (Ed.). *Speech Perception and Production in L2* (p. 236). Cambridge Scholars Publishing.

Lepola, J., Lynch, J., Laakkonen, E., Silvén, M., & Niemi, P. (2012). The Role of Inference Making and Other Language Skills in the Development of Narrative Listening Comprehension in 4-6-Year-Old Children. *Reading Research Quarterly*, *47*(3), 259–282. https://doi.org/10.1002/rrq.020

Li, Y., & Somlak, T. (2017). The effects of articulatory gestures on L2 pronunciation learning: A classroom-based study. *Language Teaching Research, 23*(3), 352–371. https://doi.org/10.1177/1362168817730420

Lipari, L. (2010). Listening, thinking, being. *Communication Theory*, *20*, 348-362. doi:10.1111/j.1468–2885.2010.01366.x

Lively, S. E., Logan, J. R., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America*, *94*(3), 1242–1255. https://doi.org/10.1121/1.408177

Marian, V., Hayakawa, S., Lam, T., & Schroeder, S. (2018). Language Experience Changes Audiovisual Perception. *Brain Sciences, 8*(5), 85. https://doi.org/10.3390/brainsci8050085

McGurk, H., & MacDonald, J. M. (1976). Hearing lips and seeing voices. *Nature, 264*(5588), 746–748. https://doi.org/10.1038/264746a0

Morreale, S., Rubin, R. B., & Jones, E. (Eds.). (1998). *Speaking and listening competencies for college students*. Washington, DC: National Communication Association.

Munhall, K. G., Jones, J., Callan, D. E., Kuratate, T., & Vatikiotis-Bateson, E. (2004). Visual prosody and speech intelligibility. Psychological Science, 15(2), 133–137. https://doi.org/10.1111/j.0963-7214.2004.01502010.x

Nadig, A. (2013). Listening Comprehension. In F. R. Volkmar (Ed.), *Encyclopedia of Autism Spectrum Disorders*. Springer. New York, NY. https://doi.org/10.1007/978-1-4419-1698-3_349

Nichols, R. (1948). Factors in listening comprehension. *Speech Monographs, 15*, 154–163. doi:10.1080/03637754809374953

Olson, K. (2003). LSAT listening assessment: Theoretical background and specifications. *Law School Admission Council (LSAC) Research Report 03–02*. http://www.lsac.org/lsacresources/Research/rr/pdf/RR--03--02.pdf

Oltedal, L., & Hugdahl, K. (2017). Opposite brain laterality in analogous auditory and visual tests. Laterality Asymmetries of Body Brain and Cognition, 22(6), 690–702. https://doi.org/10.1080/1357650x.2016.1269335

Pennington, M. C., & Rogerson-Revell, P. (2019). *English Pronunciation Teaching and Research: Contemporary Perspectives*. Palgrave Macmillan London. https://doi.org/10.1057/978-1-137-47677-7

Purdy, M. (2006, April). Listening and gender: Stereotypes and explanations. Paper presented at the International Listening Association annual convention, Salem, OR.

Rankin, P. T. (1926). The measurement of the ability to understand spoken language. Dissertation Abstracts, 12(1952), 847.

Rost, M. (2011). *Teaching and researching listening* (2nd ed.). Harlow: Longman.

Rubin, J. (1994). A Review of Second Language Listening Comprehension Research. *The Modern Language Journal*, *78*(2), 199–221. https://doi.org/10.2307/329010

Rubin, R. B., & Feezel, J. D. (1986). A needs assessment of research in communication education. *Central States Speech Journal, 37*, 113–118. doi:10.1080/10510978609368210

Sabourin, L., Leclerc, J. C., Lapierre, M., Burkholder, M., & Brien, C. (2016). The language background questionnaires in L2 research: Teasing apart the variables. In L. Hracs (ed.), *Proceedings of the 2016 Annual Conference of the Canadian Linguistics Association*. Calgary, AB. https://cla-acl.ca/pdfs/actes-2016/Sabourin_etal_CLA2016_proceedings.pdf

Safronova E. (2016). *The role of cognitive ability in the acquisition of second language perceptual phonological competence*. PhD dissertation. Barcelona, Spain: University of Barcelona.

Sekiyama, K. (1994). Differences in auditory-visual speech perception between Japanese and Americans: McGurk effect as a function of incompatibility. *Journal of the Acoustical Society of Japan (E), 15*(3), 143-158.

Sekiyama, K. (1997). Cultural and linguistic factors in audiovisual speech processing: The McGurk effect in Chinese subjects. *Perception & Psychophysics 59*, 73–80. https://doi.org/10.3758/BF03206849

Sekiyama, K., & Tohkura, Y. (1991). McGurk effect in non-English listeners: Few visual effects for Japanese subjects hearing Japanese syllables of high auditory intelligibility. *The*

Journal of the Acoustical Society of America, *90*(4), 1797–1805. https://doi.org/10.1121/1.401660

Sha, Z., Pepe, A., Schijven, D., Carrión-Castillo, A., Roe, J. M., Westerhausen, R., Joliot, M., Fisher, S. E., Crivello, F., & Francks, C. (2021). Handedness and its genetic influences are associated with structural asymmetries of the cerebral cortex in 31,864 individuals. Proceedings of the National Academy of Sciences, 118(47). https://doi.org/10.1073/pnas.2113095118

Spehar, B., Goebel, S., & Tye-Murray, N. (2015). Effects of Context Type on Lipreading and Listening Performance and Implications for Sentence Processing. *Journal of speech, language, and hearing research: JSLHR, 58*(3), 1093–1102. https://doi.org/10.1044/2015_JSLHR-H-14-0360

Spitzberg, B., & Cupach, W. R. (1984). *Interpersonal communication competence*. Beverly Hills, CA: Sage.

Steil, L.K., Barker, L.L., & Watson, K.W. (1983). *Effective listening: Key to your success*. New York: Random House.

Strand, J., Cooperman, A., Rowe, J., & Simenstad, A. (2014). Individual Differences in Susceptibility to the McGurk Effect: Links with lipreading and detecting audiovisual incongruity. *Journal of Speech, Language, and Hearing Research, 57*(6), 2322–2331. https://doi.org/10.1044/2014_jslhr-h-14-0059

Suvorov, R. (2009). Context visuals in L2 listening tests: The effects of photographs and video vs. audio-only format. In Chapelle, C., Jun, H.G., & I. Katz (Eds.), *Developing and evaluating language learning materials* (pp. 53–68). Iowa State University.

Sypher, B. D., Bostrom, R. N., & Seibert, J. H. (1989). Listening, communication abilities, and success at work. *Journal of Business Communication, 26,* 293–303. doi:10.1177/002194368902600401

Treisman, A.M. (1960). Contextual cues in selective listening. *The Quarterly Journal of Experimental Psychology*, 12, 242-248. https://doi.org/10.1080/17470216008416732

Treisman, A.M. (1964). Selective attention in man. *British Medical Bulletin*, 20(1), 12-16. https://doi.org/10.1093/oxfordjournals.bmb.a070274

Trenholm, S. & Jensen, A. (2011). *Interpersonal Communication* (7th ed.). Oxford University Press. https://doi.org/10.1111/j.1468-2958.1994.tb00342.x

Tucker, W. (1925, April). Science of listening. *19th Century, 97,* 548–557.

Tyagi, B. (2013). Listening: An Important Skill and Its Various Aspects. *The Criterion: An International Journal in English, 4*(12), 1–8. https://the-criterion.com/V4/n1/Babita.pdf

Voyer, D., & Boudreau, V. G. (2003). Cross-modal correlation of auditory and visual language laterality tasks: A serendipitous finding. Brain and Cognition, 53(2), 393–397. https://doi.org/10.1016/s0278-2626(03)00152-0

Wagner, E. (2002). Video listening tests: A pilot study. *Working Papers in TESOL & Applied. Linguistics*, *2*(1), 1-39. https://doi.org/10.7916/salt.v2i1.1651

Wagner, E. (2007). Are they watching? Test-taker viewing behavior during an L2 video listening test. *Language Learning & Technology*, *11*(1), 67–86. http://dx.doi.org/10125/44089

Wang, Y., Behne, D. M., & Jiang, H. (2009). Influence of native language phonetic system on audio-visual speech perception. *Journal of Phonetics, 37*(3), 344–356. https://doi.org/10.1016/j.wocn.2009.04.002

Wanzer, M. B., Booth-Butterfield, M., & Gruber, K. (2004). Perceptions of health care providers' communication: Relationships between patient-centered communication and satisfaction. *Health Care Communication, 16,* 363–384. doi:10.1207/ S15327027HC1603_6

Weaver, C.H. (1972). *Human listening: Processes and behavior*. Indianapolis, IN: Bobbs-Merrill.

Wellmon, T. A. (1988). Conceptualizing organizational communication competence: A rules-based perspective. *Management Communication Quarterly, 1,* 515–534. doi:10.1177/0893318988001004004

Wiemann, J. M., & Backlund, P. (1980). Current theory and research in communicative competence. *Review of Educational Research, 50*, 185 – 199. doi:10.3102/00346543050001185

Wilson, S. R., & Sabee, C. M. (2003). Explicating communicative competence as a theoretical term. In J. O. Greene & B. R. Burleson (Eds.), *Handbook of communication and social interaction skills* (pp. 3 – 50). Mahwah, NJ: Erlbaum.

Wolvin, A. D. (1989). Models of the listening process. In C. W. Roberts & K. W. Watson (Eds.), *Intrapersonal communication processes: Original essays* (pp. 508–527).New Orleans, LA: SPECTRA.

Wolvin, A. D., & Coakley, C. G. (1996). *Listening* (5th ed.). Dubuque, IA: Brown & Benchmark.

Worthington, D. L. (2001). Exploring juror's listening processes: The effect of listening style preference on juror decision making. *International Journal of Listening, 15,* 20–35. doi:10.1080/10904018.2001.10499043

Worthington, D. L. (2017). Modeling and Measuring Cognitive Components of Listening. In D. L. Worthington & G. D. Bodie (Eds.), *The Sourcebook of Listening Research: Methodology and Measures* (1st ed., pp.70-96). https://doi.org/10.1002/9781119102991.ch4

Worthington, D. L., & Bodie, G.D (2017). Defining Listening A Historical, Theoretical, and Pragmatic Assessment. In D. L. Worthington & G. D. Bodie (Eds.), *The Sourcebook of Listening Research: Methodology and Measures* (1st ed., pp.3-17). https://doi.org/10.1002/9781119102991.ch1

APPENDIX A

BACKGROUND QUESTIONNAIRE

Biographical Information

Question 1: Birth year

Question 2: Current age

Question 3: Gender

Question 4: Your native language(s)

Question 5: Mother's native language(s)

Question 6: Father's native language(s)

Question 7: Nationality

Question 8: Occupation

Linguistic Information

Question 1: What is the current proficiency level of your native language? (multiple choice)

Question 2: What other language are you most proficient in? Please give only one answer.

Question 3: How many years have you been studying this foreign language?

Question 4: What is your current proficiency level in this foreign language? (multiple choice)

Question 5: Do you speak more foreign languages? Please provide another foreign language you speak if you do. You may leave it blank if you do not speak more foreign languages.

Question 6: What is your current proficiency level in this foreign language? You may leave this question blank if you do not speak more foreign languages.

Question 7: Please write down the most recent standardized English language test you have taken, its score, and the test date.

APPENDIX B

LISTENING COMPREHENSION MATERIALS

Simple Passages (Dupuis, 2011)

Practice Items

a. Cynthia wants to quit smoking. // She bought a book on how to quit smoking cold turkey.// She threw away the book within 3 days,// and bought some Nicorette gum instead.// She knows herself too well.

(detail) Did Cynthia buy nicotine patches? (no)

(inf.) Was Cynthia able to quit cold turkey? (no)

b. Martha enjoys woodworking.// One morning, a rusty nail went through the wood and into her finger.// Later in the day, her finger became very swollen.// She worried she might be getting an infection.

(detail) Did Martha hurt herself with scissors? (no)

(inf.) Will Martha likely need medical attention? (yes)

Test Items

1. David wanted to go for a mountain bike ride. // It had been raining for six days in a row.// Now, even though it had finally stopped raining, the trails were all muddy.// So, David decided to read a book instead.

(detail) Did David want to go for a run? (no)

(inf.) Had the weather been nice lately? (no)

2. Ben is very nervous. //He has to write a document for work,//and his boss gave him a deadline.//It is far from being done.// Ben decided to stay up all night to finish it.

(detail) Is the document for Ben's work? (yes)

(inf.) Did Ben still have a lot of work to do that night? (yes)

3. Erika lives in Canada.// She went to Greece for her Christmas holidays.// The temperature was between 5 and 10 degrees// and it rained the whole time.// Sadly, she spent almost all of her time in the hotel room.

(detail) Does Erika live in Greece? (no)

(inf.) Did Erika have a good time on her trip? (no)

4. Andrew works as a mechanical engineer.// He has been working for the same company in Vancouver for 17 years.// His employer offered him a higher position in Calgary.// He declined the offer.

(detail) Is Andrew a lawyer? (no)

(inf.) Is Andrew likely happy at his current job location? (yes)

5. Oliver agreed to go on his first blind date.// At the restaurant, he looked for a woman with black hair and a red dress.// He spotted her from behind// and walked over to her table.// He was shocked to see his ex-wife sitting at the table.

(detail) Did the woman have blond hair? (no)

(inf.) Did Oliver already know his blind date? (yes)

6. Allison is sick.// She spent the night vomiting.// She took some medicine// and tried to sleep.// She remembered her dinner had tasted funny last night.

(detail) Did Allison take any medicine? (yes)

(inf.) Did Allison likely catch a cold? (no)

7. Luke loves to play poker.// Unfortunately, he is not a very good player.// However, his friends love to play with him.// They often win a lot of money.

(detail) Does Luke like to play poker? (yes)

(inf.) Does Luke win often at poker? (no)

8. Jill decided to buy a new car for her teenage son.// She found a used Toyota for an excellent price.// She bought it and took it home.// Two weeks later, the car broke down.

(detail) Was the Toyota expensive? (no)

(inf.) Was the car likely in bad condition? (yes)

9. Peter bought a new plant for his office.// He put it on his brand new desk.// On Friday, before heading home, he generously watered his plant.// On Monday, Peter was disappointed to find water stains on his desk.

(detail) Did Peter put his plant on his desk? (yes)

(inf.) Did Peter overwater his plant? (yes)

10. Harold decided to go bungee jumping.// The bungee instructor tied him securely to the body harness and rope// and got him all ready for the big jump.// Harold walked to the edge of the platform.// He screamed and ran back.

(detail) Did Harold wear a body harness? (yes)

(inf.) Was Harold brave enough to jump? (no)

11. Claire decided to go back to college.// She wanted to get a nursing degree.// She applied to the Vancouver Community College,// but she was too late.// The program was already full.

(detail) Did Claire want to get a degree in nursing? (yes)

(inf.) Was Claire able to enroll in the program? (no)

12. Sally loves to cook.// Yesterday, she decided to make a lemon pie.// Her husband tasted it and made a funny face.// Sally had forgotten to add sugar.

(detail) Did Sally forget to add flour? (no)

(inf.) Did the pie taste good? (no)

13. Bill visited his sister and brother-in-law yesterday.// They had dinner together.// They looked at family photos and laughed a lot.// He was back home at 2:00 A.M.

(detail) Did Bill go home before midnight? (no)

(inf.) Did Bill enjoy visiting his sister? (yes)

14. Rob and Kristen decided to go camping.// They set up their tent and went fishing.// When they came back, they saw many tears and cuts on their tent.// The campground staff sent out a bear warning.

(detail) Did Rob and Kristen go fishing? (yes)

(inf.) Did a bear damage Rob and Kristen's tent? (yes)

15. Sandra started her new job as a waitress.// Her first night was a disaster.// She dropped a plate full of food// and spilled water on a customer.// She quit her job at the end of her shift.

(detail) Did she spill water on a customer? (yes)

(inf.) Was Sandra clumsy? (yes)

16. Quinn and his friend decided to go fishing.// Quinn bet that he would catch more fish.// They sat by the lake all afternoon.// At the end of the day, neither one had caught anything.

(detail) Did Quinn go fishing with his brother? (no)

(inf.) Did Quinn win the bet? (no)

17. Margaret went to the public library.// She browsed the fiction section,// but could not find the book she wanted.// She asked the librarian for help.// The librarian said the book had been damaged and was no longer available.

(detail) Did Margaret browse the reference section? (no)

(inf.) Did Margaret borrow the book? (no)

18. Little Matthew begged his mom to go to the zoo.// They packed a lunch and headed to the Greater Vancouver Zoo.// Matthew spent 2 hours watching the monkeys.// He begged his mom to buy him a monkey for his next birthday.

(detail) Did Matthew and his mom pack a lunch? (yes)

(inf.) Does Matthew like monkeys? (yes)

19. When Marty got to his car, he noticed a piece of paper on the windshield.// It was a parking ticket.// Marty suddenly realized this was a 30-minute parking stall.// His car had been parked there since morning.

(detail) Did Marty park his car there in the morning? (yes)

(inf.) Was Marty's car parked for less than 30 minutes? (no)

20. Lily tried sushi for the first time yesterday.// She liked the California rolls very much.// She ordered a few more,// and ate her friend's as well.// At the end of the night, she felt very sick.

(detail) Did Lily order California rolls? (yes)

(inf.) Did Lily likely eat too much sushi? (yes)

Intermediate Passages (Dupuis, 2011)

Practice Items

a. When Sean's computer got a virus,// he decided to go to the store and get it fixed right away.// At the store, Sean was told that it would cost him at least $500 to get his computer fixed,//

since the virus had damaged many parts of the system.// Sean thought of buying a new computer instead,// since he could probably get a fairly good computer for the same price.// He therefore decided to buy himself a new laptop for under $500.

(detail) Did the virus damage many parts of the system? (yes)

(inf.) Did Sean get his computer fixed? (no)

b. Chad headed to the airport extra early to make sure he wouldn't miss his plane.// When he got to the airport, he realized he had forgotten his passport,// and therefore decided to hurry back home and get it.// After spending a half hour looking for his passport,// he started wondering if perhaps he had dropped it on the floor somewhere at the airport.// Worried he would never find it,// he opened his suitcase to reach for his cell phone and found his passport neatly tucked under his phone.

(detail) Did Chad spend 3 hours looking for his passport? (no)

(inf.) Did Chad have his passport with him all along? (yes)

Test Items

1. On Tuesday morning, Greg headed to the dentist's office for a checkup.// He hadn't been to the dentist in over ten years,// and decided that it was finally time to get a thorough cleaning.// After spending more than 3 hours cleaning his teeth, the staff told Greg that his mouth was full of cavities.// Greg promised himself that he would get a regular checkup every year from now on.

(detail) Did the staff spend more than 3 hours cleaning Greg's teeth? (yes)

(inf.) Did Greg wait too long to get his teeth checked? (yes)

2. Every year, the Johnsons plant six different kinds of tomatoes in their garden.// They use them fresh in salads, soups, and sauces in the summer,// and when fall comes they can the rest of the tomatoes for the winter.// Mr. Johnson also makes homemade ketchup and tomato marmalade

which he sells at the local farmers' market for a very reasonable price.// For years, the Johnsons have been known in the neighborhood as the "tomato couple", a name they have proudly embraced.

(detail) Is Mr. Johnson's ketchup expensive? (no)

(inf.) Do the Johnsons like their nickname? (yes)

3. Sasha, a busy 32-year-old businesswoman with 4 young children,// decided to hire a nanny to help out around the house and take care of the kids.// The nanny had been recommended by a friend,// so Sasha felt comfortable hiring her without contacting her references.// After the first two weeks, Sasha started noticing that some items were missing from her bedroom,// including a very expensive pearl necklace.// She decided to call one of her nanny's references,// and soon realized that she had made a big mistake by hiring her.

(detail) Did Sasha's pearl necklace go missing? (yes)

(inf.) Was the nanny likely a thief? (yes)

4. Ever since Sam was a young boy, he has always been a big hockey fan.// As a child, he would dress up in his older brother's hockey shirt and helmet,// go down to the community rink in the winter// and slide down the ice on his boots while pretending to be Wayne Gretsky.// Today, Sam's wife says that things haven't really changed.// He still dresses up in his Canucks jersey,// drives down to GM place during NHL seasons,// and sits in the stands, eats popcorn, and cheers for his team.

(detail) Did Sam used to pretend he was Bobby Orr? (no)

(inf.) Is Sam an NHL hockey player? (no)

5. Recently, Geoff noticed that his vision had become increasingly blurry,// and thought that it might be time to get his eyes checked again and perhaps get a new pair of glasses.// During his appointment, the optometrist told Geoff that his vision had actually gotten better.// Relieved, he

bought new frames and lenses,// went home, and happily told his wife that his vision had improved.// She smiled and said that it was great news,// but couldn't help laughing at his new glasses and told him they were way too big for his face.

(detail) Did Geoff's vision get worse? (no)

(inf.) Did Geoff's wife like his new glasses? (no)

6. Every year, Kim worries about her teenage son Jake, especially when winter comes around.// Jake loves snowboarding,// but often chooses to go snowboarding without a helmet, even though he is well aware of the risks involved.// He insists that he is an excellent snowboarder// and that nothing will ever happen to him because he doesn't do any of the risky moves many of his friends do.// Kim is convinced she will one day get a phone call telling her that her son is in critical condition at the hospital.

(detail) Does Jake always wear a helmet? (no)

(inf.) Did Jake have a serious snowboarding accident? (no)

7. The Browns decided it was time to redecorate their living room,// which of course inevitably involved changing the colors of the walls.// They went to Home Depot// and picked 2 new colors for their living room,// a dark red and a light forest green.// They started painting the next day,// immediately pleased with the results of the dark red color.// When both colors were applied, however,// they were disappointed to see that the combination of red and green made the room look like it was decorated for the Christmas holidays.

(detail) Did the Browns buy two different colors of paint? (yes)

(inf.) Did the Browns like the look of their new living room? (no)

8. Although Anne wasn't supposed to give birth until March,// she went into labour one month earlier while making dinner.// Panicked, she called her husband at work, who rushed home as soon

as possible.// They hopped into their car and drove to the hospital,// worried that they wouldn't make it in time.// Anne gave birth to a healthy baby boy less than 20 minutes after arriving at the hospital.

(detail) Did Anne go into labor while eating breakfast? (no)

(inf.) Did Anne and her husband arrive at the hospital in time? (yes)

9. Cassie decided that it was time for her to learn how to be a good photographer,// and therefore bought herself a new high-end digital camera.// She started practicing using her camera by going out to the park and taking pictures of people, animals, trees, and flowers.// She tried the many different buttons and options// but remained unhappy with the results.// She therefore decided to switch to the automatic mode// and was instantly pleased with how her pictures turned out.

(detail) Did Cassie practice taking pictures in the park? (yes)

(inf.) Was Cassie skilled in using all the different options on her camera? (no)

10. Every Thursday night, Loretta goes to the community center to play bingo.// She goes partly because she enjoys playing bingo with her friends,// but also because she secretly hopes to win money or some of the other prizes offered.// Last Thursday, Loretta almost won twice that night,// but every time, someone else was one step ahead of her and claimed the prize.// Because next week is Loretta's birthday, she strongly believes it will be her lucky week,// and she'll win for the first time since she started playing at the community center.

(detail) Does Loretta play bingo with her friends? (yes)

(inf.) Has Loretta ever won at bingo at the community center? (no)

11. On Saturday, Mike decided to go to the pet store and buy a puppy for his seven-year-old daughter's birthday.// Unfortunately, the only puppy available at the store was a young bulldog.//

Mike knew that this was not the type of dog his daughter was hoping to get,// but he also knew that she would be very upset if she didn't get a pet for her birthday.// He then saw a cute orange tabby kitten and hoped that his daughter would be equally happy with a cat.

(detail) Was the only puppy available at the store a bulldog? (yes)

(inf.) Did Mike buy his daughter a cat? (yes)

12. One of Mary Ann's New Year resolutions was to learn to cook.// Last weekend, she bought a new seafood cookbook// and decided to try a very tasty shrimp and scallop soup.// She followed all the steps outlined in the recipe, and her soup looked similar to the picture in the book.// She cautiously took a spoonful,// and decided that her New Year resolution was off to a good start.

(detail) Did Mary Ann buy a new dessert cookbook? (no)

(inf.) Did Mary Ann's soup taste bad? (no)

13. Heather and Karl have always wanted a baby girl,// but were instead blessed with 3 young healthy, and very active boys.// They were considering having another baby and hoping for a girl,// but Karl worried that 4 kids would be too much to handle,// and that they might end up with another boy instead.// Soon after, Heather found out she was pregnant,// and the couple was full of hopes and dreams for this last child.// At the ultrasound a few weeks later, they were shocked to find out they were having twin boys.

(detail) Did the couple already have 4 boys? (no)

(inf.) Will Heather and Karl likely try again for a girl? (no)

14. As a child, Lynn always dreamed of becoming a famous singer and signing autographs for thousands of fans.// At the age of 8, she participated in a singing contest at her school,// but unfortunately finished second to last.// Determined to make it as a singer and convinced that she had a hidden talent as a performer,// she enrolled in singing classes in her community.// Eleven

years later, she was named the winner of the TV show Canadian Idol// and pursued a very successful career in entertainment.

(detail) Did Lynn win the singing contest at her school when she was 8 years old? (no)

(inf.) Is Lynn a good singer today? (yes)

15. Janice wondered why her clothes kept disappearing from her closet and reappearing days later.// She sometimes became very upset in the morning when getting ready for work when she couldn't find the clothes or shoes she was looking for.// She initially blamed her teenage daughter Katherine,// who firmly denied it and insisted she would never wear those types of clothes.// One day when shopping at the mall, Janice spotted her daughter with her friends on the other side of the store,// and realized that her daughter had lied to her about sneaking into her closet.

(detail) Did Janice spot her daughter with her friends in the park? (no)

(inf.) Was Katherine taking her mom's clothes without her permission? (yes)

16. Tania had been wanting to go work in Mexico for a few months,// and decided that this year was finally the year she would go.// To get ready for this new adventure,// she enrolled in Spanish classes and familiarized herself with authentic Mexican dishes.// After a month of eating spicy tacos and enchiladas and struggling with her Spanish,// she decided to rethink her decision to go to Mexico.// Perhaps Australia would be a better option, she decided.

(detail) Was Tania planning on going to Mexico to work? (yes)

(inf.) Did Tania feel at home eating Mexican food and speaking Spanish? (no)

17. Fred's dog has always been terrified of thunderstorms.// When Fred first adopted him 9 years ago, the little puppy would run under the kitchen table and bark until the storm passed,// unable to stop shaking.// Over the years, Fred's dog learned that thunder was not as threatening as

it seemed,// and that it never actually hurt him.// Now, when a thunderstorm starts, he sits in the kitchen, never barks,// but still cannot stop himself from shaking.

(detail) Does Fred's dog sit in the kitchen during thunderstorms? (yes)

(inf.) Does Fred's dog still fear thunderstorms as much as when he was little? (no)

18. For the past few months, Blake had been trying to convince his wife to switch to satellite TV,// insisting that it wasn't too expensive and that they would get a lot of different channels.// When Blake's wife finally agreed,// he immediately called Star Choice and got it all set up.// Thrilled, he spent the next 3 weeks sitting on the couch, watching sports, movies, and reality TV shows.// After trying unsuccessfully to get her husband off the couch,// Blake's wife called Star Choice and canceled their contract.

(detail) Did Blake spend a lot of time watching sports? (yes)

(inf.) Did Blake's wife regret getting satellite TV? (yes)

19. One morning while making his bed, Lionel noticed tiny black dots on his mattress.// Confused, he picked up a flashlight to take a closer look,// and realized that they were moving around on his mattress and bedsheets.// He let out a scream,// took a step back,// and picked up the telephone to call the exterminator.// In the end, it took the exterminator close to a month to get rid of all the bedbugs that had taken over Lionel's bed, couches, and carpet.

(detail) Were the dots on Lionel's bed moving? (yes)

(inf.) Were the bedbugs difficult to get rid of? (yes)

20. Althea, a high school English teacher, was very excited about the new pants she had bought on sale at the mall the previous weekend.// Yesterday, she decided to wear them for the first time.// She headed to school and walked into her classroom,// put her bag down by her desk, and headed over to the blackboard to write the daily schedule.// One girl raised her hand and told Althea that

she had forgotten to take the price tag off her new pants,// and the other students started laughing when Althea started blushing.

(detail) Did a boy in the class tell Althea about the price tag? (no)

(inf.) Was Althea embarrassed? (yes)

Complex Passages (Kamiya, 2022)

Passage 2

The Buy Nothing Movement

Social media, magazines, and shop windows bombard people daily with things to buy, and British consumers are buying more clothes and shoes than ever before. Online shopping means it is easy for customers to buy without thinking, while major brands offer such cheap clothes that they can be treated like disposable items – worn two or three times and then thrown away.

In Britain, the average person spends more than £1,000 on new clothes a year, which is around four percent of their income. That might not sound like much, but that figure hides two far more worrying trends for society and for the environment. First, a lot of that consumer spending is via credit cards. British people currently owe approximately £670 per adult to credit card companies. That is 66 percent of the average wardrobe budget. Also, not only are people spending money they do not have, they are using it to buy things they do not need. Britain throws away 300,000 tons of clothing a year, most of which goes into landfill sites.

People may not realize they are part of the disposable clothing problem because they donate their unwanted clothes to charities. But charity shops cannot sell all those unwanted clothes. 'Fast fashion' goes out of fashion as quickly as it came in and is often too poor quality to recycle; people do not want to buy it second-hand. Huge quantities end up being thrown away, and a lot of clothes

that charities cannot sell are sent abroad, causing even more economic and environmental problems.

However, a different trend is springing up in opposition to consumerism – the 'buy nothing' trend. The idea originated in Canada in the early 1990s and then moved to the US, where it became a rejection of the overspending and overconsumption of Black Friday and Cyber Monday during Thanksgiving weekend. On Buy Nothing Day people organize various types of protests and cut up their credit cards. Throughout the year, Buy Nothing groups organize the exchange and repair of items they already own.

The trend has now reached influencers on social media who usually share posts of clothing and make-up that they recommend for people to buy. Some YouTube stars now encourage their viewers not to buy anything at all for periods as long as a year. Two friends in Canada spent a year working towards buying only food. For the first three months, they learned how to live without buying electrical goods, clothes, or things for the house. For the next stage, they gave up services, for example, haircuts, eating out at restaurants, or buying petrol for their cars. In one year, they had saved $55,000.

The changes they made meant too fewer cars on the roads, a reduction in plastic and paper packaging, and a positive impact on the environment from all the energy saved. If everyone followed a similar plan, the results would be impressive. But even if you cannot manage a full year without going shopping, you can participate in the anti-consumerist movement by refusing to buy things you do not need. Buy Nothing groups send a clear message to companies that people are no longer willing to accept the environmental and human cost of overconsumption.

1. What are some of the reasons why people buy things that <u>are NOT listed?</u>

   A. Internet.

B. Window shopping.

C. Commercial.

D. Magazine.

2. Which of the following is NOT TRUE in terms of the consumption activities of the British?

A. Spends 4% of the income on clothes.

B. Many people buy with credit cards.

C. Throw away 66% of their clothes.

D. Spends £1,000 on clothes every year.

3. What happens to 300,000 tons of clothes every year in the UK?

A. It has been donated.

B. Discarded.

C. It is being bought.

D. Sold.

4. What activities does the Buy Nothing Group NOT do?

A. Destruction of credit cards.

B. Protests.

C. Replacement or repair of goods.

D. Donation of clothes abroad.

5. What did the two Canadians NOT do in the first three months?

A. Don't buy things for home.

B. Don't drive a car.

C. Don't buy appliances.

D. Don't buy clothes.

6. Which one does the speaker propose?

   A. Not shopping for 1 year.

   B. Joining the Buy Nothing Group.

   C. Protest against companies.

   D. Don't buy things you don't need.

Answers

1. C: Commercials.
2. C: Throw away 66% of their clothes.
3. B: Discarded.
4. D: Donation of clothes abroad.
5. B: Don't drive a car.
6. D: Don't buy unnecessary things.

Passage 3

The Sharing Economy

If we look around us at the things we have purchased at some point in our lives, we would no doubt notice that not everything we own is being put to good use: the thick woolen coat which we thought looked trendy despite the fact that we live in a tropical country, the smartphone that got put away when we bought ourselves the newest model, the car that only gets used at the weekends, or even the guest room in our house that somehow got turned into a storeroom.

Those underutilized items may seem useless to some but could be an asset to others. With the advent of the internet, online communities have figured out a way to generate profit from the sharing of those underused assets. Using websites and social media groups that facilitate the buying

and selling of second-hand goods, it is now easier than ever for peer-to-peer sharing activities to take place. And this is known as the sharing economy.

These democratized online platforms are providing a chance for people to make a quick buck or two. To give an example, busy parents previously might not have bothered with setting up a stall at the local market or car boot sale to sell their children's old equipment, but with online marketplaces, parents are now able to sell on those hardly worn baby clothes that their children have outgrown, the expensive pushchairs and baby toys and foods they have invested in, so as to put some cash back into their pockets.

Businesses have also caught on to the profitability of the sharing economy and are seeking to gain from making use of those underutilized resources. A business model that has rapidly risen in popularity sees companies providing an online platform that puts customers in contact with those who can provide a particular product or service. For example, Uber encourages people to use their own personal cars as taxis to make some extra cash in their free time.

This move towards a sharing economy is not without criticism. Unlike businesses, unregulated individuals do not have to follow certain regulations and this can lead to poorer and inconsistent quality of goods and services and a higher risk of fraud. Nevertheless, in the consumerist society we live in today, the increased opportunities to sell on our unwanted and underused goods can lead to a lesser impact on our environment.

1. The sharing economy does not involve …

    A. people selling their used things to others.

    B. people offering their services to others.

    C. businesses selling their goods to people.

D. businesses acting as a middleman for people who want to sell a product and people who want to buy it.

2. People can now sell things more easily because …

   A. people nowadays buy more things.

   B. businesses want to buy the things they don't use.

   C. there are now more market stalls and car boot sales.

   D. there are now online platforms where they can meet people who want to buy their goods and services.

3. Parents might want to sell their baby clothes and baby equipment because …

   A. they want to make back some of the money they spent on those baby purchases.

   B. they don’t like the baby items they have bought.

   C. the baby clothes and equipment are old and worn out.

   D. they need the money for other investments.

4. Which of these is something that the author says we might underutilize?

   A. A thick coat in a cold country

   B. The latest smartphone

   C. Clothes our babies don't or can't wear anymore

   D. The storeroom in our house

5. It might be a problem for unregulated individuals to sell to others because …

   A. they have to follow certain regulations.

   B. what they sell might be of a lower quality.

   C. they don't have a business license.

   D. they like to criticize their buyers.

6. What might be a good title for this article?

   A. The consumerist society

   B. Parents who need money

   C. The rise of the sharing economy

   D. Why we buy things we don't need

Answers

1. C. businesses selling their goods to people.
2. D. there are now online platforms where they can meet people who want to buy their goods and services.
3. A. they want to make back some of the money they spent on those baby purchases.
4. C. Clothes our babies don't or can't wear anymore
5. B. what they sell might be of a lower quality.
6. C. The rise of the sharing economy

Passage 4

A Threat to Bananas

In the 1950s, Central American commercial banana growers were facing the death of their most lucrative product, the Gros Michel banana, known as Big Mike. And now it is happening again to Big Mike's successor – the Cavendish.

With its easily transported, thick-skinned, and sweet-tasting fruit, the Gros Michel banana plant dominated the plantations of Central America. United Fruit, the main grower and exporter in South America at the time, mass-produced its bananas in the most efficient way possible: it cloned shoots from the stems of plants instead of growing plants from seeds and cultivated them in densely packed fields.

Unfortunately, these conditions are also perfect for the spread of the fungus *Fusarium oxysporum f. sp. cubense*, which attacks the plant's roots and prevents it from transporting water to the stem and leaves. The TR-1 strain of the fungus was resistant to crop sprays and traveled around on boots or the tires of trucks, slowly infecting plantations across the region. In an attempt to escape the fungus, farmers abandoned infected fields, flooded them, and then replanted crops somewhere else, often cutting down rainforest to do so.

Their efforts failed. So, instead, they searched for a variety of bananas that the fungus did not affect. They found the Cavendish, as it was called, in the greenhouse of a British duke. It was not as well suited to shipping as the Gros Michel, but its bananas tasted good enough to keep consumers happy. Most importantly, TR-1 did not seem to affect it. In a few years, United Fruit had saved itself from bankruptcy by filling its plantations with thousands of the new plants, copying the same monoculture growing conditions Gros Michel had thrived in.

While the operation was a huge success for the Latin American industry, the Cavendish banana itself is far from safe. In 2014, Southeast Asia, another major banana producer, exported four million tons of Cavendish bananas. But, in 2015, its exports had dropped by 46 percent thanks to a combination of another strain of the fungus, TR-4, and bad weather.

Growing practices in South East Asia have not helped matters. Growers cannot always afford the expensive lab-based methods to clone plants from shoots without spreading the disease. Also, they often are not strict enough about cleaning farm equipment and quarantining infected fields. As a result, the fungus has spread to Australia, the Middle East and Mozambique – and Latin America, heavily dependent on its monoculture, Cavendish crops, could easily be next.

Racing against the inevitable, scientists are working on solving the problem by genetically modifying the Cavendish with genes from TR-4-resistant banana species. Researchers at the

Queensland University of Technology have successfully grown two kinds of modified plant which have remained resistant for three years so far. But some experts think this is just a sophisticated version of the same temporary solution the original Cavendish provided. If the new bananas are planted in the same monocultures as the Cavendish and the Gros Michel before it, the risk is that another strain of the disease may rise up to threaten the modified plants too.

1. Mass-produced bananas are …
    A. grown from seeds because it's efficient.
    B. cloned because it is a fast and cheap way to grow them.
    C. sweeter than other bananas.
    D. exported to Central America.
2. The spread of the TR-1 strain was …
    A. caused by lack of water.
    B. speeded up by the flooding of banana fields.
    C. slowed down by crop spraying.
    D. helped by the movement of people and vehicles.
3. Which sentence is NOT true?
    A. The Cavendish replaced the Gros Michel.
    B. The Cavendish bananas were easier to transport than the Gros Michel.
    C. The Cavendish was resistant to the fungus.
    D. The Cavendish stopped United Fruit from losing more money.
4. South East Asia's Cavendish exports fell in 2015 because …
    A. a new strain of the fungus has developed.
    B. farmers can't afford new farming technology.

C. they had to quarantine their fruit.

D. they depended too much on other countries.

5. Genetically modifying bananas may …

A. mean farmers can grow the Gros Michel again.

B. cause farmers to repeat the mistakes of the past.

C. encourage farmers to try new growing methods.

D. only be a short-term solution.

6. How would you describe the writer's opinion about the future of the Cavendish?

A. Optimistic

B. Pessimistic

C. Cautious

D. Uninterested

Answers

1. B. cloned because it is a fast and cheap way to grow them.
2. D. helped by the movement of people and vehicles
3. B. The Cavendish bananas were easier to transport than the Gros Michel.
4. A. A new strain of the fungus has developed.
5. D. only be a short-term solution.
6. C. Cautious

Passage 5

Cultural Behavior in Business

Much of today's business is conducted across international borders, and while the majority of the global business community might share the use of English as a common language, the

nuances and expectations of business communication might differ greatly from culture to culture. A lack of understanding of the cultural norms and practices of our business acquaintances can result in unfair judgments, misunderstandings, and breakdowns in communication. Here are three basic areas of differences in the business etiquette around the world that could help stand you in good stead when you next find yourself working with someone from a different culture.

First, addressing someone. When discussing this topic in a training course, a German trainee and a British trainee got into a hot debate about whether it was appropriate for someone with a doctorate to use the corresponding title on their business card. The British trainee maintained that anyone who wasn't a medical doctor expecting to be addressed as 'Dr.' was disgustingly pompous and full of themselves. The German trainee, however, argued that the hard work and years of education put into earning that PhD should give them full rights to expect to be addressed as 'Dr'.

This stark difference in opinion over something that could be conceived as minor and thus easily overlooked goes to show that we often attach meaning to even the most mundane practices. When things that we are used to are done differently, it could spark the strongest reactions in us. While many Continental Europeans and Latin Americans prefer to be addressed with a title, for example, Mr. or Ms., and their surname when meeting someone in a business context for the first time, Americans, and increasingly the British, now tend to prefer using their first names. The best thing to do is to listen and observe how your conversation partner addresses you and, if you are still unsure, do not be afraid to ask them how they would like to be addressed.

Second, smiling. A famous Russian proverb states that 'a smile without reason is a sign of idiocy' and a so-called 'smile of respect' is seen as insincere and often regarded with suspicion in

Russia. Yet in countries like the United States, Australia, and Britain, smiling is often interpreted as a sign of openness, friendship, and respect, and is frequently used to break the ice.

In a piece of research done on smiles across cultures, the researchers found that smiling individuals were considered more intelligent than non-smiling people in countries such as Germany, Switzerland, China, and Malaysia. However, in countries like Russia, Japan, South Korea, and Iran, pictures of smiling faces were rated as less intelligent than the non-smiling ones. Meanwhile, in countries like India, Argentina, and the Maldives, smiling was associated with dishonesty.

Third, eye contact. An American or British person might be looking their client in the eye to show that they are paying full attention to what is being said, but if that client is from Japan or Korea, they might find the direct eye contact awkward or even disrespectful. In parts of South America and Africa, prolonged eye contact could also be seen as challenging authority. In the Middle East, eye contact across genders is considered inappropriate, although eye contact within a gender could signify honesty and truthfulness.

Having an increased awareness of the possible differences in expectations and behavior can help us avoid cases of miscommunication, but it is vital that we also remember that cultural stereotypes can be detrimental to building good business relationships. Although national cultures could play a part in shaping the way we behave and think, we are also largely influenced by the region we come from, the communities we associate with, our age and gender, our corporate culture, and our individual experiences of the world. The knowledge of the potential differences should therefore be something we keep at the back of our minds, rather than something that we use to pigeonhole the individuals of an entire nation.

1. The British trainee felt that the people who want to be addressed as 'Dr.' must be …

A. Hard-working.

B. conceited and self-important.

C. doing a medical degree.

D. from Germany.

2. If you are not sure how to address someone, you should …

A. use the title you see on their business card.

B. make your decision based on cultural stereotypes about their country.

C. address them the way you'd like to be addressed.

D. ask them what they would like you to call them.

3. There might be a misunderstanding if an American smiles at a Russian business associate because the Russian might think that the American is ...

A. being fake.

B. challenging their authority.

C. trying to break the ice.

D. disrespectful.

4. The Japanese, South Koreans, and Iranians might interpret a smiling face as being …

A. friendlier.

B. less open.

C. not as intelligent.

D. dishonest.

5. Americans and British people sometimes use eye contact to show that they ...

A. like the speaker.

B. are really listening to what is being said.

C. are honest and truthful.

D. are attending to every need of the speaker.

6. The last paragraph warns the reader not to ...

A. engage in international business.

B. let national cultures shape the way we behave and think.

C. let miscommunication damage our business relationships.

D. overgeneralize using our knowledge of cultural stereotypes.

Answers

1. B. conceited and self-important.
2. D. ask them what they would like you to call them.
3. A. being fake.
4. C. is not as intelligent.
5. B. are really listening to what is being said.
6. D. overgeneralize using our knowledge of cultural stereotypes.