

TEST TAKERS' ATTENTION TO SPEAKERS' NONVERBAL BEHAVIORS IN AN  
INTEGRATED, VIDEO-BASED LISTENING TEST: AN EYE-TRACKING AND  
INTERVIEW STUDY

By

Elena Gorshkova

A THESIS

Submitted to  
Michigan State University  
in partial fulfillment of the requirements  
for the degree of

Teaching English to Speakers of Other Languages – Master of Arts

2024

## **ABSTRACT**

This study investigates what specific nonverbal behaviors L2 language speakers attend to in a video-based listening test. Thirty-two participants from a large Midwestern university were recruited to participate in this study. They were asked to watch three short video-based lectures that contained different types of nonverbal behavior, complete a written free recall immediately after each video, and participate in an individual interview. Eye tracking results demonstrate that participants fixated mostly on the lecturers' faces (71.12%) while watching the video, then eyes (41.48%), mouths (13.63%), with gestures occupying the least of dwell times (2.03%). Interview data combined with free recall scores revealed that participants paid attention to gestures in varying degrees. Different types of nonverbal behaviors differentially affected how the participants engaged with the lectures. The study corroborates others that conclude that nonverbal behavior is a natural part of processing aural information and should be included as a part of L2 listening assessment.

This thesis is dedicated to my family.  
Thank you for always believing in me.

## **ACKNOWLEDGEMENTS**

I would like to express my deepest appreciation and gratitude to Professor Paula Winke for her guidance throughout this research project and beyond. Her expertise and passion for language testing have been a constant source of inspiration. I am also deeply grateful to Professor Charlene Polio for her invaluable support during this academic journey and for her insightful comments during master's thesis defense that helped me strengthen my work.

This project would not have been possible without the constant support of my family and Sam, my partner, who continue to support and encourage me along this journey. For this, I am forever grateful.

## TABLE OF CONTENTS

LIST OF TABLES.....	vi
LIST OF FIGURES.....	vii
INTRODUCTION.....	1
LITERATURE REVIEW.....	2
METHODOLOGY.....	10
DATA ANALYSIS.....	14
RESULTS.....	16
DISCUSSION.....	26
CONCLUSION.....	29
REFERENCES.....	30

## LIST OF TABLES

Table 1. Characteristics of the videos and gestures.....	12
Table 2. Nonverbal behaviors dwell times in percentages.....	16
Table 3. Descriptive statistics of free recall performance.....	24

## LIST OF FIGURES

<b>Figure 1.</b> Lecturer's example of the "skin gesture" .....	18
<b>Figure 2.</b> Participants' perception of the gestures in the videos.....	21
<b>Figure 3.</b> B2 level participants' perception of the gestures in the videos.....	21
<b>Figure 4.</b> C1 level participants' perception of the gestures in the videos.....	22

## INTRODUCTION

For several decades, video has been a popular tool for language learners and educators in the classrooms. As a multimodal tool, it contains a wide range of semiotic resources such as verbal, nonverbal and auditory channels as well as textual information (Perez, 2020). The inclusion of these semiotic resources in the video has long been considered to enhance authenticity especially considering that language is inherently multimodal and that information in real life settings is expressed using different channels (Perniss, 2018). Driven by this argument, several researchers (Wagner 2008, 2010, 2013; Ockey, 2007) in the field of language assessment, in particular, listening assessment advocated for inclusion of nonverbal cues in the test. Most recently, Batty (2021) concluded that test takers paid attention to nonverbal behaviors in different degrees, for example, participants attended to the faces the most, then the eyes and mouth behaviors, with gestures drawing little measurable attention. Negligible attention to gestures is in line with the findings by Gullberg and Holmqvist (2006). At the moment, there are only few studies that have comprehensively investigated test takers' attention to nonverbal behaviors, with the majority of these studies using multiple choice questions as a measure of listening comprehension. In the current study, I aim to explore test takers' attention to nonverbal cues using eye tracking methodology and free recall protocol as a different measure of listening comprehension test to fill in the gaps in the research.



## LITERATURE REVIEW

In this section, I overview three main areas of theory that inform this current study: (1) how nonverbal communication has historically been viewed as part of human communication, (2) benefits and pitfalls of incorporating different types of visuals in the listening assessment. I also review (3) various listening assessment measures. As I will explain below, these areas of the literature influenced the research questions at hand, which appear in the next section after this literature review.

### **Nonverbal communication**

Nonverbal communication describes the act of transmitting messages through nonlinguistic behaviors such as gestures, facial expressions, gaze, touch, and personal space, otherwise known as proxemics (Hall et al., 2019; Burgoon, 1983; Duncan, 1969). Although nonverbal communication does not involve spoken utterances, it is necessary to consider verbal and nonverbal channels together to understand the intended meaning as they are often co-constructed (Hall et al., 2019). Nonverbal behaviors (NVB) are particularly important given the research that suggests that NVB are usually unconscious and less controllable, and therefore, are more accurate in conveying speaker's emotions and feelings especially if the utterance is in conflict with the gestures (Gregersen, 2005; Burgoon et al., 2011). Nonverbal communication includes multiple cue modalities that range from body language to the use of interpersonal space; however, in this paper, the following nonverbal behaviors are considered and discussed in detail: gestures and facial cues.

#### *Gestures*

McNeill (2012, p.4) defined gestures as “expressive actions that enact imagery (not necessarily through hands alone) and are generated during the process of speaking”. Urbanski and Stam (2023) outlined that gestures can involve different body parts such as lips, eyebrows, legs; but it is the

hand or arm movements that are usually studied and referred to as gestures in literature. When talking about different hand movements, it is equally important to draw the distinction between hand actions and hand gestures, both of which have distinct practical and communicative functions, respectively (Bavelas & Chovil, 2006; Krauss et. al., 1996). The communicative functions that gestures perform are well known in the second language acquisition (SLA) literature, yet gestures are also studied in such domains as psychology and neurocognition (Gullberg & McCafferty, 2008). For example, some studies suggest that there is a link between language and action, that is, certain parts of the brain are activated during action-related language processing (Pulvermüller, 2005; Willems & Hagoort, 2007). Other studies (So et al., 2012; Cook et al., 2010) concluded that different types of gestures used in the stimuli facilitate recall of the information. Thus, gestures, irrespective of their form and function, improve the recall of information and facilitate comprehension. Here, I would like to segue to gesture classifications. I will first examine gestures detailed in pioneering work by Ekman and Friesen (1969) and supplement their framework with another. In their work, Ekman and Friesen (1969) identified four major gesture types: emblems, illustrators, regulators, and adapters. Emblems are these types of gestures that have a direct translation. Similar to emblems, illustrators share their functions in awareness and intentionality, but their primary function is in complementing the message and “visually” depicting it. Regulators, such as nods and shifts in posture, maintain conversation and signal listener’s engagement. Lastly, adapters convey emotional states. McNeill (1992, 2005) further classified gestures into the following types: iconic, metaphoric, beat, and deictic. Thus, iconic gestures directly represent the objects, while metaphoric gestures represent the utterances that are abstract. Beat gestures help outline the rhythm of speech, while deictic are the pointing gestures. Crucially, there are other (sub)categories within the gestures, but within the scope of this

paper, only the frameworks introduced by Ekman and Friesen (1969) and McNeill (1992, 2005) are used.

### *Facial cues*

Of all nonverbal behaviors, the face is considered as one of the most dynamic and prominent cues that facilitates communication, provides information about person (from emotional state to personality traits) and is overall a conduit for a lot of information that is communicated nonverbally (Jack & Schyns, 2015; Matsumoto & Hwang, 2013). Therefore, it is not surprising that face and facial behavior in particular is given a lot of attention from the interlocutor or listener (Ekman & Friesen, 1969). In fact, researchers believe that facial behavior sometimes gives more information than certain other gestures. Among such facial behaviors are facial expressions of emotion (affect displays). In general, face serves as a primary display of emotion, although people may consider other paralinguistic and nonverbal cues to understand person's emotional state (Burgoon et al., 2011). Recent scholarship also suggests that facial expressions differ in the way they are perceived and generated by individuals from different cultural groups (Marsh et al., 2003; Elfenbein, 2013). For example, Yuki et al. (2007) found that participants from Japan paid attention to eyes while they were viewing the faces, whereas participants from US interpreted emotions based on the mouth. Apart from affective states, attention to facial cues can show how a person processes information and is engaged with the content. Several eye tracking studies (Vo et al., 2012; Scott et al., 2018) demonstrate that gaze is oriented toward different facial cues and depending on the behavior in the stimuli, such as making eye contact or speaking, participants fixate on the eyes or the mouth respectively. Eyes and eye contact tend to receive much attention from the interlocutor (Burgoon et al., 1984) and in general eyes are more indicative of the person's true intentions in comparison to mouth as the muscles around the eyes are difficult to control (Yuki et al., 2007; Ekman et al.,

1988). Allocation of attention to mouth shows that participants focus on content and the words that are being spoken (Scott et al., 2018). Eye gaze patterns also demonstrate that mouth is looked at more often especially if the language is unfamiliar (Barenholtz et al., 2016) or when auditory information is unavailable (Buchan et al., 2007). All of the literature reviewed in this section demonstrates that nonverbal behaviors are one of the most prominent and important cues in communication that perform different functions.

### **Visual information in listening comprehension**

The importance of incorporating visuals in listening tests has long been argued in the literature, especially when researchers were comparing audio-only and audiovisual listening modes. While some studies found the positive impact of visuals in tests (Sueyoshi & Hardison, 2005; Wagner, 2013), other studies (Suvorov, 2009; Ockey, 2007; Coniam, 2001) concluded that visual information might have a slightly debilitating effect on test takers' performance. For example, the obvious benefits of incorporating video stimuli were found in the work by Sueyoshi and Hardison (2005). The researchers compared students' listening comprehension in three different stimuli (audio only, audiovisual with speaker's facial cues, and audiovisual with speaker's facial cues and gestures). Generally, they concluded that the students who were exposed to visual input performed better on the tests than those who received audio-only input regardless of their level of English proficiency. In line with Sueyoshi and Hardison's (2005) findings, Wagner (2013) found that audiovisual input contributed to listeners' comprehension of the content in the video, resulting in higher scores in the tests that contained visual channel in comparison with audio-only mode. In his earlier study Wagner (2010) reported that test-takers themselves viewed visuals in the listening test positively and considered them as an interesting addition to the test. Several other researchers found that the inclusion of visual stimuli in the listening tests might differentially affect students'

test performance. For example, Ockey (2007), employed still images and video in a listening test and found different degrees of interaction with visual input: still images were helpful to most students at the onset as they provided additional information about the context, but overall students had low engagement with still images. They also differed in the degrees they interacted with video – while some considered it helpful, a few students found the input “constantly distracting.” In a study by Suvorov (2009), test takers’ performance and scores on the multiple choice were lower when they were presented with video-based input as opposed to the scores in the audio- and picture-based test format. Cubilo and Winke (2013) found that test format did not yield significant differences in the test scores, yet participants were in favor of the video lecture and as authors pinpointed, visual input might have helped students understand the content. Similarly, in Hsieh and Davis’ study (2019) participants in the audiovisual and audio-only groups did not differ on the test performance on the following measures: fluency, complexity, and the number of produced in the integrated listen-to-speak TOEFL test. However, authors found a correlation between the test type and students’ proficiency: highly proficient test takers produced more accurate answers than audio-only group with the same proficiency, while participants with low and medium proficiency in the audio-only test format outperformed those in the audiovisual group.

Another field of L2 listening research has focused on the impact of different visuals on students’ test performance and comprehension. For example, Ginther (2002) examined the effects of content (supplemented by visual aids and related to the narrative) and context (actual narrative, without visuals) videos and found that overall content-only videos aided comprehension of the visual input, while context videos had little facilitative effect. As research within attention to visuals in context and content videos started gaining traction, researchers (e.g., Suvorov, 2014) used eye-tracking to quantify students’ attention to visuals and their preference to different types of videos. Suvorov

concluded that students interacted with both content (58%) and context (51%) videos, even though they spent slightly more time watching content videos due to their informative nature. These findings are in line with Ginther's study (2002), who found that students interacted with content videos more, thus making it rather clear that having auditory and visual aids is beneficial in multimodal assessment.

Finally, other researchers examined the impact of nonverbal information in listening tests. For example, Batty (2021) quantitatively measured test takers' attention to nonverbal behaviors and other visual information in a video-based listening test that included two interlocutors (speaker and a listener) and found that test takers paid the most attention to the faces (81.74%) while equally splitting attention between speaker's mouth (20.61%) and eyes (19.89%). In another study, Wagner (2008) found that nonverbal behaviors present in the listening test impacted test takers' comprehension and helped them remember (visually) key gestures. Wagner reported that six participants mentioned one particular salient gesture in the video and three participants referred to this gesture to help them answer comprehension questions. Lastly, test takers in Ockey's study (2007) commented on the presence of nonverbal behaviors in the video stimuli, which ultimately led to different degrees of engagement. For example, two students relied on a lot of visual cues (e.g., lip movement, hand motion, facial gestures) to understand information in the video, while one student did not employ any nonverbal behaviors in doing so.

It seems obvious that the inclusion of any kind of visual information in the listening test does affect test takers' performance and comprehension of the material in different ways, and as a result, researchers diverge in their opinions in incorporating visual information into listening assessment. Buck (2001) wrote that visuals distort the listening construct, that is, instead of processing linguistic information only, test takers take in visual input, which according to Buck might increase

the cognitive load. Other researchers (Wagner 2008, 2010, 2013; Ockey, 2007) have argued for the expansion of the listening construct, in particular, they suggested including nonverbal component in the listening test as a way to increase authenticity of the video material and thus help test takers utilize this real-world information in the test.

### **Measuring listening comprehension**

Listening assessment includes a variety of testing measures. One of the most common tests of listening comprehension is multiple-choice questions. The ubiquity of multiple-choice questions that have been used in large-scale assessments for decades is apparent: they allow for automatic and objective scoring (Rukthong, 2020; Clapham, 2000;) and are usually administered to measure many aspects of the language, including test takers' receptive skills (Brown & Hudson, 1998; McNamara, 2000). Multiple choice questions are also one of the independent forms of testing, which means other skills that can be assessed are kept to a minimum, thus eliminating construct-irrelevant variance (Brunfaut, 2016). However, despite decades of administration in the classrooms and high-stakes environment, multiple choice questions have received criticism, such as the lack of authenticity, test takers' use of test-wise strategies or guessing the answers without necessarily reading or understanding the input (Rukthong, 2020; Cheng, 2004). Other common types of listening assessment include constructed response items such as notetaking, gap filling, sentence completion, short answer responses (Brunfaut, 2016) and selected response items such as true/false and matching questions (Ockey, 2020). Another form of assessment that has gained traction in reading assessment research first and is now slowly implemented in listening assessment - free recall. According to Bernhardt (1991), free recall is a "pure" measure of comprehension and presents itself as an "integrative" task, and as such, it involves several linguistic skills. For example, researchers (Rukthong & Brunfaut, 2019; Brunfaut, 2016;) believe that integrated tasks

are more representative of the real-life communication as listening goes side by side with other skills. With regard to free recall, while some researchers differ in administering it (Ableeva & Lantolf, 2011; Sakai, 2009; Winke & Gass), comprehension is generally measured by the total number of idea units recalled (Riley & Lee, 1996). Importantly, the modality of free recall can be different, that is, it can be done verbally and in written form, and administered immediately after the test or be delayed.

While a lot of studies have been conducted to investigate the role of different types of visual input in the listening assessment, at the time of writing this paper there are only few studies that examine the impact of nonverbal behaviors in the video-mediated listening tests. Additionally, previous studies have primarily utilized multiple choice questions to measure listening comprehension. The current study seeks to examine the attention to nonverbal behaviors in a video-based listening test by employing eye tracking technology and free recall as a different measure of listening comprehension. By using free recall in the study, my hypothesis was that participants will vary in their degree of information recalled and details produced, but this type of test would make the participants pay attention to nonverbal behaviors more, or at least allow them to pay more attention to them. As evidenced by Wagner (2008), test takers made more references to nonverbal behaviors when they were viewing the video as opposed to when they were answering multiple choice questions, thus suggesting that different types of listening assessment types affects the ways participants interact with visual information and what nonverbal behaviors they attend to. With this mind, I seek to answer the following research question in the study:

1. What nonverbal behaviors do test takers pay attention to in a video-based listening test?
2. Do nonverbal behaviors aid the comprehension of lectures?



## METHODOLOGY

### Participants

Thirty-two second language speakers of English (15 male, 17 female), whose ages ranged from 18 to 33 ( $M = 25.15$ ), participated in this study. Out of these, 27 were enrolled in different academic degrees and programs in a large Midwestern university in the United States. The remaining participants were exchange students and visiting scholars at that same university. Participants came from 13 different countries and reported that they studied English since childhood ( $M = 15.34$ ). Their length of stay in the United States varied from 2 months to 10 years. Each participant submitted their language proficiency scores from one of the standardized English proficiency tests (IELTS, TOEFL, Duolingo) that are used for college admissions. Their individual scores were applied to the Common European Framework of Reference (CEFR) concordance table (Duolingo English Test, n.d.), and as a result, three different levels were represented: B2 ( $n=14$ ), C1 ( $n=16$ ), C2 ( $n=2$ ). Due to the nature of research and prior to the data collection, participants were asked to report if they were wearing glasses or contact lenses, or otherwise had any vision problems. Out of 32 participants, two were wearing contact lenses and two were wearing glasses with thin lenses during the experiment. The other three participants reported that they wear glasses for long distance but conducted the experiment without glasses after successful calibration.

### Materials

The instrument consisted of three video lectures taken from an online open resource hosted by Yale University (<https://oyc.yale.edu/courses>). The website contains 22 field-specific courses from different departments with more than twenty lectures per each course and features different modes of information presentation (content and context videos, see Suvorov, 2014). Within the context of this research, there were several criteria for video selection. First, all videos were viewed to

identify the lecturers displayed from head to waist, without the use of different visual aids (e.g., presentation slides, photos) thus allowing participants to see nonverbal behaviors clearly. It was also crucial that the camera stayed fixed on the lecturer to ensure the accuracy of participants' recorded eye movements. Second, I aimed to find two lecturers who used different gestures to varying degrees and one lecturer who hardly used any gestures. Three videos that met these criteria were selected from the following courses: The American Revolution (Freeman, 2010), Frontiers of Biomedical Engineering (Saltzman, 2008), and Epidemics in Western Society since 1600 (Snowden, 2010). A fragment was edited from each video lecture, resulting in three short videos that contained different gestures and sufficient content coverage (enough to be used as isolated texts) for subsequent free recall. In these videos, lecturers relied on notes to present the material, yet largely their speech was organized, fluent, and free of long pauses. These video fragments featured different topics, namely, the life of the president, infrared and ultraviolet radiation, and contagionism theory. Detailed characteristics of the videos and gestures used in them are outlined in Table 1.

To quantify the difficulty of information presented in the videos, transcripts of the videos were uploaded to Lexile Text Analyzer (<https://hub.lexile.com/analyzer>) and the Lexile scores of the texts were mapped onto CEFR concordance scale (Wei & Van Moere, 2021). All three videos fell within B2 level, although the Lexile score of the Epidemics video was slightly higher than the videos about American Revolution and Biomedical Engineering.

Table 1. Characteristics of the videos and gestures

Course title, video content, and its length with the timestamp	Lexile score	Characteristics	Gestures and their count
Video 1 (The American Revolution): The life of the president -- 167 seconds, [12:54-15:41]	1010-1200L	Female lecturer behind the podium; frequent eye contact with the audience; facial expressions	Illustrators (39), Self-adaptors (3), Beat (7), Emblem (1), Deictic (1), Iconic (2)
Video 2 (Frontiers of Biomedical Engineering): Infrared and ultraviolet radiation -- 124 seconds, [7:02-9:06]	1010-1200L	Male lecturer behind the podium; computer in front of him; addresses the audience	Illustrators (7), Iconic (8), Deictic (3), Metaphoric (1)
Video 3 (Epidemics in Western Society since 1600 ): Contagionism theory -- 99 seconds, [14:39-16:18]	1210-1400L	Male lecturer behind the podium; infrequent eye contact	Illustrators (3)

*Note:* The videos were presented in the same order in the experiment as they are in the table.

For the video-based listening test, participants watched each video once and completed a free recall after each video was finished. Similar to Winke and Gass (2016), the instructions for the free recall were as follows:

*Please write down everything you remember from the video and include as many details as you can. There is no time limit.*

## Procedure

After the study was approved by my university's Institutional Review Board (IRB), the materials were piloted with 8 participants, and instructions for the experiment were modified. New instructions ensured that the participants focus on watching the video first and then writing down their recalls. On the data collection day, I conducted individual sessions with each participant in the eye tracking lab. They received consent forms and filled out background questionnaire. In this

study, I used a head-stabilized eye tracker EyeLink 1000 with average accuracy of 0.25° to 0.5° and 1000Hz sampling rate. The participants were seated approximately 60 cm from the monitor with their head positioned in the head mount to reduce movement and potential data loss. They performed a routine calibration check that consisted of a 9-point grid calibration and a fixation cross. The instructions about the study were provided on the screen, but I repeated them in between the videos. Participants were instructed to watch the videos and complete a free recall after each video was finished. While they were watching the videos, I monitored data quality on a separate monitor that was occluded from participants' view and noted down any behavior to bring up in the interview. For the free recall part, I assured participants that I was not testing them on grammatical and lexical accuracy, instead, I wanted to see how much and what they retained from each video. They were not timed and were free to write in their preferred format (e.g., graph, summary, notes). At the end, participants took part in the interview where they were asked about what they noticed about the videos. I audio recorded the interviews and kept the notes of any behaviors that participants reenacted after watching three videos. Once participants completed all stages in the study, they received monetary compensation (40\$) and had a chance to ask about the purpose of the study.

## DATA ANALYSIS

Videos were uploaded to DataViewer (SR Research Ltd., version 4.3.210) for further analysis and data cleaning. I filtered the data by using interest periods, which allowed me to work on eye tracking data from the videos only. I then drew manually elliptical areas of interest (AOIs) for each video. Since the study deals with attention to nonverbal behaviors, I had four predefined AOIs: face, eyes, mouth, and hand gestures. Lastly, I used the mean of dwell times across the four different nonverbal behaviors.

Free recalls were analyzed using the idea units. Following recommendations provided by Winke and Gass (2016) in segmenting the ideas collaboratively, I recruited a rater, who is a recent graduate from a master's program in Applied Linguistics and has experience in rating. Similar to procedures outlined in Winke and Gass (2016) and Riley and Lee (1996), I first segmented the texts into main, supplemental ideas, and details. The rater performed a similar procedure, then we discussed any differences in assigning the ideas to "supplemental" and "details" category. When we reached the agreement, I segmented idea units into a coding sheet. Overall, there were 19 and 18 idea units for the first and the second video respectively, and 13 idea units for the third video. Continuous scores of 0.5 and 1 were assigned to each participant if the idea unit was (partially) represented, the score of 0 was given when no idea unit was written. Similar to Ableeva and Lantolf (2011), logical inferences were not taken into consideration, but paraphrases were. I did not count any grammatical and spelling mistakes. Generally, the participants focused on the content, but some participants included other contextual details. For that purpose, I kept the comment sheet for the participants who mentioned details rather than the content. The rater and I coded 96 free recalls separately. I calculated interrater reliability for each set, and overall Cronbach's alpha value was

.98. Therefore, I used the mean of two scores assigned by myself and the rater for each participant as a measure for listening comprehension.

Thirty-two interviews were transcribed and imported to MAXQDA for thematic analysis. I used predefined themes from the literature, which is called deductive thematic analysis (Braun & Clarke, 2006). However, as I was coding the themes, new themes emerged, which will be covered in the analysis section.

## RESULTS

### *Research question 1*

To answer the first research question which deals with test takers' attention to nonverbal behaviors, two separate sets of data were used: eye tracking metrics and qualitative data from the interviews. Table 2 demonstrates the total dwell times participants spent fixating on areas of interest (AOIs). Overall, face occupied 71.12% of total dwell time in three conditions. The second most visited nonverbal behavior was eyes with a median of 41.48%. Participants fixated on the mouth for 13.63% of the total dwell time and 2.03% on gestures. To put this into perspective, participants directly watched each lecturer's face on average 77.7 seconds, lecturer's eyes for 44.6 seconds, lecturer's mouth for 14.8 seconds, and gestures for 2.1 seconds.

The nonverbal behaviors of the second lecturer were viewed with a higher rate, especially his gestures accounted for 5.21% of the total dwell time. The dwell times of the first video were viewed with a higher rate as well, with the only difference being eyes behavior (36.91%). Given that the lecturer in the first video was making eye contact with the audience more frequently than the lecturer in the third one, the distribution of attention is interesting.

Table 2. Nonverbal behaviors dwell times in percentages

	Gestures	Face	Mouth	Eyes
Video 1	0.75	70.07	13.17	36.91
Video 2	5.21	76.51	14.67	46.59
Video 3	0.15	66.78	13.05	40.94
<i>Mean</i>	2.03	71.12	13.63	41.48

For the specific nonverbal behaviors that participants noticed in these three videos, I turn to thematic analysis of the themes identified within one-on-one interviews. I first detail gestures identified in each video, then I focus on participants' perception of the presence or lack of gestures in the videos. I also focus on participants' perceptions of the facial cues that are present in the

videos. It should be first noted that the interview questions were structured in a way so as not to sway participants to the study's focus and each interview started with general questions. Participants were not asked about gestures until they explicitly mentioned them or at the end of the interview when they did not mention them at all. Thus, 27 participants noticed lecturers' nonverbal behaviors at different parts of the interview process, and they referred to them differently "hand motion," "gestures," "body language," "hand movements," "hand actions". When this number of participants were prompted to recall what types of lecturers' gestures they noticed, 18 of them mentioned particular gestures used in the videos and reenacted them. Five remaining participants did not mention the gestures until they were prompted directly at the end of the interview.

The most common types of gestures recalled were iconic gestures from the second video (Frontiers of Biomedical Engineering) where the lecturer described ultraviolet and infrared radiation. Figure 1 demonstrates one of the gestures that the lecturer was using in the video. Among the iconic gestures that the lecturer used in the video and the participants remembered visually were (numbers in parentheses indicate the number of gestures recalled and reenacted):

- 1) "skin gesture" – the lecturer is touching one hand with another to demonstrate potential effects of sunburn (9 times)
- 2) "heat visualization goggles" – the lecturer cups his hands and shows the gesture resembling binoculars (4 times)
- 3) "satellite pictures" – the lecturer is stretching his hands (3 times)
- 4) "heated objects" – the lecturer places his right hand in front of him, palm down, to show the distribution of heat (2 times)





**Figure 1.** Lecturer’s example of the “skin gesture” (used with permission from Dr. Saltzman, Yale University)

Iconic gestures that the lecturer was using to supplement the content helped participants visualize the information and associate it with gestures. For example, the following participant described his recollection of the first gesture and indicated that this gesture helped him remember the chain of events:

*Whereas in case of wavelength, I just remember one thing that the person who was speaking, he said that, you know, your skin got burned (shows the gesture). And that helped me in remembering that story that, because of, ultraviolet.*

When recollecting the gestures used in the video, another participant remembered not only the gesture but also the number of times it was repeated in the video. Later, he emphasized that the gestures helped him remember and connect information to the spoken utterance:

*When he talked about oil companies...he made like this, this big, wide gesture, (shows the gesture) and I think he repeated it three times because he, he didn't finish the sentence, but the gesture, so he started all over again and did it three times.*

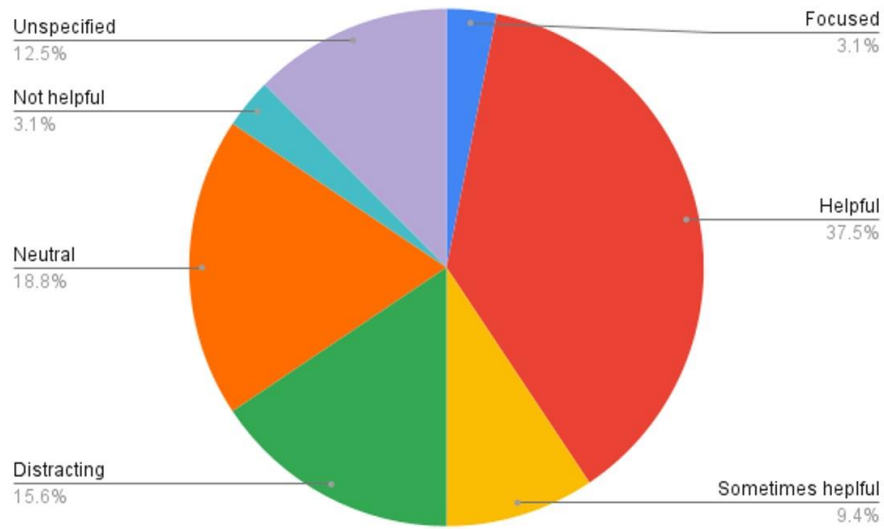
The gestures of the first lecturer (American Revolution lecture) did not draw much attention. Even though her speech was abundant with different types of gestures, largely, there were a lot of illustrator gestures that supplemented the content, but would have no meaning if they were separated from the actual words. There were only three participants who reported and reenacted her illustrator gestures. There was one iconic behavior when the lecturer described the behavior of the person hitting the pen repeatedly on the table and that was visually salient to two participants: *She used her pen to hit on the table (shows the gesture).* Interestingly, lecturer's self-adaptor gestures of adjusting her hair were also noticeable to participants as one provided the following interpretation of the gesture:

*She was doing her hair a lot (shows the gesture). It's a gesture that seems more personal because like when you're trying to explain something with your hands, it's just like for the other person. But when you're replacing hair, like she's careful about what she looks like, she and stuff like that.*

Lastly, the lecturer in the third video hardly had any hand movements and it was a stark contrast between the first two videos. In addition to outlining the differences in presenting information, the participants explicitly mentioned that they would prefer for this lecturer to have gestures while explaining the content. One participant referenced commonly presented virus videos that are animated and gave suggestions for the current video:

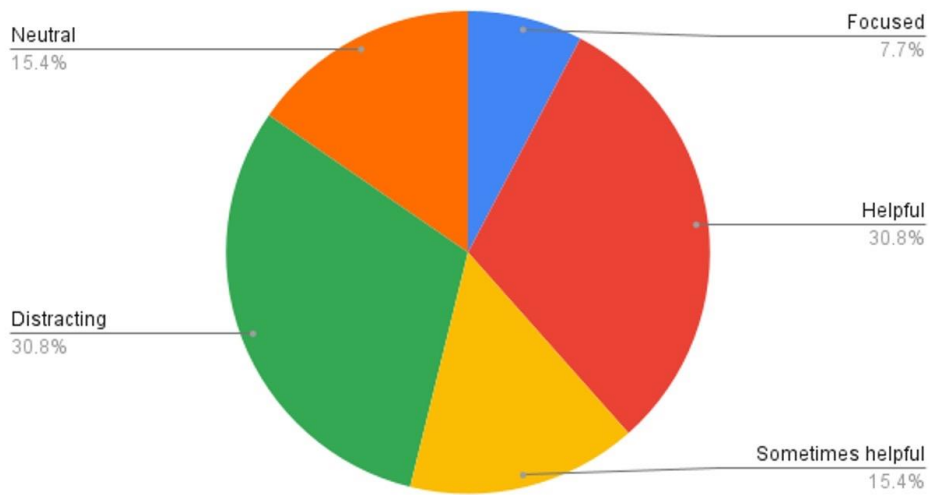
*Like how the one virus touching another (shows the gesture of one palm touching the other). I guess those could be used as gestures, right? But yeah, he, he didn't do anything.*

It should be noted, however, that even though a lot of participants could recall the gestures used in the videos, they engaged with the videos and nonverbal behaviors in them differently. Figure 1 showcases data from all participants, including the breakdown by the level, and presents their engagement with gestures. Regardless of whether the participants mentioned the gestures themselves or were prompted directly by me at the end of the interview, at the end, all of them were asked how they felt about the gestures being present in the videos. Based on participants' responses, the following major categories were singled out: "helpful", "sometimes helpful", "distracting", "neutral". Other categories included "focused", "not helpful", and "unspecified". Participants who fell into "unspecified" category did not provide any opinions on the presence of gestures, but three out of four participants in this category mentioned them in the interview. Thus, 12 participants found gestures helpful, with 4 participants indicating that the gestures helped them visualize the content described in the lectures and build associations necessary for understanding minute details. Five participants found gestures to be a distraction for their perception and retention of the content. They all noticed nonverbal behaviors in the video stimuli, but could not recall any particular gestures. One participant later outlined that it was hard for her to "follow verbal and nonverbal at the same time," which is similar to another participant in the study who indicated that he would rather listen to the lecturer than watch the person moving. Other participants referred to gestures as neutral (n=6), sometimes helpful (n=3), and not helpful (n=1). Lastly, one participant outlined that nonverbal cues made her more focused and attentive to the lectures.



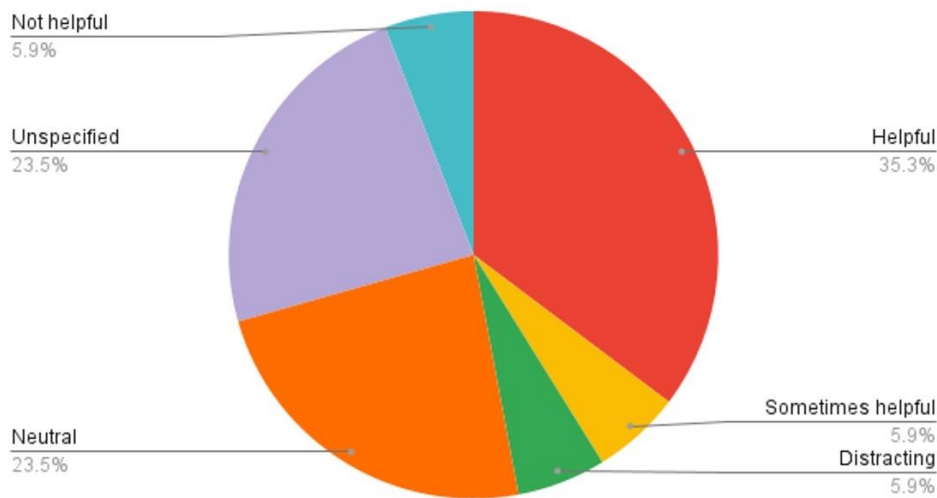
**Figure 2.** Participants' perception of the gestures in the videos

#### B2 level participants



**Figure 3.** B2 level participants' perception of the gestures in the videos

#### C1 level participants



**Figure 4.** C1 level participants' perception of the gestures in the videos

*Note:* There were two participants with C2 level, and they indicated that gestures were helpful.

#### *Facial cues*

I examined the co-occurring patterns in this category, and I detail participants' perceptions of facial expressions and eye contact. Overall, 11 participants commented on the presence or absence of lecturers' facial expressions. In particular, female lecturer was very expressive while talking about the life of the president and had different affect displays (e.g., laughs, smiles, eyebrow raise), and that was salient to participants. For example, one participant noted that:

*And I remember from, from the woman for sure, the first lecturer, um, that she used a lot of facial expressions when she talked about George Washington being that he was still a human or he cared a lot of a lot about his reputation.*

In general, participants indicated that facial expressions helped them to be more focused and better understand what content is emphasized in the lecture. Other behaviors that participants noticed were smiling, laughing, and paralinguistic (tone and voice). Tone and voice were important factors for determining emotions and "passion" about the context and speaker's attitude to it. Additionally,

four participants emphasized the importance of eye contact in the lectures as nonverbal cues that help focus on the lecture content and feel more “connected” to the lecturer. When directly asked about what facial cues they were looking at while watching the video, participants differed in their opinions. For example, 14 participants reported that they paid attention to lecturers’ faces, with 4 participants outlining that they also looked at other behaviors such as mouth and eyes, or both. Four participants indicated that they watched lecturers’ eyes, with one splitting the attention between eyes and mouth. Seven participants said that they were watching mouth and lip movements for better comprehension of speech and accent. Other participants either reported that they looked at facial expressions, including eye contact, or did not mention particular facial cues they were focusing on.

Overall, participants were positive about facial behaviors present in the videos, but similar to gestures, they engaged with facial cues differently. Even though some indicated that certain facial cues made them more engaged and focused, for few participants it was distracting. For example, two participants mentioned that eye behavior of the first lecturer was distracting in a way that it might have given them a wrong message.

### *Research question 2*

To answer the second research question, I turn to free recall scores and idea units that were recalled across three visual stimuli. There were 19, 18 and 13 possible points participants could receive for each video respectively. Participants completed free recalls in different formats. Whereas summary was the most common way to approach the task, other participants used bullet points and one used graph. Regardless of the format they chose to complete free recall, points were awarded if either idea unit (main, supplemental, details) was present or partially present. Descriptive statistics of the average number of ideas participants recalled are presented in Table 3. The results indicate that

participants produced more ideas (main, supplemental, details) for the second video ( $M=6.21$ ), which together with the eye tracking data might explain the higher rate of engagement with this video and retention of more details. The total number of idea units from the first and third videos differed from that of the second, but not to a substantial degree.

Table 3. Descriptive statistics of free recall performance

Video	Maximum	Mean	<i>SD</i>	Maximum points awarded for all idea units recalled
1	12	5.18	3.25	19
2	16	6.21	3.57	18
3	9.50	3.49	2.05	13

In the score sheet for the first and second video, information that was supplemented by gestures was coded as supplemental and mean of ideas recalled was calculated. Thus, in the first video there were two ideas where gestures were integral for comprehension – when the lecturer was describing visually how the person was sitting at the head of a table and hitting the piece of silverware against it. The means for recalling these ideas were 0.34 and 0.25 (out of 1.0 points possible) respectively. With regard to the second video, the recall of three ideas was significantly higher. The following ideas were identified that were accompanied with gestures and were important to the context: heat visualization goggles, satellite pictures, skin turning red/ get a sunburn. The means for these ideas were 0.53, 0.39 and 0.76 (out of 1.0 points possible). Since the third lecturer did not have any gestures to supplement the content, particular idea units were not analyzed afterward. Combined with the results from eye tracking sessions, interview data and free recall it can be inferred that participants paid to attention to meaning-enhanced gestures in the second video, therefore, the recall of these details was produced at a higher rate. The first lecturer’s nonverbal behaviors consisted of primarily illustrator gestures that did not carry any meaning per se. The only distinct

visual cue of hitting the piece of silverware, which in this case was the pen, was barely noticeable to participants.



## DISCUSSION

Guided by two research questions, this study provided the following findings. Participants engaged with videos differently, with face behavior occupying the largest dwell time ( $M=71.12\%$ ) and with eye behavior being the most prominent after it ( $M=41.48\%$ ). Attention to faces is in alignment with the findings from Batty (2021) and Gullberg and Holmqvist (2006), who found that test takers fixated largely on the face of the speaker. Similar to their findings, this study also found that participants spent the least amount time fixating directly on the gestures ( $M=2.03\%$ ), although the attention to gestures of the second lecturer was significantly higher than in other two video stimuli (5.21%). However, interview data combined with free recall details clearly show that participants paid attention to gestures although in varying degrees. Overall, 27 participants noted at different parts of the interview that gestures were prominent to them and considering eye tracking metrics results, which show the mean of 2.03% for attention, the findings might suggest that participants attended to gestures in their peripheral vision as evidenced by one of the participants: *I would say I was looking most of the time maybe at the face, while at the corner of my eyes I was also looking at the hands*. Buchan et al. (2007) indicated that it may not be necessary to foveate a certain facial cue to gather information necessary for understanding the content of speech or interpreting emotion. There also have been several studies suggesting that information can be processed covertly (Carrasco & McElree, 2001; Posner, 1980). Thus, it is possible that the participants noticed gestures in their parafovia, but did not fixate on them long enough or directly; rather, they took body language in as the whole picture and the gestures added to this whole picture.

Another finding that was consistent with Wagner's study (2008) was that participants remembered particular gestures and reenacted them. More than half participants (18 out of 32) who took part in the study were able to recall exact gestures, relate them to the video and reenact

them unconsciously. Wagner (2008) mentioned that participants made more references to gestures when viewing the video as opposed to answering multiple choice questions. Therefore, it is possible that participants in the present study were able to attend to nonverbal behaviors, including facial expressions and gestures, due to a different form of listening comprehension – free recall. The obvious benefits of this test modality were that participants were not timed and were allowed to complete a free recall in their preferred format, thereby allowing participants to visualize information in their heads or reflect on the video more thoroughly. The results of free recall showed that gestures helped participants remember particular details especially if they were supplemented by gestures as seen in the second video and to a lesser degree in the first video.

Lastly, participants had different engagement with gestures and facial cues that were present in the video stimuli. For example, the majority of participants (12 out of 32) found gestures to be helpful, with four participants noting that gestures enhanced their ability to visualize information and make relevant associations with the content. One participant specifically reported that gestures helped her focus on the material of the lectures more. However, other participants reported that gestures were sometimes helpful (n=3), neutral (n=6), distracting (n=5) or not beneficial (n=1) to them. With regard to facial cues, participants appreciated different aspects of facial behavior such as facial expressions, eye contact, and paralanguage. Their interaction with videos varied, for example, with some participants attending to face, mouth, or eyes.

### **Implications for language testing and test design**

The current study provides several implications for language testing. First, when administering the listening test, it is important to consider the purpose of the test and what it tries to measure (Suvorov, 2009), otherwise it might inadvertently create construct-irrelevant variance and assess other components beyond listening (Wagner, 2008; Brunfaut, 2016). For instance, as academic

lectures are deemed to be more authentic stimuli than audio or still images and academic lecture is the scenario listeners are likely to encounter in their lives (Ockey, 2007), it would be crucial to include such components as speaker and any additional materials that are prevalent in the lectures. That is, while it makes sense not to include visual stimuli for a listening test that features a phone call, it will only be beneficial in the communicative and authentic situations where listeners see the interlocutor. The second implication includes the importance of giving test takers more than one option of testing their listening comprehension. In the current study, while the majority of students engaged with the video stimuli in different degrees, there were a couple of students who preferred listening instead of watching the visual information on the screen. This could be due to their academic training: If listening was taught as an audio-only process, the participants may cognitively desire to focus on the audio-input only (a strategy or process that is familiar to them), which may limit them from achieving fuller comprehension of the material at hand, especially if research is right that visuals aid or augment listening comprehension.

## CONCLUSION

The purpose of this study was to a) investigate test takers' attention to nonverbal behaviors in the video-based listening test and b) see whether nonverbal behaviors are beneficial for comprehension of material in the lectures. Taking into consideration different views on inclusion of visual input in the listening assessment and given a few studies that incorporated nonverbal cues in the test, I decided to conduct this study to see whether nonverbal behaviors are salient and beneficial to participants. In summary, results of this study demonstrate that participants interacted with nonverbal information in the videos to different degrees. While gestures were recalled at a higher rate than facial cues, it was the gestures that produced a wide range of responses regarding their effectiveness and impact on participants' comprehension. In contrast, facial cues such as affect displays, eye contact, and paralanguage were regarded positively by a majority of the participants. These findings therefore suggest that even though nonverbal behaviors, including gestures and facial cues, were attended to frequently in the video-based listening test, it is important to consider participants' interaction with visual input. For example, with the exception of a few students, facial cues did not produce mixed responses as gestures did, despite them being recalled more frequently in the interview and free recall samples.

One limitation of the present study deals with free recalls. That is, even though free recall has advantages in terms of authenticity and integrated nature of the task, prompting students to recall everything they remember from the video might be taxing on their memory (Heinz, 1993), and each participant's level of detail might be impacted by the different proficiency levels. This explains the variation and different score distribution of free recalls. Therefore, a more structured version of free recall might yield different results.

## REFERENCES

- Ableeva, R., & Lantolf, J. (2011). *Mediated dialogue and the microgenesis of second language listening comprehension. Assessment in Education: Principles, Policy & Practice*, 18(2), 133–149. <https://doi.org/10.1080/0969594x.2011.555330>
- Batty, A. O. (2021). An eye-tracking study of attention to visual cues in L2 listening tests. *Language Testing*, 38(4), 511–535. <https://doi.org/10.1177/0265532220951504>
- Barenholtz, E., Mavica, L., & Lewkowicz, D. J. (2016). Language familiarity modulates relative attention to the eyes and mouth of a talker. *Cognition*, 147, 100–105. <https://doi.org/10.1016/j.cognition.2015.11.013>
- Bavelas, J. B., & Chovil, N. (2006). Nonverbal and verbal Communication: Hand gestures and facial displays as part of language use in face-to-face dialogue. In V. Manusov & M. L. Patterson (Eds.), *The Sage handbook of nonverbal communication* (pp. 97–115). Sage Publications. <https://doi.org/10.4135/9781412976152.n6>
- Bernhardt, E. B. (1991). *Reading development in a second-language*. Ablex
- Braun V., Clarke V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Brown, J. D., & Hudson, T. (1998). The alternatives in language assessment. *TESOL Quarterly*, 32(4), 653–675. <https://doi.org/10.2307/3587999>
- Brunfaut, T. (2016). Assessing listening. In D. Tsagari, & J. Banerjee (Eds.), *Handbook of Second Language Assessment* (pp. 97–112). (Handbooks of Applied Linguistics; Vol. 12). Mouton de Gruyter.
- Buchan, J. N., Paré, M., & Munhall, K. G. (2007). Spatial statistics of gaze fixations during dynamic face processing. *Social Neuroscience*, 2(1), 1–13. <https://doi.org/10.1080/17470910601043644>
- Buck, G. (2001). *Assessing listening*. Cambridge University Press.
- Burgoon, J. K. (1983). Nonverbal communication. In M. L. Knapp & G. R. Miller (Eds.), *Handbook of interpersonal communication* (pp. 344–390). Sage.
- Burgoon, J. K., Buller, D. B., Hale, J. L., & DeTurck, M. A. (1984). Relational messages associated with nonverbal behaviors. *Human Communication Research*, 10(3), 351–378. <https://doi.org/10.1111/j.1468-2958.1984.tb00023.x>
- Burgoon, J. K., Guerrero, L. K., & Manusov, V. (2011). Nonverbal signals. In M. L. Knapp & J.A. Daly (Eds.). *Interpersonal Communication* (4th ed., pp. 239–280). Sage.

- Carrasco, M., & McElree, B. (2001). Covert attention accelerates the rate of visual information processing. *Proceedings of the National Academy of Sciences*, 98(9), 5363–5367. <https://doi.org/10.1073/pnas.081074098>
- Clapham, C. (2000). Assessment and Testing. *Annual Review of Applied Linguistics*, 20, 147–161. <https://doi.org/10.1017/s0267190500200093>
- Cook, S. W., Yip, T. K., & Goldin-Meadow, S. (2010). *Gesturing makes memories that last. Journal of Memory and Language*, 63(4), 465–475. <https://doi.org/10.1016/j.jml.2010.07.0>
- Coniam, D. (2001). The use of audio or video comprehension as an assessment instrument in the certification of English language teachers: a case study. *System*, 29(1), 1–14. [https://doi.org/10.1016/s0346-251x\(00\)00057-9](https://doi.org/10.1016/s0346-251x(00)00057-9)
- Cubilo, J., & Winke, P. (2013). Redefining the L2 listening construct within an integrated writing task: Considering the impacts of visual-cue interpretation and note-taking. *Language Assessment Quarterly*, 10, 371–397. <https://doi.org/10.1080/15434303.2013.824972>
- Duncan, S., Jr. (1969). *Nonverbal communication. Psychological Bulletin*, 72(2), 118–137. <https://doi.org/10.1037/h0027795>
- Duolingo English test*. (n.d.). <https://englishtest.duolingo.com/scores>
- Ekman, P., & Friesen, W.V. (1969). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica*, 1(1), 49–98. <https://doi.org/10.1515/semi.1969.1.1.49>
- Ekman, P., Friesen, W. V., & O’Sullivan, M. (1988). Smiles when lying. *Journal of Personality and Social Psychology*, 54, 414–420. <https://doi.org/10.1037/0022-3514.54.3.414>
- Elfenbein, H. A. (2013). Nonverbal dialects and accents in facial expressions of emotion. *Emotion Review*, 5(1), 90–96. <https://doi.org/10.1177/1754073912451332>
- EyeLink Data Viewer 4.3.210 [Computer software]. (2023). Oakville, Ontario, Canada: SR Research Ltd.
- 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>
- Gregersen, T. (2005). Nonverbal cues: Clues to the detection of foreign language anxiety. *Foreign Language Annals*, 38(3), 388–400. <https://doi.org/10.1111/j.1944-9720.2005.tb02225.x>
- Gullberg, M., & Holmqvist, K. (2006). What speakers do and what addressees look at: Visual attention to gestures in human interaction live and on video. *Pragmatics and Cognition*, 14(1), 53–82. <https://doi.org/10.1075/pc.14.1.05gul>

- Gullberg, M., & McCafferty, S. G. (2008). Introduction to gesture and SLA: Toward an integrated approach. *Studies in Second Language Acquisition*, 30(2), 133–146. <https://doi.org/10.1017/S0272263108080285>
- Hall, J. A., Horgan, T. G., & Murphy, N. A. (2019). Nonverbal communication. *Annual Review of Psychology*, 70(1), 271–294. <https://doi.org/10.1146/annurev-psych-010418-103145>
- Heinz, P.J. (1993). *Towards enhanced, authentic second-language reading comprehension assessment, research, and theory building: The development and analysis of an automated recall protocol scoring system*. Unpublished doctoral dissertation, The Ohio State University, Columbus, Ohio.
- Hsieh, C.-N., & Davis, L. (2019). The effect of audiovisual input on academic listen-speak task performance. In S. Papageorgiou & K. M. Bailey (Eds.), *Global perspectives on language assessment: Research, theory, and practice* (pp. 18-31). Routledge.
- Jack, R. E., & Schyns, P. G. (2015). The human face as a dynamic tool for social communication. *Current Biology*, 25(14), 621–634. <https://doi.org/10.1016/j.cub.2015.05.052>
- Freeman, J. (2010). *Lecture 16 – The importance of George Washington* [Video]. Open Yale Courses. <https://oyc.yale.edu/history/hist-116/lecture-16>
- Krauss, R. M., Chen, Y., & Chawla, P. (1996). Nonverbal behavior and nonverbal communication: What do conversational hand gestures tell us? In M. P. Zanna (Ed.), *Advances in experimental social psychology* (pp. 389–450). Academic Press. [https://doi.org/10.1016/S0065-2601\(08\)60241-5](https://doi.org/10.1016/S0065-2601(08)60241-5)
- Marsh, A. A., Effenbein, H. A., & Ambady, N. (2003). Nonverbal “accents”: Cultural differences in facial expressions of emotion. *Psychological Science*, 14, 373–376. <https://doi.org/10.1111/1467-9280.24461>
- Matsumoto, D., & Hwang, H. S. (2013). Facial expressions. In D. Matsumoto, M. G. Frank, & H. S. Hwang (Eds.), *Nonverbal communication: Science and applications* (pp. 15–52). Sage Publications, Inc. <https://doi.org/10.4135/9781452244037.n2>
- McNamara, T. (2000). *Language testing* (1st ed.). Oxford University Press.
- McNeill, D. 1992. *Hand and Mind*. University of Chicago Press.
- McNeill, D. 2005. *Gesture and Thought*. University of Chicago Press.
- McNeill, D. (2012). *How language began: Gesture and speech in human evolution*. Cambridge University Press.
- Ockey, G. J. (2007). Construct implications of including still image or video in computer-based listening tests. *Language Testing*, 24(4), 517–537. <https://doi.org/10.1177/0265532207080771>

- Ockey, G. J. (2020). Assessment of listening. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics*. John Wiley & Sons. <https://doi.org/10.1002/9781405198431.wbeal0048.pub2>
- Perez, M. M. (2020). Multimodal input in SLA research. *Studies in Second Language Acquisition*, 42(3), 653–663. <https://doi.org/10.1017/s0272263120000145>
- Perniss, P. (2018). Why we should study multimodal language. *Frontiers in Psychology*, 9. <https://doi.org/10.3389/fpsyg.2018.01109>
- Posner, M. I. (1980). Orienting of attention. *The Quarterly Journal of Experimental Psychology*, 32(1), 3–25. <https://doi.org/10.1080/00335558008248231>
- Pulvermüller, F. (2005). *Brain mechanisms linking language and action*. *Nature Reviews Neuroscience*, 6(7), 576–582. <https://doi.org/10.1038/nrn1706>
- Riley, G. L., & Lee, J. F. (1996). A comparison of recall and summary protocols as measures of second language reading comprehension. *Language Testing*, 13(2), 173–189. <https://doi.org/10.1177/026553229601300203>
- Rukthong, A. (2020). MC listening questions vs. integrated listening-to-summarize tasks: What listening abilities do they assess? *System*, 102439. <https://doi.org/10.1016/j.system.2020.102439>
- Rukthong, A., & Brunfaut, T. (2019). Is anybody listening? The nature of second language listening in integrated listening-to-summarize tasks. *Language Testing*, 37(1), 31–53. <https://doi.org/10.1177/0265532219871470>
- Sakai, H. (2009). Effect of repetition of exposure and proficiency level in L2 listening tests. *TESOL Quarterly*, 43(2), 360–372. <https://doi.org/10.1002/j.1545-7249.2009.tb00179.x>
- Saltzman, W.M. (2008). *Lecture 20 – Bioimaging* [Video]. Open Yale Courses. <https://oyc.yale.edu/biomedical-engineering/beng-100/lecture-20>
- Scott, H., Batten, J. P., & Kuhn, G. (2018). Why are you looking at me? It's because I'm talking, but mostly because I'm staring or not doing much. *Attention, Perception & Psychophysics*, 81(1), 109–118. <https://doi.org/10.3758/s13414-018-1588-6>
- Snowden, F. (2010). *Lecture 13 – Contagionism and Anticontagionism* [Video]. Open Yale Courses. <https://oyc.yale.edu/history/hist-234/lecture-13>
- So, W. C., Sim Chen-Hui, C., & Low Wei-Shan, J. (2012). Mnemonic effect of iconic gesture and beat gesture in adults and children: Is meaning in gesture important for memory recall? *Language and Cognitive Processes*, 27(5), 665–681. <https://doi.org/10.1080/01690965.2011.573220>



- Sueyoshi, A., & Hardison, D. M. (2005). The role of gestures and facial cues in second language listening comprehension. *Language Learning*, 55(4), 661–699. <https://doi.org/10.1111/j.0023-8333.2005.00320.x>
- Suvorov R. (2014). The use of eye tracking in research on video-based second language (L2) listening assessment: A comparison of context videos and content videos. *Language Testing*, 32(4), 463–483. <https://doi.org/10.1177/0265532214562099>
- Suvorov, R. (2009). Context visuals in L2 listening tests: The effects of photographs and video vs. audio-only format. In C. A. Chapelle, H. G. Jun, & I. Katz (Eds.), *Developing and evaluating language learning materials* (pp. 53-68). Iowa State University.
- Urbanski, K. (B.), & Stam, G. (2023). Overview of multimodality and gesture in second language acquisition. In G. Stam & K. (B.) Urbanski (Eds.), *Gesture and multimodality in second language acquisition: A research guide* (pp. 1–25). Routledge.
- Vo, M. L. H., Smith, T. J., Mital, P. K., & Henderson, J. M. (2012). Do the eyes really have it? Dynamic allocation of attention when viewing moving faces. *Journal of Vision*, 12(13), 3. <https://doi.org/10.1167/12.13.3>
- Wagner, E. (2008). Video listening tests: What are they measuring? *Language Assessment Quarterly*, 5(3), 218–243. <https://doi.org/10.1080/15434300802213015>
- Wagner, E. (2010). Test-takers' interaction with an L2 video listening test. *System*, 38(2), 280–291. <https://doi.org/10.1016/j.system.2010.01.003>
- Wagner, E. (2013). An investigation of how the channel of input and access to test questions affect L2 listening test performance. *Language Assessment Quarterly*, 10(2), 178–195. <https://doi.org/10.1080/15434303.2013.769552>
- Wei, J., & Van Moere, A. (2021). *Aligning the Lexile framework for reading to the Common European Framework of Reference*. MetaMetrics. <https://metametricsinc.com/wp-content/uploads/2018/07/Aligning-the-Lexile-Framework-to-the-CEFR.pdf>
- Willems, R. M., & Hagoort, P. (2007). *Neural evidence for the interplay between language, gesture, and action: A review*. *Brain and Language*, 101(3), 278–289. <https://doi.org/10.1016/j.bandl.2007>
- Yuki, M., Maddux, W. W., & Masuda, T. (2007). Are the windows to the soul the same in the East and West? Cultural differences in using the eyes and mouth as cues to recognize emotions in Japan and the United States. *Journal of Experimental Social Psychology*, 43(2), 303–311. <https://doi.org/10.1016/j.jesp.2006.02.004>