Pertanika J. Soc. Sci. & Hum. 30 (S1): 63 - 79 (2022)

SOCIAL SCIENCES & HUMANITIES

Journal homepage: http://www.pertanika.upm.edu.my/

The Effects of Visual Input and Text Types on the Listening Comprehension of EFL Students in China

Tan Shaojie[1]*, Arshad Abd Samad[2] and Lilliati Ismail[3]

[1]*School of English, Anhui International Studies University, 231201, Anhui, China*
[2]*School of Education School of Education, Taylor's University, 1, Lrg DK Senza, 47500 Subang Jaya, Selangor, Malaysia*
[3]*Faculty of Educational Studies, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia*

ABSTRACT

In recent years, there has been an interest in using visual input in multimodal applications for language learning. However, the effects of visual input in listening materials remain to be discussed. Past literature has shown no unified answer to the effects of video input in improving listening comprehension scores. Since there are many proposals for the diversified reform of English teaching methods, it is worth examining whether using traditional audio listening only or using different video inputs can bring more significant benefits to students. The subjects of this study are 30 non-English majors in Chinese universities. This paper applied the quantitative research method, testing students' performance using different listening visual inputs (content, context, and audio only) and different listening text types (monologue and dialogue). Data were analysed using the two-way repeated measures within groups ANOVA. The interaction effects and the main effect of variables on listening performance were examined to explore the impact of visual input and text types on English listening comprehension. The following conclusions are drawn from the data analysis: (1) The interaction effect shows that video and text types significantly affect students' listening scores. Videos that contain authentic listening scenes and content elements are beneficial to promote listening comprehension as they support students' interpretation and understanding of what they hear, and (2) It was noted that grouping students by listening proficiency and examining their cultural background would expand the study in the future.

Keywords: English as a foreign language, listening performance, multimodal, text type, visual input

ARTICLE INFO

Article history:
Received: 17 July 2022
Accepted: 09 November 2022
Published: 22 December 2022

DOI: https://doi.org/10.47836/pjssh.30.S1.04

E-mail addresses:
787788545@qq.com (Tan shaojie)
arshad.abdsamad@taylors.edu.my (Arshad Abd Samad)
lilliati@upm.edu.my (Lilliati Ismail)
* Corresponding author

ISSN: 0128-7702
e-ISSN: 2231-8534

INTRODUCTION

As an input skill, listening plays a vital role in students' language development. The development of students' listening comprehension in the classroom is closely related to the listening materials they use (Richards, 2008; Sadiku, 2015; Vandergrift, 1999). Listening is examining verbal information and interpreting non-verbal information (Chion, 2019; Guillebaud, 2017). However, in traditional teaching and assessment of listening, students' listening comprehension is tested based solely on listening to audio input. With the development of multimedia technology, language learners can now access audio-visual materials that integrate listening and visuals. Although most researchers affirm its positive effects on listening comprehension (Sueyoshi & Hardison, 2005; Wagner, 2010), some studies point out that the use of videos does not have a positive effect on listening comprehension, and sometimes it even hinders learners' listening comprehension (Cubilo & Winke, 2013; Gruba, 2006; Suvorov, 2009). In summary, research results on the effects of video and audio on listening comprehension have yet to reach consistent experimental results.

Most real-life listening processes are multimodal (Campoy-Cubillo & Querol-Julián, 2015; Guichon & McLornan, 2008). The listeners can view the scene of the event and observe the speaker's facial expressions and body movements. In contrast, students can only hear processed sounds in traditional listening tests. They are expected to develop inferences or predictions based on what they hear to interpret the intended message without supporting information, such as the speaker's identity, facial expressions, and speech situation. It should not be overlooked, however, that in traditional audiometric examinations, the listening test is highly well-structured at the level of comprehension of phonetic information. The addition of visual information changes the listening test from a test of sound decoding to a test that includes sound decoding and non-speech information interpretation. Nevertheless, for non-traditional English teaching and learning, using internet video resources to update information is necessary to provide an authentic and vivid language cognitive environment. As a result, in recent years, more and more attention has been dedicated to determining ways to integrate video resources into English education (Harmer, 2001; Hung, 2015; Richards, 2008).

Text types or listening texts refer to the text used in teaching listening comprehension. The main body is the language material, but because of the nature of the language material as the listening text, it also involves language accuracy, intonation, tone, and speed of the language material. According to Atay (2005), the teaching interpretation of any language text has three dimensions: the dimension of language itself, the dimension of language as a teaching material, and the dimension of language used in teaching activities. The dimension of language as a listening teaching material includes the language difficulty, the graphical difficulty, and the genre of the listening texts. Concerning the listening text genre, Lu (2016) divided it into two types: dialogue and monologue. The dialogue covers life conversations, teaching conversations, TV programmes, and radio programmes, while monologues include life monologues, speeches, and TV and radio programmes, among others. Listening texts can also be classified based on their contents, complexity, authenticity, sphere, theme, number, and other factors. According to Neri et al. (2003), the basic types of oral expression are dialogue and monologue, broadly recognised.

Context visuals depict details about the context of a verbal encounter, such as the participants, location, and text type. For example, a photograph of a man and a woman chatting in a classroom may be used as a background visual to the dialogue being heard. According to Ginther (2002), background images have two fundamental functions: (a) setting the stage for the verbal exchange and (b) signalling a change in speakers in dialogue. Finally, content visuals are visuals relevant to the content of the verbal interaction, which can include still images, videos, drawings, diagrams, and others. An example would be the inclusion of a photograph of Leonardo Da Vinci's Mona Lisa in a lecture on Renaissance art.

This research incorporates multimodal teaching into classroom listening by providing students with different video types and exploring the impact of varying video and textual inputs on students' listening effects. Visual input is categorised into content video and context video as used in the literature (Suvorov, 2008, 2009, 2015).

In the current study, student performance on these types of visual input and audio-only input is examined. The text types consist of two: dialogue and monologue. In this paper, there are three specific research questions:

a. Will the different video inputs—context video, content video, and audio-only— affect students' listening comprehension?
b. Will the different textual inputs—monologue and dialogue—affect students' listening comprehension?
c. Are there any interaction effects between the two independent variables (visual and textual input)?

LITERATURE REVIEW

Multi-Modality

"Multi-modality" is also called multi-symbol, which refers to the phenomenon where people use multiple senses to communicate through language, text, images, sounds, expressions, actions, symbols, and other means (Yuan & Fengping, 2021). In the communication process, a large part of the meaning is not reflected by language but by other factors. For example, some symbols appear with language, and in terms of sound, some are expressed through speech rate, intonation, and stress. Physically, they are expressed through body movements, expressions, and gestures. Therefore, communication is not only carried out by using a single sense but also through multiple senses at the same time. Whether it involves a single medium, dual media, or multimedia, it can be called multimodal if its content is presented in more than one medium (Li et al., 2021).

Some scholars believe that compared to pure audio-monomodal materials of the same content, audio-visual materials which involve audio and visual input can reduce the difficulty of listening (Li, 2019). Ginther (2002), for example, showed that video could supplement audio information in the context of a scene. To explore the effects of video (audio-visual) and pure audio on listening comprehension, some scholars in China have made further discoveries through empirical research. Delu (2009), who conducted empirical research, found that in learning English, a multimodal combination of audio and video with English subtitles has the most significant effect on students' listening comprehension, followed by audio and video with Chinese subtitles, and audio and video without subtitles, while audio alone has a minor effect. The experimental research conducted by Woottipong (2014) showed a big difference between the scores of the video group and the audio group, indicating that video can promote listening comprehension. Maleki and Rad (2011) investigated the effects of visual and textual to verbal stimuli on listening comprehension performance. They found that visual aids were more advantageous to listeners with low proficiency. In contrast, textual aids were more beneficial to listeners with higher proficiency, and tests with more static images yielded much better performance than those with fewer static images.

Other researchers have also obtained many valuable results in the empirical research of multimodal teaching of listening. Baltova's (1994) research mainly focused on French listening, and the target population was Canadian non-French majors. Experiments showed that French subtitles helped students recall the content of listening materials in teaching listening. Vandergrift (2004) also took French listening lessons as the research object. Through investigation and research, it was found that students showed a high interest in multimodal listening teaching and were very willing to accept it. The study of Romero and Arévalo (2010), which focused on the role of multimodal teaching of listening, found that the multimodal teaching of a listening model can promote a better understanding of listening materials among students.

Text Types

Ginther (2002) researched the effects of the presence or absence of different types of stimuli (dialogues, short conversations, academic discussions, and mini talks) and proficiency on students' performance on the TOEFL. He used a video format for dialogue and lectures in his study. He noted that the total score of the dialogue and lecture videos was noticeably lower than those presented in the pictorial and audio formats. This difference was so significant that it must not be ignored. The focus of another study conducted by Wagner (2008) investigated the effects on L2 listeners watching a video monitor when presented with different listening video text types, such as academic lectures and dialogues. The test consisted of six tasks: three dialogue and lecture texts. Statistical analysis revealed that the time subjects focused on the video monitor during the three dialogue texts was higher (72%) than when they focused on the three lecture texts (67%).

Moreover, Amiri and Saberi (2017) explained whether dialogues and lectures, the two primary sources of textual materials, affect listening comprehension tests differently. The participants in their study were 60 male English language learners. To examine the influence of the different text types on the participants' listening comprehension scores, the researcher used an internet-based listening test designed by Suvorov (2008). The passages consisted of different text types, including dialogues and lectures, in various input formats (audio, video, and image). The results showed that the participants' test scores on dialogue passages did not differ from their scores on lecture passages in all input formats. This result provides a new idea for the variables of the two perspectives in this paper.

Types of Visual Input and Listening Comprehension

It is necessary to consider the different types of visual information to estimate the role of visuals in L2 listening more accurately (Lesnov, 2018). It would lead to a more meaningful analysis of the impact of the visuals on both the lower-level and higher-level processes in L2 listening. Most studies have shown that video materials can improve listening to a certain extent (Picou et al., 2011). The reason is that the richness and authenticity of video content can significantly stimulate learners' interest in learning; the relevant background knowledge provided in the video also helps learners grasp the overall content. In addition, video can embody some virtual abstract concepts and construct a schematic model of the information in the brain, significantly reducing listening difficulty (Gruba, 2006).

Scholars hold three views on the role of video in listening comprehension. One group of scholars found through experiments that the brain's simultaneous processing of auditory and visual information can interfere with listening comprehension because the insertion of a motion or still picture will distract attention, which increases the cognitive burden and makes it easier to cause the loss of listening information (Pusey & Lenz, 2014; Seeber, 2017). Another school of scholars indicates that the scenes conveyed in the video, the speaker's gestures, body posture, and paralanguage, among others, can provide learners with more clues and help activate existing schemas and establish the connection between the new and old information, stimulating the learners to recall the information they heard, thereby promoting understanding (Ginther, 2002; Guichon & McLornan, 2008). The last group of scholars found that the information in the video had no significant effect on the listening comprehension scores when they compared the listening scores of the candidates in the video group and the audio group (Sarani et al., 2014).

The reasons why scholars hold different opinions can be summarised in the following paragraph.

First, some researchers needed to classify the type of video information involved. It is important as different types of video information may affect students' listening ability. According to Ginther (2002) and Ockey (2007), video information is mainly divided into context and content. A context video provides listeners with scenes of the video material and the speakers, but the scenes in the video are fixed (Wijnants et al., 2019, October). For example, in a classroom lecture, listeners can see a teacher teaching the students, but it is difficult to guess the main topic of the lecture. On the contrary, a content video shows the scenes and the person talking and allows listeners to guess the main topic through the constantly switching screens. For example, listeners watching a TV news announcer reporting on war or natural disaster can guess the main content of the news through the series of images presented even though they do not hear the specific details.

Second, some researchers did not group the foreign language proficiency of the subjects (Sueyoshi & Hardison, 2005). Their study investigated the contribution of gestures and facial cues to second-language learners' listening comprehension of a videotaped lecture by a native speaker of English. A total of 42 learners of English as a second language were randomly assigned to three stimulus conditions: AV-gesture-face (audio-visual including gestures and face), AV-face (no gestures), and Audio-only. The result showed that AV-gesture-face showed the best results. However, the shortcoming of this paper is that there did not divide students into groups according to their different language proficiency. Therefore, video information may have different effects on learners of different foreign language proficiency, i.e., although it may benefit some candidates, it may have no significant effect on others.

Suppose researchers have not classified different listening tasks and question types. In that case, different video information may significantly affect certain tasks and question types, but it has no significant effect on other tasks and question types. Whether video information promotes learners' listening comprehension may be related to their ability to interpret the images in the video or their attitude towards the video (Wagner, 2008; 2010). This paper, therefore, attempts to make up for the shortcomings of existing research and more comprehensively investigate the impact of different video types and text types on their listening performance.

METHOD

The research questions were addressed quantitatively using a within-subjects quasi-experimental design. In this research, the independent variables measured throughout the experiment were types of visual input (context visual, content visual, and audio-only) and text types (dialogue and monologue). In contrast, the dependent variable was the listening performance score on the online test.

Participants

The 30 students who took the listening test were all second-year students from different majors at Anhui International Studies University. Twelve males and 18 females took English listening courses in the 2021–2022 school year. In addition, they must take the English listening comprehension course offered in the second-year student semesters. Due to the epidemic's impact, the convenience sampling method was used, a type of non-probability sampling involving the sample being drawn from that part of the population close to hand.

Materials

An online listening comprehension test (OLCT) was developed to investigate the students' performance on visual and textual input types. The online listening test took two weeks for the teachers at the Language Testing Centre of Anhui Foreign Language Institute, who selected appropriate video material from the news and then chose six different areas of video news from the alternative 10. Then reviewed by relevant experts, it was posted on the Chaoxing platform. Finally, students logged in to their accounts and answered the questions on the platform. The test consisted of six listening passages and 30 multiple-choice questions and lasted 40 minutes. Each listening passage has one of the two text types: dialogue and monologue. In addition, the researchers used one of the three types of visual input in the test: a context visual, a content visual, and no visual (i.e., audio-only format) with the listening comprehension passages. Table 1 outlines the structure of the OLCT, and the sequence of the passages played in the online test. To ensure that students can avoid the impact of fatigue on their performance during the listening process, the playback order of the test is as follows: AD-XM-TD-AM-TM-XD.

The input was selected from Voice of America (VOA) news channels and other American-based mainstream media. The topics covered culture, history, politics, and others. After watching each video clip, test-takers are expected to respond to the questions by selecting the best answer from the four options. According to the classification of visual types by Bejar and Ginther (2002), a context video contains visual information about the context of the lecture. It mainly serves three purposes: (1) It is focused on the situation, (2) it sets the scene for verbal exchange, and (3) it gives a cue to the viewers on a change of speakers in the conversation. For example, a context news clip can show a journalist and a US security advisor talking about American foreign policy in a studio. In the studio, listeners can only see the two people exchanging information verbally. A screenshot of the video, which illustrates

Table 1
Structure of the OLCT

Audio-only		Context		Content	
Dialogue	Monologue	Dialogue	Monologue	Dialogue	Monologue
AD	AM	XD	XM	TD	TM

A = Audio, X = context video, T = content video, M = monologue, D = dialogue a context video from the OLCT where the message is conveyed through verbal information between the host and the US security advisor, is provided in Figure 1.

A content video, on the other hand, provides visual information besides oral input. The visual information can be a photo, a graph, or a drawing related to the content of the verbal stimulus. These visuals are referred to as content visuals (Ginther, 2002). For example, in a content monologue, the News comes with some visual stimuli, such as the video scene of the Covid-19 vaccination appearing in the video. Figure 2 provides a screenshot of the content monologue. The video shows the global spread of Covid-19 and the response strategies, such as vaccinations adopted by many countries. In this screenshot, the image shown in the video is the back of a man who is being vaccinated, so the content video can provide some listening information to help listeners better understand the listening materials.

The selection of video clips for the OLCT to ensure consistency in the content of the video was based on the following criteria:

1. Each video comes from news broadcasted by mainstream television broadcast news channels in the United States.
2. Each piece of news is limited to topics related to social life, such as economics, education, and politics.
3. Each video's content difficulty is similar and based on the Flesch-Kincaid grade level measurement. All videos are "fairly difficult" in terms of their level of the grade level test.
4. Speakers in all videos use American pronunciation.
5. The length of the video is about 3 minutes.
6. All videos have the same sound and picture quality.
7. All content videos provide pictures and videos related to the content. The context video provides sound and picture images but does not display other relevant prompt information about the content of the conversation.
8. All dialogue videos are between two people with no third person involved.

Figure 1. A screenshot of a page with a context video from the OLCT

Figure 2. A screenshot of a page with a content monologue video from the OLCT

The researcher utilised the Flesch Reading Ease and Flesch-Kincaid Grade Levels to determine the difficulty level of the language used in each video clip. These two formulas are the most extensively used readability formulas for determining the difficulty level of written texts. Despite some concerns about a lack of "empirical validations of the listenability/readability equation", according to Suvorov and Hegelheimer (2013), some researchers have used Flesch's readability formulas to assess "listenability" (i.e., external factors that make listening difficult or easy) (Rubin, 1994, p. 263). Flesch Reading Ease ratings are given on a scale of one to one hundred, with lower scores indicating difficult readability. Flesch Reading Ease scores and Flesch-Kincaid Grade Levels for each video clip in the OLCT are provided in Table 2.

As can be seen in Table 2, the difficulty of all video and audio is similar. Therefore, the score description is fairly difficult.

Procedure

Before the study began, the researcher introduced the project to the students, requested them to sign the agreement, and informed them of the online test time and location. Then, two days before the pilot test, the researchers and an IT specialist inspected the equipment, such as the earphones and screen connections in the multimedia classroom, to ensure that the listening test would proceed smoothly.

On the day of the listening test, the three teachers arrived at the multimedia language room one hour in advance. They placed a piece of white paper and a pencil on each student's desk to allow them to take notes during the test and use them when answering the questions. Before the formal test began, the invigilator read out the online test instructions. Since the Chaoxing test platform has a memory function, all the questions answered were recorded even if the students quit halfway. Chaoxing is the campus teaching platform, so all students are given a student ID upon enrolment. It allows them to use the software to submit their homework after signing in, so students know its login operation. The researcher distributed the test papers before the start of the test to each student based on their student ID. When the students logged in to

Table 2
Readability statistics for the scripts of video clips in the OLCT

Video/Audio Clip	Visual type	Word Count	Flesch Reading	Score Description	Flesch-Kincaid Grade Levels
Video1	Content Dia	550	58.411	Fairly difficult	11.59
Video2	Content mono	414	57.61	Fairly difficult	10.94
Video3	Context Dia	555	52.125	Fairly difficult	10.4
Video4	Content mono	488	57.315	Fairly difficult	8.056
Audio1	Audio	463	52.72	Fairly difficult	9.19
Audio2	Audio	551	58.76	Fairly difficult	9.62

their accounts, they could see a notification of the listening test in the message column. They then clicked to enter and started to answer the questions. A countdown clock also appeared on the exam page to remind students of the remaining time. Figure 3 is a screenshot of a webpage with a test item.

Data Analysis

The repeated measures within-subjects Two Way Analysis of Variance (ANOVA) was used to examine the interaction effects and the main effects of the factors in the study to answer the three research questions. The two within-subject factors are visual and textual input with three (content, context, and audio only) and two (monologue and dialogue) levels, respectively. If a statistically significant difference is found, the main effects of visual input and textual input will be analysed through a one-way repeated measures ANOVA for visual input (3 levels) and Paired Samples t-test for textual input (2 levels), respectively. Before the use of ANOVA, descriptive statistics involving the means, standard deviations, and values of skewness and kurtosis were calculated for overall test scores, as well as for scores on the two factors on the OLCT. The assumption about the validity of the procedure (i.e., normal distribution of scores and Skewness and Kurtosis of the data is satisfied).

RESULTS AND DISCUSSION

Table 3 presents the descriptive statistics under different video input modalities (i.e. audio-only, context video, and content video).

Results shown in Table 3 reveal that performance on the content visual has the highest mean (M=7.9, SD=1.06). The mean for listening with context video is slightly lower (M=7.16, SD=1.08), while the mean

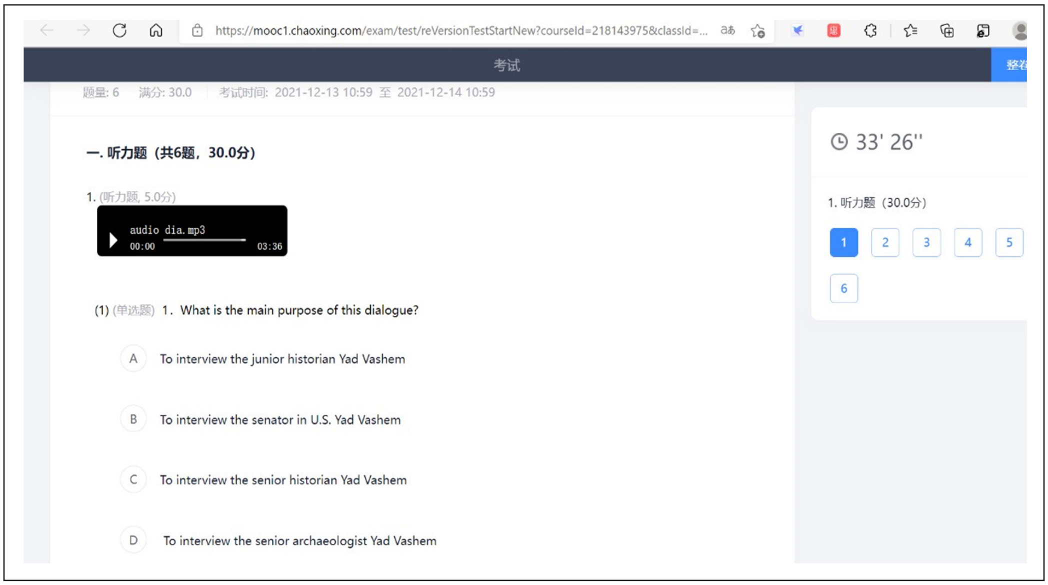

Figure 3. A screenshot of a webpage with a test item from OLCT

for performance on audio-only stimulus has the lowest score (M=7.13, SD=0.78).

The descriptive statistics under different text types are shown in Table 3.

Results shown in Table 4 reveal that the mean is higher in dialogue text types than in monologue text types which are (M=11.3, SD=1.41) and (M=10.9, SD=1.30), respectively.

The descriptive statistics under different text types and visual input are below in Table 5.

Table 5 reveals that the mean is highest in content visual input with dialogue text type (TD) (M=4.17, SD=0.65), and the lowest score is in audio-only with dialogue (M=3.53, SD=0.5).

Because students were within a group, they were repeatedly measured based on different video and text modalities. General linear repeated measures were performed to determine whether an interaction between the two independent variables existed, and the P-value of the interaction effect was

Table 3
Descriptive statistics for types of visual input

Visual	N	Mean	Std. Deviation	Skewness		Kurtosis	
	Statistics	Statistics	Statistics	Statistics	Std. Error	Statistics	Std. Error
AUDIO	30	7.1333	.77608	-.242	.427	-1.261	.833
CONTEXT	30	7.1667	1.08543	.514	.427	.496	.833
CONTENT	30	7.9000	1.06188	-.159	.427	-.769	.833

Table 4
Descriptive statistics for different text types

Textual	N	Mean	Std. Deviation	Skewness		Kurtosis	
	Statistics	Statistics	Statistics	Statistics	Std. Error	Statistics	Std. Error
DIA	30	11.3000	1.41787	.282	.427	-.407	.833
MONO	30	10.9000	1.29588	.299	.427	-.914	.833

Table 5
*Descriptive statistics for different visual input * text types*

	N	Mean	Std. Deviation	Skewness		Kurtosis	
	Statistics	Statistics	Statistics	Statistics	Std. Error	Statistics	Std. Error
AD	30	3.5333	.50742	-.141	.427	-2.127	.833
AM	30	3.6000	.56324	.198	.427	-.835	.833
XD	30	3.6000	.72397	.210	.427	-.234	.833
XM	30	3.5667	.62606	.635	.427	-.453	.833
TD	30	4.1667	.64772	-.166	.427	-.502	.833
TM	30	3.7333	.63968	.291	.427	-.554	.833

AD = Audio dialogue, AM = Audio monologue, XD = context Dialogue, XM = context monologue, TD = Content Dialogue, TM = Content Monologue observed. The result of this interactive effect is shown in Table 6.

The results of the two-way repeated-measures ANOVA revealed that there was a significant interaction effect between visual input and different textual types (F (2,58) = 4.09,p < .05, $\eta p2$ = .124). The interactive effects of visual input and text type influenced the participants' performance. The main effect of visuals was also noted to be significant, while the main effect of textual stimuli was not. The main effect of two independent variables, i.e., visual and textual, was also examined. The repeated measure was conducted again, and Bonferroni was chosen to compare the main effect between the two variables. The results are shown in Table 7.

The pairwise comparisons showed a significant difference in the performance on content and audio-only stimuli as well as on context and audio-only stimuli. There was, however, no significant difference between performance on content and context visual stimuli. It shows that students' performance is better when listening comprehension materials are presented with visuals, regardless of whether they are content or context visuals than when there is no visual.

Results of pairwise comparisons to observe the main effect of text types are provided in Table 8.

The results shown in Table 8 indicate that the text type has no significance in the main effect (p > 0.05).

Table 6
The test of within-subjects effect between visual and textual input

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Visual	Sphericity Assumed	5.633	2	2.817	9.782	.000	.252
Error(visual)	Sphericity Assumed	16.700	58	.288			
textual	Sphericity Assumed	.800	1	.800	2.275	.142	.073
Error(textual)	Sphericity Assumed	10.200	29	.352			
Visual * textual	Sphericity Assumed	2.100	2	1.050	4.087	.022	.124
Error(visual*textual)	Sphericity Assumed	14.900	58	.257			

Table 7
Pairwise comparison for the main effect of visual input

(I) visual	(J) visual	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
1	2	-.017	.094	1.000	-.256	.223
	3	-.383*	.092	.001	-.617	-.150
2	1	.017	.094	1.000	-.223	.256
	3	-.367*	.107	.006	-.639	-.095
3	1	.383*	.092	.001	.150	.617
	2	.367*	.107	.006	.095	.639

N/B: 1 = content, 2 = context, 3 = audio only

Table 8
Pairwise comparisons for the main effect of textual input

(I) textual	(J) textual	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Differences	
					Lower Bound	Upper Bound
1	2	.133	.088	.142	-.047	.314
2	1	-.133	.088	.142	-.314	.047

N/B: 1 = dialogue, 2 = monologue

In summary, for Research Question One, i.e., 'Will the different video inputs (context video, content video, and audio-only) have different effects on students' listening comprehension?', results show that the students' listening scores for the content video input mode are the highest. There is almost no difference between the student's scores in the context video mode and the pure audio input mode. There was a significant main effect in the visual input and pairwise comparisons, which indicates that this effect was due to the significant difference between the content visual input and the audio-only input and between the content visual input and the audio-only input. There was no significant difference in the visual input pairwise comparison between the content and context type.

As for Research Question Two, in terms of text types, the results of descriptive statistics show that the mean scores on dialogue text type are higher than on monologue text type among students. However, in the within-group main effect test, it was found that the effect between listening to the dialogue and monologue stimuli was not significant. Lastly, for Research Question Three, i.e., "Are there any interactive effects between the two independent variables," ANOVA results indicate that there is a significant interaction effect of visual and textual input on student performance ($F(2,58) = 4.09$, $p < .05$). Performance was highest when the video involved content and was presented in a dialogue. Performance was lowest when it involved audio-only presented in monologue. The interaction between visual input and text types suggests that teachers should be fully aware of the features of a listening text and the nature of videos when teaching or assessing listening comprehension.

CONCLUSIONS AND RECOMMENDATIONS

The conclusions are summarised as follows: From the perspective of video type, the highest scores are in content video type, followed by context video, and finally, audio-only listening. Regarding text type, participants' scores in dialogue text-type videos were higher than those in monologue videos. The interaction of the two shows that video and text types significantly affect students' listening scores.

This study is based on the impact of different video types and text types on listening, which results in several effects. First, experimental data show that content video media has a significant impact on listening comprehension. Thus, it is recommended that teachers incorporate content videos into classroom listening training in the future. Videos containing authentic listening scenes and content elements promote listening comprehension as they support students' interpretation and understanding of what they hear.

Second, experimental results show a significant effect on the interaction of visual and text types; therefore, teachers should consider this. Finally, it is worth noting that the highest scores were observed in content-type videos involving dialogues. Contrary to the belief that content in content-type videos may distract students in their listening comprehension task and that those dialogues would do the same because of the presence of an additional speaker, students seem to have performed well under these conditions. In listening comprehension, students use as much information as they see and hear, which does not distract them from their task.

In a theoretical sense, the results of this study will contribute to the multimodal theories in learning and contribute to a more effective English teaching method. The results are designed to contribute to the representational features of auditory and visual senses to provide theoretical explanations for multimodal processing. In a practical sense, it provides insight for selecting and designing various types of English video materials required by teachers and test developers to assess students' English listening ability in authentic settings accurately.

As shown by the experiment conducted in this study, when the visual modality provides background knowledge that directly corresponds to the auditory information, it will promote students' cognition, thereby improving students' listening comprehension. However, if the visual information does not correspond to the auditory information, it will form a cognitive load and distract students' attention. Receiving different simultaneous information can result in excessive cognitive costs, preventing students from processing the information effectively (Kirschner et al., 2018). Therefore, in listening classes, teachers must master the synergistic, reinforcing, or complementary relationship between the various modalities to improve their teaching quality. If the different modes are contradictory, irrelevant, and disconnected from each other, it may affect teaching. In this sense, mode selection should be based on the principle of increasing positive effects (Ruan, 2015).

This paper has the following limitations: Firstly, this research focused on the performance and achievement of students under different teaching methods in the classroom, and classroom interaction is not within the scope of the study. The sample analysed in the study consists of 30 Chinese students studying at Anhui International Studies University in Anhui Province. They are non-English majors between 19 and 20 years old and have been learning English as a foreign language for 8-10 years. Therefore, the study findings may not be generalised to other samples or populations, especially those with significantly different cultural and educational backgrounds. In addition, the failure to carry out high-level and low-level grouping is also a major defect of the paper. It is recommended that later researchers discuss the effects of different proficiency levels among students on visual and text types of stimuli.

ACKNOWLEDGEMENT

The authors would like to thank Liwei of the Computer Science Department of AFLU for helpful discussions on topics related to this work.We would like to thank Anhui International Studies University for its financial support for this project, the training center of the school for its strong support during the experiment, and the director of the data processing center, Mr. Liu Wei, for his comments and suggestions on data processing.

REFERENCES

Atay, D. (2005). Reflections on the cultural dimension of language teaching. *Language and Intercultural Communication, 5*(3-4), 222-236. https://doi.org/10.1080/14708470508668897

Amiri, F., & Saberi, L. (2017). The impact of learner-centered approach on Learners' motivation in Iranian EFL students. *International Academic Journal of Social Sciences, 4*(1), 99-109. https://doi.org/10.9756/iajss/v6i1/1910015

Baltova, I. (1994). The impact of video on the comprehension skills of core French students. *Canadian modern language Review*, *50*(3), 507-531. https://doi.org/10.3138/cmlr.50.3.507

Campoy-Cubillo, M. C., & Querol-Julián, M. (2015). Assessing multimodal listening comprehension through online informative videos: The operationalisation of a new listening framework for ESP in Higher Education. In Diamantopoulou, S. & Ørevik, S. (Eds). *Multimodality in English Language learning*. Routledge. https://doi.org/10.4324/9781003155300-17

Chion, M. (2019). The three listening modes. In *Audio-vision: Sound on screen* (pp. 22-34). Columbia University Press. https://doi.org/10.7312/chio18588-004

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*(4), 371-397. https://doi.org/10.1080/15434303.2013.824972

Delu, Z. (2009). Multimodal discourse theory and its application to foreign language teaching with modern media technology. *Foreign Language Education*, *30*(4), 15-20. https://doi.org/10.16362/j.cnki.cn61-1023/h.2009.04.006.

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

Gruba, P. (2006). Playing the videotext: A media literacy perspective on video-mediated L2 listening. *Language Learning & Technology*, *10*(2), 77-92. Http://llt.msu.edu/vol10num2/gruba/

Guichon, N., & McLornan, S. (2008). The effects of multimodality on L2 learners: Implications for CALL resource design. *System*, *36*(1), 85-93. https://doi.org/10.1016/j.system.2007.11.005

Guillebaud, C. (2017). Introduction: Multiple listenings. Anthropology of sound worlds. In Guillebaud, C. (Ed.) *Toward an anthropology of ambient sound* (pp. 1-18). Routledge. https://doi.org/10.4324/9781315755045-1

Harmer, J. (2001). The practice of English language teaching. *London/New York*, 401-405. https://doi.org/10.1177/003368820103200109

Hung, H. T. (2015). Flipping the classroom for English language learners to foster active learning.

Computer Assisted Language Learning, *28*(1), 81-96. https://doi.org/10.1080/09588221.2014.967701

Kirschner, P. A., Sweller, J., Kirschner, F., & Zambrano, R. J. (2018). From cognitive load theory to collaborative cognitive load theory. *International Journal of Computer-Supported Collaborative Learning, 13*(2), 213-233. https://doi.org/10.1007/s11412-018-9277-y

Lu, Z. (2016). Teaching interpretation of listening text: framework and cases. *English Learning* (05), 4-7.

Li, C. H. (2019). Using a listening vocabulary levels test to explore the effect of vocabulary knowledge on GEPT listening comprehension performance. *Language Assessment Quarterly, 16*(3), 328-344.

Li, Z., Zhu, J., & Li, X. (2021). Factors influencing the behavior of multi-modal information search. *Library Hi Tech, (ahead-of-print).*

Lesnov, R. O. (2018). *The role of content-rich visuals in the L2 academic listening assessment construct* [Doctoral dissertation, Northern Arizona University].

Maleki, A., & Rad, M. S. (2011). The effect of visual and textual accompaniments to verbal stimuli on the listening comprehension test performance of Iranian high and low proficient EFL learners. *Theory and Practice in Language Studies*, *1*(1), 28-36. https://doi.org/10.4304/tpls.1.1.28-36

Neri, A., Cucchiarini, C., & Strik, H. (2003, August). Automatic speech recognition for second language learning: how and why it actually works. *In Proc. ICPhS* (pp. 1157-1160). https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.6.7376&rep=rep1&type=pdf

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

Picou, E. M., Ricketts, T. A., & Hornsby, B. W. (2011). Visual cues and listening effort: Individual variability. *Journal of Speech, Language, and Hearing Research, 54*(5), 1416-1430. https://doi.org/10.1044/1092-4388(2011/10-0154)

Pusey, K., & Lenz, K. (2014). Investigating the interaction of visual input, working memory, and listening comprehension. *Language Education in Asia, 5*(1), 66-80. https://doi.org/10.5746/leia/14/v5/i1/a06/pusey_lenz

Richards, J. C. (2008). Teaching listening and speaking. In Bailey, K. M. (Ed.) *Teaching listerning and speaking in second and foreign language contexts* (p. 48). Cambridge university press. https://doi.org/10.5040/9781350093560.ch-007

Ruan, X. (2015). The role of multimodal in Chinese EFL students' autonomous listening comprehension & multiliteracies. *Theory and Practice in Language Studies*, *5*(3), 549. https://doi.org/10.17507/tpls.0503.14

Rubin, J. (1994). A review of second language listening comprehension research. *The modern language journal, 78*(2), *199-221.* https://doi.org/10.1111/j.1540-4781.1994.tb02034.x

Romero, E. D., & Arévalo, C. M. (2010). Multimodality and listening comprehension: testing and implementing classroom material. *Language Value*, (2), 100-139.

Sadiku, L. M. (2015). The importance of four skills reading, speaking, writing, listening in a lesson hour. *European Journal of Language and Literature*, *1*(1), 29-31. https://doi.org/10.26417/ejls.v1i1.p29-31

Sarani, A., Behtash, E. Z., & Arani, S. M. N. (2014). The effect of video-based tasks in listening comprehension of Iranian pre-intermediate EFL learners. *Gist: Education and Learning Research Journal*, (8), 29-47.

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. S. (2008). *Context visuals in L2 listening tests: The effectiveness of photographs and video vs. audio-only format* [Master's thesis, Iowa State University]. Iowa State University. https://doi.org/10.31274/rtd-180813-16671

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. https://doi.org/10.31274/rtd-180813-16671

Suvorov, R. (2015). 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., & Hegelheimer, V. (2013). Computer-assisted language testing. *The companion to language assessment, 2*, 594-613. https://doi.org/10.1002/9781118411360.wbcla083

Seeber, K. G. (2017). Multimodal processing in simultaneous interpreting. *The handbook of translation and cognition*, 461-475. https://doi.org/10.1002/9781119241485.ch25

Vandergrift, L. (1999). Facilitating second language listening comprehension: Acquiring successful strategies. *ELT Journal, 53*(3), 168-176. https://doi.org/10.1093/elt/53.3.168

Vandergrift, L. (2004). 1. Listening to learn or learning to listen? *Annual review of applied linguistics*, *24*, 3-25. https://doi.org/10.1017/s0267190504000017

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

Wijnants, M., Coppers, S., Rovelo Ruiz, G., Quax, P., & Lamotte, W. (2019, October). Talking video heads: Saving streaming bitrate by adaptively applying object-based video principles to interview-like footage. In *Proceedings of the 27th ACM International Conference on Multimedia* (pp. 2449-2458).

Woottipong, K. (2014). Effect of using video materials in the teaching of listening skills for university students. *International Journal of Linguistics, 6*(4), 200.

Yuan, T., & Fengping, Y. (2021). Construction and Application of Multi-modal Translation Teaching Mode Under Media Turn. *International Journal of Education, Culture and Society*, *6*(6), 198.