Effects of Textual Visuals on L2 Academic Listening Difficulty

Roman Lesnov

Northern Arizona University

Abstract

This study compared second language test-takers' performance on an academic listening test in an audio-only versus an audio-video mode. A new method of classifying video-based visuals was developed and piloted, which used L2 experts' opinions to place the video on a continuum from being content-deficient (not helpful for answering comprehension items) to content-rich (very helpful for answering comprehension items). The video for one testlet contained only the speaker's non-verbal cues and was found to be content-deficient. The other video contained non-verbal cues overlapping with PowerPoint text and was deemed content-rich. Seventy-three ESL learners participated in the study. The video type classification method was shown to be reliable and practical. The results of the Rasch analysis showed no significant impact of condition, either the content-deficient or the content-rich, either at the testlet level or at the item level. Possible reasons and implications of these findings are discussed.

Effects of Textual Visuals on L2 Academic Listening Difficulty

Background

Modern technology has dramatically altered the way second language (L2) academic listening is taught worldwide. Reflecting visually rich characteristics of the target language use (TLU, Bachman & Palmer, 2010) domain, academic L2 listening classes nowadays are filled with different kinds of new media such as videos and power point presentations (Lynch, 2011). New media are now ubiquitous in L2 education practices such that they change the nature of L2 education, making it more interactive and multimodal (Royce, 2007).

While the multimodal nature of the L2 *listening* competence is mainly accepted by L2 scholars, the field of L2 listening assessment keeps operationalizing L2 listening as an auditory skill. For instance, existing standardized high-stakes tests of academic English proficiency are mainly visual-free (Kang, Gutierrez Arvizu, Chaipuapae, & Lesnov, 2016). To eliminate this mismatch, a growing number of researchers have advocated for the inclusion of visuals in L2 listening tests (e.g., Ockey, 2007; Suvorov, 2015, Wagner, 2008). Researchers' arguments stemmed from the effects of videos on the difficulty of a listening message.

Research into video effects on L2 listening comprehension has three major gaps. First, no attempts have been made to unravel the role of textual visuals in L2 listening comprehension construct. Studies that did investigate the effects of textual visuals were mainly situated in the context of first language (L1) listening comprehension, or L2 language teaching rather than testing, or were entwined with the concept of advance organizers and captions (e.g., Baltova, 1994; Blockzijl & Andeweg, 2007; Chung, 1999).

Second, all previous comparative studies failed to control for the relationship of comprehension questions to video input. None of the studies specified to what extent video-based clues could lead test-takers' to the correct choices. One way to account for this would have been to quantify the degree of individual items' dependence on the video stimulus. This could have made comparisons of listening difficulty by mode more informative and advanced the field in terms of investigating the role of individual items in video-based listening test difficulty.

Third, previous research has barely attempted to investigate how individual listening items functioned under the video condition. Batty (2015) and Wagner (2010b) seem to be the only studies that explored the effect of videos on L2 listening comprehension at the item level. As part of the Batty's study, 164 EFL university students of different proficiency levels were administered a listening comprehension test in the two modes – audio and video. The Multi-Faceted Rasch Analyses yielded no general effect of mode. Nor were interactions with text type or proficiency detected. However, the subsequent bias analysis discovered that four items displayed mode-based differences in difficulty. Two items were easier under the video condition while the other two were easier in the audio mode. Among possible reasons for these interactions, Batty mentions gestures, facial expressions, and poor acting as exerting either facilitating or debilitating effect on the comprehension of the items.

Research Questions

1. How does listening difficulty depend on delivery mode, video type, and L2 listening proficiency?

2. How do individual item difficulties depend on delivery mode, degree of item video-dependence, and L2 listening proficiency?

Methods

Participants

Archived data was obtained that came from de-identified students in the Program in Intensive English (PIE) in the Northern Arizona University. The data were the students' scores on the listening section of the Fall 2016 PIE placement test. The overall group of the PIE placement test-takers ($n = 73$) is believed to be typical of students' population in similar programs across the USA. The majority of the test-takers were 18 to 25 years old Chinese and Arabic international students.

Instruments

Listening section of the placement test. The listening section contained seven passages of three text types – conversations, lectures, and news reports. Each passage was read once and was followed by five to six 4-option multiple-choice questions assessing students' ability to identify main ideas (20.51%), details (41.03%), and make inferences (38.46%). Each question was dichotomously scored. Then the scores for the questions were added up, and the overall result was converted to the 30-point scale.

Pilot test. Two listening testlets were developed by the researcher and included into the placement tests as pilots. The scripts and test questions for these two testlets are provided in the Appendix. Test-takers' scores on these two testlets were the primary source of data. The strategies used for developing these testlets are described below.

Videos were searched for among authentic academic video materials on YouTube. The two videos had to contain different kinds of visuals. One video contained elements of non-verbal communication (NVC), allowing no or a minimal amount of content-related clues. This video was labeled as NVC; it was on the topic of cybersecurity. The other video contained mostly

content-related clues in the form of PowerPoint text. The content-related clues might overlap

with the speaker's non-verbal cues (i.e., displayed simultaneously). This video was labeled

NVC+T (non-verbal cues plus text) and was about children's first language acquisition (see the

Appendix). The use of these videos was in keeping with Paragraph 107 "Fair Use" of the US

Copyright Law ("Copyright Law of the US", 2011).

Procedures

  Test-takers' scores. The whole group of 73 test-takers took the placement test in four

different classrooms, with about 20 students in each. After the test-takers finished the main part

of the placement test, each of the four groups were administered the pilot test consisting of the

two developed testlets. Two versions of the pilot test were developed that were counter-balanced

for the order of delivery mode (i.e., audio-only first vs audio-video first). Two of the four groups

took version 1 while the other two were administered version 2. Audio-only equivalents of the

testlets were created by removing the video channel in the Mac iMovie software (Version

10.1.3).

  ESL experts' judgements on items' video-dependence. To determine the degree to

which visual input could help answer each individual item within the pilot listening test (i.e.,

item video-dependence), a survey was developed. The survey contained the following question:

To what degree can the video-based visual clues help a test-taker to obtain a correct answer for

each individual item? This question was to be answered by ESL experts with a number on a

semantic differential scale ranging from 1 ("Not helpful") to 7 ("Very helpful"). Three ESL

teachers from the PIE were recruited who held their Master's degrees in Teaching English as a

Foreign Language or a related field and were professionally involved. All of them were pursuing

a Ph.D. degree in applied linguistics. The percent of agreement between the three experts was 87% for the Cybersecurity testlet and 77% for the Language testlet.

Analysis

For RQ 1, video type was a two-dimensional variable. The first dimension measured how rich a video was in content-related clues. It was the experts' judgements of the helpfulness of the video for answering comprehension questions, averaged over the set of items in a corresponding testlet. These average values were 1.27 for the five items in the Cybersecurity testlet, and 6.13 for the five items in the Language testlet. On the "content-deficient (1) – content-rich (7)" continuous scale, Cybersecurity and Language were situated nearly at the two bipolar ends. The second dimension was nominal, defining the kind of visuals in a video. Since the Cybersecurity video consisted of the speaker's non-verbal cues (NVC) only, it was designated as NVC on the second dimension. The Language video had textual visuals (T) overlapping with non-verbal cues, and was labeled as NVC+T. Two resulting values of video type were 1.27 NVC and 6.13 NVC+T for Language and Cybersecurity testlets respectively.

Test-takers' listening proficiency ranged from low ($n = 17$) through medium ($n = 41$) to high ($n = 15$). This categorization was based on test-takers' scaled scores on the placement test, excluding the scores on the Cybersecurity and Language testlets. The following ranges of scores were chosen to determine each of the three proficiency levels: 0-16 (low), 17-21 (mid), and 22-30 (high).

To answer RQ1, Rasch interaction/bias analysis was modeled with interactions between the facets set as follows: listening difficulty, delivery mode, and video type, and listening difficulty, delivery mode, video type, and proficiency. In regards to RQ2, the ten dependent variables were Rasch difficulty logits for the ten individual items. The independent variables

were identical to RQ1 except for item video-dependence. Item video-dependence was ratings of how helpful a video was for answering a particular item, averaged for each item across the three experts' ratings. Video-dependence values for items 1 to 5 as part of the Cybersecurity testlet were 1.00, 1.67, 1.00, 1.67, and 1.00. Video-dependence values for items 6 to 10 in the Language testlet were 7.00, 4.33, 7.00, 5.67, and 6.67. These values were not part of the Rasch model; they served for descriptive analysis of how interaction between items and delivery mode could be associated with item video-dependence.

Results

First, Rasch interaction analyses between video type and delivery mode revealed no significant differences. Table 1 shows the following information for each of the video types: listening difficulty when video is absent (audio target measure) and present (video target measure), target contrast (difference between the target measures), and its significance. We can see that videos of neither 1.27 NVC nor 6.13 NVC+T type affected test-takers' listening comprehension.

Table 1

Bias/Interaction Analysis: Listening Difficulty, Delivery Mode, and Video Type

Video Type (testlet)	Target measure (S.E.)		Target contrast	Joint S.E.	t	Welch *d.f.*	*p*
	Audio	Video					
1.27 NVC (Cybersecurity)	10.10 (.33)	9.88 (.36)	.21	.49	.44	346	.66
6.13 NVC+T (Language)	10.11 (.35)	9.91 (.32)	.20	.47	.42	346	.67

Note: S.E. = Standard Error

Second, there was no significant interactions among listening difficulty, delivery mode, video type, and test-takers' proficiency. Looking at Table 2, we can see that both Cybersecurity and Language testlets were somewhat easier in the video condition for each of the proficiency group (all target contrasts are positive), though not to the point of statistical significance.

Table 2

Bias/Interaction Analysis: Listening Difficulty, Delivery Mode, Video Type, and Proficiency

Video Type (testlet)	Proficiency	Target measure (S.E.)		Target contrast	Joint S.E.	t	Welch $d.f.$	p
		Audio	Video					
1.27 NVC	High	10.31 (.69)	10.23 (1.06)	.08	1.26	.06	45	.95
(Cybersecurity)	Mid	10.13 (.54)	9.92 (.60)	.21	.81	.26	125	.79
	Low	9.94 (.53)	9.79 (.49)	.15	.72	.21	146	.83
6.13 NVC+T	High	9.80 (.99)	9.75 (.63)	.05	1.17	.05	44	.96
(Language)	Mid	10.07 (.56)	9.88 (.52)	.19	.76	.25	126	.81
	Low	10.22 (.50)	10.06 (.54)	.16	.73	.21	146	.83

Third, there was no significant interaction between individual item difficulties and delivery mode. Likewise, interactions between item difficulties, delivery mode, and test-takers' proficiency did not reach statistical significance. Since there were no significate effects of videos on any item, analysis of the interaction between item difficulties and item video-dependence was not needed.

Relevance to PIE and Second Language Learning

The results of this study may not be conclusive. On the one hand, they do not seem to empirically support the inclusion of digital videos, either content-deficient or content rich, in L2 academic listening tests. On the other hand, the item video-dependence analysis showed that video-based textual visuals have a considerable potential to facilitate comprehension of individual listening items. Thus, the Language video was content-rich, with visual text being the abundant source of content-related clues. Though almost each item in the Language testlet had a high degree of video-dependence, no effect of the Language video on any item was found.

One explanation to this conundrum may be that test-takers did not pay much attention to the video. According to Wagner (2010a), test-takers' viewing behavior can vary to a large extent, which could be the case in this study. The two pilot testlets came at the end of the listening section of the placement test, the time when test-takers' fatigue was the highest. Also, all the non-pilot placement testlets had been delivered in the audio-only mode. Even though test-takers' were told that one of the last two testlets would be video-based, it is possible that some test-takers did not bother to adjust their viewing behavior. Having become accustomed to the audio-focused listening while looking down on the test paper, they might have found it unnecessary to attentively watch the video. Future studies should do a better job controlling for test-takers' viewing behavior and balancing audio- and video-based stimuli.

This study may have an implication for PIE. It seems that the use of either content-deficient or content-rich videos in PIE placement tests is not justifiable at this point. New studies are needed that would overcome the abovementioned limitation and generate more credible conclusions. This implication may not be applicable to the administration of the program's achievement tests, however. Since PIE students are used to watching videos of different types during their listening and speaking courses as well as achievement tests, videos can be expected to generate a more focused viewing behavior and, thus, exert a noticeable effect on achievement listening difficulty. This finds support in Lesnov's (2016) PIE internal research report.

References

Bachman, L. F., & Palmer, A.S. (2010). *Language Assessment in practice.* Oxford: Oxford

    University Press.

Baltova, I. (1994). The impact of video on the comprehension skills of core French students.

    Canadian Modern Language Review, *50*, 507–531.

Batty, A. O. (2015). A comparison of video- and audio-mediated listening tests with many-facet

    Rasch modeling and differential distractor functioning. *Language Testing*, *32*, 3-20. doi:

    10.1177/0265532214531254

Blockaijl, W. & Andeweg, B. (2007, October 1-3). The Effect of text slides compared to

    visualizations on learning and appreciation in lectures. Paper presented at the 2007 IEEE

    Professional Communication Conference (USA). Retrieved from

    https://www.researchgate.net/profile/W_Blokzijl/publication/4322061_The_Effect_of_Te

    xt_Slides_Compared_to_Visualizations_on_Learning_and_Appreciation_in_Lectures/lin

    ks/54197bff0cf25ebee98861e0.pdf

Chung, J. (1999). The effects of using video texts supported with advance organizers and

    captions on Chinese students' listening comprehension: An Empirical Study. *Foreign

    Language Annals, 32,* 295-308.

Copyright Law of the United States and Related Laws Contained in title 17 of the United States

    Code. (2011). Retrieved from http://www.copyright.gov/title17/circ92.pdf

Kang, T., Gutierrez Arvizu, M. N., Chaipuapae, P., & Lesnov, R.O. (2016). Reviews of

    Academic English Listening Tests for Non-Native Speakers. *International Journal of

    Listening, 0,* 1-38.

Lynch, T. (2011). Academic listening in the 21st century: Reviewing a decade of research. *Journal of English for Academic Purposes, 10,* 79-88.

Ockey, G. J. (2007). Construct implications of including still image or video in computer-based listening tests. *Language Testing, 24,* 517–537.

Lesnov, R. (2016). *Using videos in ESL listening achievement tests: The effects on difficulty.* PIE Internal Research Report. Flagstaff: Northern Arizona University. Retrieved from https://nau.edu/uploadedFiles/Academic/CAL/PIE/2015-2016_Research/Lesnov_video%20and%20test%20difficulty.pdf

Royce, T. D. (2007). Multimodal communicative competence in second language contexts. In T. D. Royce & W. Bowcher (Eds.), *New directions in the analysis of multimodal discourse* (pp. 361-90). New York, NY: Erlbaum.

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, 21,* 1-21.

Wagner, E. (2008). Video listening tests: What are they measuring? *Language Assessment Quarterly*, *5*, 218–243.

Wagner, E. (2010a). Test-takers' interaction with an L2 video listening test. *System*, *38,* 280-291.

Wagner, E. (2010b). The effect of the use of video texts on ESL listening test-taker performance. *Language Testing, 27*, 493-513.

Appendix

Listening Testlets: Sources, Scripts, and Comprehension Items

Testlet 1. Cybersecurity Challenges

Source title: Security expert Michael Chertoff discusses cybersecurity challenges, solutions
Speaker: Michael Chertoff
Link: https://www.youtube.com/watch?v=3MkFO6EALI8
Clip time: 17:45 – 22:00

First draft: 07/2016
Piloted: 08/2016
Completed:

Script:

The reality is when you are online there is no way to be sure that the person you think you are communicating with or the website you're … are going to is really that person or that website. There is no 100% certainty with the basic architecture of the network. And so … we've had to think about how do you manage this problem. The problem is maybe best encapsulated by a New York … New Yorker magazine cartoon – I think it goes back fifteen to twenty years – it's back in the days of big clunky ah PCs on the desk. And there's a drawing of a PC and there's two dogs talking to each other. And one dog says to the other: "On the Internet nobody knows you're a dog." And in many ways that sums up the problem. So with this lack of trust and with the ability of people to mascaraed as others and use it as a way to gain entry to our own networks, what we've seen again and again is the capability that people have if they're bad actors to corrupt information, to steal information, to deny access or introduce latency or delay in the transmission of information, to destroy and overwhelm networks and of course to steal all kinds of inf information for financial gain. If I would group … these types of consequences, I would say in the main they fall into three main categories. The one that's maybe the most long-standing set of security challenges and the one that we still read about the most and probably the one that touches us personally the most is the use of the network to steal financial information for the purposes of committing fraud – identity information, credit card information, access to bank accounts. Ah as you've read there've been literally millions of dollars stolen in this way. In the last couple of years, for example, there was one organized criminal effort to gain access to ATMs. What they did was they hacked into a couple of firms overseas, they were managing debit cards and ATM withdrawal cards, and they had the withdrawal limits on those cards removed. Then, on a single day, ahhp individuals working as part of this conspiracy were sent out to ATM machines all over the world to withdraw all the money from the machines. Because the withdrawal limits were gone, they could take every cash bit of cash that was in those machines. And on a single day before it was shut down tens of millions of dollars were stolen. So, that's a classic example of the fact that because the Internet is now where the money is, it's like [??]. To paraphrase [??], you don't have to rob banks any more by going in with a gun – you just rob it through the ATM or the credit card. A second area of things that we have seen are denial of service attacks. Ah these aren't maybe the most sophisticated attacks, they don't

ultimately destroy ah systems or networks, they don't kill people, but they interfere with the ability to get access to your … perhaps your bank or some other facility that you need to communicate with. And they create an enormous burden and dragging expense for enterprises. But the third and most consequential from a national security standpoint, the third type of category of attacks we worry about are attacks that actually could be corruptive or destructive. Imagine what would happen if ah malevolent actors penetrated into banks and were able able over a period of time, in a very subtle way, to change bank records. If you didn't have a back-up for transactions, you might have a crisis of confidence in banks something like what we saw in 2008 when we had our financial crisis. You could have destruction of critical infrastructure but unlike in Sony which destroyed business enterprises, tools and and and information technology architecture, you could actually have attacks on critical infrastructure that deals with transportation – the train that I came up with, the airplane I'm flying, maybe power. And that could actually cause loss of life as well as significant economical property damage.

LISTENING 1: Listen to part of a talk about cybersecurity.

1. The speaker discusses the cartoon about two dogs in order to _____.

    A. lecture about early New York internet animation
    B. illustrate the problem of trusting online resources
    C. point to the danger of the internet for animals
    D. introduce the solution for the lack of trust online

2. Based on the lecture, the cybercrime that would touch people personally the most is _____.

    A. robbing a bank's ATM machine
    B. destroying a government infrastructure
    C. stealing credit card information
    D. denying access to a bank website

3. The speaker said that individuals were able to take all the money from ATMs because _____.

    A. the power in banks was shut down
    B. bank workers robbed with a gun
    C. withdrawal limits on debit cards were removed
    D. the ATM machines were broken

4. For national security, the most serious category of cybercrimes is _____.

    A. stealing financial information
    B. denial of service attacks
    C. corruptive or destructive attacks
    D. robbing a bank with a gun

5. This lecture is mainly about the _____.

    A. access to cartoons about dogs
    B. problems of insecure ATM machines
    C. lack of trust among modern people
    D. security problems of online systems

Testlet 2. First Language Acquisition

Source title: First language acquisition
Speaker: Martin Hilpert
Link: https://www.youtube.com/watch?v=up0yVJWf9zQ
Clip time: 0:00 – 3:47

First draft: 07/2016
Piloted: 08/2016
Completed:

Script:

Hello there and welcome back to the introduction to English linguistics. In this video I'd like to talk about language acquisition - how do children learn a first language. And to start out with, let me give you a few basic facts about language learning. First of all, there is no genetic predisposition for learning any one particular language. A baby born to English-speaking parents will of course learn English but the same baby, if it grows up around people talking in Finnish or in Mandarin or in Sinhalese or in Welsh, will acquire any of those languages with the same speed and ease. All human languages are equally easy to acquire as a first language and not only that - children can acquire two or more first languages with ease. Yeah. Ahh having two or more first languages – that's called bilingualism or multilingualism and it has been shown that there are strong cognitive advantages to being bilingual. Bilinguals they have two language systems in their mind and in order to use one, they have to inhibit the other, so they have to concentrate on one thing and defocus another thing. And you can imagine that this helps in a whole lot of other cognitive tasks – you concentrate on one thing and selectively ignore the other thing. Right. More facts about language acquisition. Ahhm I said that the process seems to be effortless - very easy and very rapid so that all essential parts of language – the grammatical structures, pronunciations, all of that, is in place by age five to six, so there kids talk pretty much like adults. Now of course they don't talk completely like adults – they don't have the same capabilities that adults have. Think of telling a good joke or understanding irony. There kids catch up over the years, but in terms of grammatical rules, pronunciations, knowledge of different words – the basics really are in place by age five to six. All this happens without formal instruction. You don't have to tell kids: this is right, this is wrong, this is what the rules are. They figure that out by themselves and, interestingly, the outcome is almost always the same. Everybody learns how to talk and ah even though there may be some people that talk really really well, that are super eloquent, that know how to talk in public, ahhm … well this is a skill that you have to learn as an adult. Yeah ahh everybody learns instinctively how to talk well enough to hold a conversation. Right. Now there are certain puzzles associated with language acquisition. For one thing, kids say things that they've never heard before. How do they do that? Kids get things right without being corrected. How is that? How do they figure that out? And then they master grammar by age 5 but they don't master things that are equally complex or comparable to language like mathematics, differential equations. Mmm they have trouble doing that at age 15 and yet at age five they chatter away, yeah, they have trouble tying their shoelaces but they use relative clauses – that seems to be remarkable. Now linguists try to explain these puzzles with theories of language acquisition.

LISTENING 2: Listen to a conversation about language.

6. According to the speaker, which statement is correct?

        A. Children learn some languages faster than others.
        B. Any human language is learned equally easily by children.
        C. Learning two first languages is difficult for children.
        D. Some children are slower at learning languages.

7. Children will most likely _____ by age 6.

        A. need instruction to speak well enough
        B. be able to tell many good jokes
        C. be able to speak quite well
        D. learn the skill of public speaking

8. One of the puzzles that linguists try to explain is how children can _____.

        A. say something they heard before
        B. understand language theories by age 15
        C. solve mathematical equations by age 5
        D. speak accurately without being corrected

9. The lecture is mainly about _____.

        A. the importance of formal language instruction
        B. how children learn their first language
        C. the benefits of learning two or more first languages
        D. how children learn to talk in public

10.The teacher will probably talk next about _____.

        A. the lives of famous linguists
        B. what languages children should learn
        C. language acquisition theories
        D. how adults learn a foreign language