

The Effects of Video Media in English as a Second Language Listening Comprehension Tests

Zsuzsa Cziráky Londe
University of Southern California

The emergence of powerful computers in language testing permits the use of video media in second language computer assisted listening comprehension tests. Little research is available on what the effects of the video media are in listening comprehension test tasks. The present study examines two video formats (close-up view of the head of the lecturer, and full body view of the lecturer) and compares these to the audio-only format in a listening comprehension test setting. A simulated UCLA classroom lecture was videotaped and used, and one hundred and one students took the test. The aim of the research was to explore whether there were any performance differences when students took these tests in the different formats. The results of the present study show that the addition of the visual channel does not contribute to or take away from the performance in English as a second language listening comprehension test.

INTRODUCTION

Powerful computers in language testing today allow the use of video media in second language (L2) computer assisted listening comprehension tests, thereby offering the potential for creating a more authentic, and more realistic test of communicative competence (Ginther, 2002). Before the practical applications of a valid test, however, the theoretical framework of the construct needs to be determined (Bachman, 1990). In the case of the video assisted listening comprehension the question becomes whether the construct of listening comprehension changes with the addition of the visual channel (versus the traditional audio-only channel) since it is possible that the visuals carry additional information that is not available in audio-only format, thereby altering the construct. A priori one would expect better performance on the video input because cues from visuals might add to comprehension (Kellerman, 1992). To test this hypothesis an audio-only input was compared with two formats of video input. The reasoning behind the two different formats of video input was to test whether a close-up version on the lecturer's head elicited different results than the whole-body format. One could hypothesize that a close-up version would add to comprehension because of better "reading" of the mouth, alternatively it could be that a whole-body format would allow test takers to better "read" the lecturer's gestures and add to comprehension (Kellerman, 1992). Empirical evidence is required to define the construct validity of score-based inferences, i.e. the relationship between test performance and the measured abilities (Bachman, 1990) in listening comprehension. Little research has been available in this area in second language assessment.

The empirical investigation of the construct and validity thus will address the possible effects of the additional channel of visuals on test takers' performance (Bachman, 1990). A news-clip, an extreme close-up of a lecturer, a professionally edited film presentation, or additional special effects accompanying the audio text might affect test takers' performance, of which test designers need to be aware. In present study the consequences of two of such effects will be investigated.

It is also possible that students will perform similarly when taking the test in the various formats but will arrive at the answers through different processes. If this were the case (and we could track the different processing methods) the construct of listening comprehension would have to be redefined.

The purpose of this study was to investigate the differences, if any, in test performance when students are given listening comprehension tests in differing formats of video and audio.

RESEARCH QUESTION

To what extent will test takers perform differently in "audio-only" versus "talking head" (TH) versus "full body" (FB) video-assisted second language comprehension tests?

To test listening comprehension two video and one audio format were used. The video formats were 1) the "talking head" (TH), which was a close-up shot of the lecturer's head and face, which filled the screen; 2) the "full body" (FB) format, which was a long shot of the whole body view of the lecturer with some background and foreground information. The "audio-only" (audio) format was the sound track from the video, without any visuals. The target language use domain (TLU) is academic English as a second language. All three formats presented the same ten-minute simulated academic lecture.

Assessing listening comprehension

The "invisible cognitive operations" (Brindley, 1998) make it difficult to define the construct of listening comprehension. In simple terms, listening comprehension consists of understanding the information input, processing that information, and giving evidence of the understanding. In real life, the evidence could be a response to a question; in testing students could be asked to answer several questions about a 3-10 minute input.

According to Buck (1990) the entire process could be best described as follows:

...any model of the normal process of listening comprehension must allow the sum total of the listener's knowledge, past experience, current thoughts feelings, intentions, personality and intelligence to interact freely with the acoustic input and with each other, to create the interpretation of a text. Processing has to be massively interactive and parallel (1990:409)

This interactive aspect of the process makes it difficult to attribute one single

ability or skill that is measured when interpreting listening comprehension test results (Brindley, 1998). Another confounding factor, Buck (2001) points out, is the individuality of the process; listening is a very individual, personal process.

Historically there have been three overall approaches to language testing, and to listening comprehension testing: discrete-point, integrative and communicative approaches. Lado (1961) was the main advocate of discrete-point testing. In this approach he suggested that listening comprehension first and foremost is the process of recognizing the sounds of a language; a test therefore should check whether the student understands the utterances or its crucial parts (1961:208). In a discrete-point test segmental phonemes, vocabulary, stress, intonation, grammatical structures would be assessed in order to evaluate student's listening comprehension ability.

In integrative testing the learner's ability to use the learned pieces of language is assessed integratively. Oller (1979), one of the main representatives of this belief suggested that the whole is greater than the sum of its part, and that using the language, rather than knowing about it is essential in language learning.

Communicative second-language teaching brought with it the trend of communicative testing. Carroll (1972) argued that the purpose of language knowledge is to use the language in the target language use situation for the purpose of communication rather than knowing about the language. This led to the distinction of competence versus performance (Morrow, 1979), the latter giving evidence of actual "use of the language in ordinary situations" (1979:148).

Bachman's (1990) definition of abilities seems to best reflect the complex, dynamic, and interactive nature of listening comprehension: by testing language knowledge (language competence and performance) and strategic competence (metacognitive strategies) we will be able to test the understanding of the input and the interpretation of meaning.

Traditionally, listening comprehension has been tested with a short audio passage (either live or from a tape), on which students were tested; i.e. questions probed students' understanding after they had heard the passage. Adding video media can change this test task. Gruba (1989) compared a seven-minute audio and video lecture and found no difference in students' performance on the two tasks. In a later study, he investigated test takers' performance on a task that involved the use of a video newsreel in combination with audio text (Gruba 1999) and found that the higher the language proficiency, the less the reliance was on visual aids. From these findings Gruba concluded that the discrepancy between the content of the visuals and the audio message can hinder test takers' understanding. This seems to correspond with April Ginther's (2002) study in which she found that content visuals have a facilitating effect, but context visuals can be debilitating. Feak and Salehzadeh (2002) found in their research that the use of video did not have any adverse effect on test takers. According to Progosh (1996), learners would prefer video-mediated listening tests to audio-only tests, and Lynch (1998) suggests that video-media use may increase students' motivation and attention. According to Kellerman (1992) body language affects listening comprehension to decode mean-

ing but there is no information how it influences second language learners (in an unknown visual context) in an assessment context.

Content of the test

A lecture is one of the possible authentic academic formats of a listening comprehension test. Olsen and Huckin (1990) suggest that ESL students in a lecture situation might understand all the words individually, but do not necessarily understand the main ideas of the lecture. Research has shown that redundancy (Chiang and Dunkel, 1992) and discourse markers (Chaudron and Richards, 1986) in a lecture facilitate both L1 and L2 listeners' understanding, thus video streaming should facilitate students' comprehension. It is therefore essential to investigate how the video format affects the listening comprehension construct.

METHODOLOGY

The goal of this study was to evaluate and compare test takers' performance in three different listening comprehension formats. A quasi experimental design was employed, in which three conditions were used. An analysis of variance (ANOVA) was applied to investigate the differences in test performance. The delivery formats were the three independent variables, and the scores on the tests were the dependent variables. The sample population could be considered fairly homogenous, given their academic background and the level of their ESL proficiency. Participants were of diverse linguistic and cultural backgrounds from diverse academic fields in undergraduate and graduate studies at a university between 19–28 years of age. Their stay in the US ranged from two weeks to ten years with an average stay of 3 years. All participants had taken some form of an English test before being accepted to the university, and all have participated in academic lectures either in their home country or in this country. One hundred and one students participated in the study.

The key materials consisted of two video and one audio recordings. All three formats were of the exact same content: a 10-minute simulated lecture. The lecture was prepared, timed and given by a lecturer in the applied linguistics department of a university. The lecturer was videotaped with two digital video cameras: one camera recorded the lecturer in an extreme close-up, which only showed her head, making her mouth-movement and facial gestures pronounced, and the other camera showed her whole body, which included some of the background- (black board) and foreground-features (three students listening) of the classroom. The subject matter of the lecture was the "models of reading," which is part of a regularly scheduled ESL class: "Advanced Reading and Vocabulary for ESL Students." The ten-minute lecture did not require or assume any previous knowledge about the subject matter; it simulated the first day of class of the course.

The test booklet contained twelve personal questions, (name, length of stay in US, etc.) and eleven open-ended test questions about the content of the lecture.

Two questions at the end of the booklet referred to their comfort level with the test. Students were assigned to the three formats from a group of ESL students of mid-high to high level of English proficiency. The proficiency level was within the same range based on their ESLP (English as a Second Language Placement) test results, which also included listening comprehension. The two levels of English proficiencies were present in all three groups, and these proficiencies were acknowledged based on the placement exam.

The goal was to have comparable number of participants in the three delivery formats. The formats were randomly assigned to three ESL classes. In this quasi experimental research design the three groups of 31, 38, and 32 participants were assigned to the audio-only, talking head, and full body formats, respectively. Students took the test once only, and in one format only.

All participants were tested in university classrooms or in university offices. First they were given the consent form to sign, and then a notepaper was distributed. Test takers were asked to take notes the same way they would in a regular classroom, knowing that they would be tested on the content of the ten-minute lecture. Next, the video/audio input was played. Promptly after the lecture was presented the test booklet was distributed and participants were asked to answer all the questions using their notes.

The tests were scored on correct answers, and partial scores were allocated to some composite answers.

RESULTS

In order to answer the question whether there was a performance difference due to the different delivery formats, an analysis of variance, (ANOVA) was used. The score distributions for the three groups are shown in Table 1:

Table 1: Descriptive Statistics

| Descriptive Statistics | | | | | | | |
|------------------------|-----------|-----------|-----------|-----------|------------|-----------|------------|
| | N | Mean | Std. Dev. | Skew-ness | Std. Error | Kurtosis | Std. Error |
| | Statistic | Statistic | Statistic | Statistic | Error | Statistic | Error |
| AUDIO | 31 | 10.65 | 4.10 | -0.76 | 0.42 | -0.20 | 0.82 |
| TALKING HEAD | 38 | 11.37 | 3.79 | -0.94 | 0.38 | 0.65 | 0.75 |
| FULL BODY | 32 | 10.81 | 3.07 | -0.15 | 0.41 | -0.99 | 0.81 |

The means of the three groups range from 10.65 to 11.37; the Talking Head format has the highest mean among the formats. The standard deviation is the low-

est in the Full Body format and highest in the Audio-only format, which suggests a wider distribution range in the Audio-only format.

ANOVA assumes a normally distributed population. To check this assumption the score distributions for the three groups were analyzed using histograms. The Audio-only scores showed a slightly negatively skewed distribution, and to test normality with parametric measure, the ratio of the statistic to its standard error, or z-score, were calculated. The values for these were -1.80 and -.23 for skewness and kurtosis, respectively. Since both of these values fall within -2 and $+2$ range, we can accept the distribution of these scores to be reasonably normal.

The distribution of the scores from the Talking Head format seemed to be slightly negatively skewed and after using the parametric measures, (the ratio of the statistic to its standard error) the z-score values were -2.47 and $.86$ for skewness and kurtosis, respectively. The skewness value shows a non-normal distribution tendency but kurtosis does fall within the -2 and $+2$ range; we can accept the distribution of these scores to be reasonably normal.

The distribution of the total scores of the Full Body video seems normally distributed, which was supported by the parametric measures, of z-score of $-.36$ and -1.22 for skewness and kurtosis, respectively. Since both of these values fall within -2 and $+2$ range, we can accept the distribution of these scores to be reasonably normal. ANOVA is a robust enough computation to tolerate these slight non-normalities of the score distribution.

The assumption of equality of variance was also checked (Levene's test), which was $.674$ ($p=.512$, $df=2, 98$, $F=3.089$). This indicates that the differences among the variances of the three groups are not significant, so homoscedasticity can be assumed. The assumption of the independence of measures was achieved by the test design, in which only one test-format was given to each participant thus did not have a chance to interact or influence each other about the test.

ANALYSIS AND RESULTS

The results of the one-way ANOVA are presented in Table 5 below.

Table 5: ANOVA

| TOTAL | Sum of Squares | df | Mean Square | F | Sig. |
|----------------|----------------|-----|-------------|-------|------|
| Between Groups | 10.097 | 2 | 5.049 | 0.372 | 0.69 |
| Within Groups | 1328.81 | 98 | 13.559 | | |
| Total | 1338.91 | 100 | | | |

According to the ANOVA analysis there is no statistically significant difference between the means of the performance scores of the three test groups. ($F(2,98) = .372, p = .69$). This means that the different treatment manipulations (the three delivery methods) did not affect the performance of the test takers; the mean score differences were not statistically significant at the .05 level. This finding suggests that test takers were not influenced by the input formats of the test; i.e. regardless of the format the three groups performed similarly.

DISCUSSION AND IMPLICATIONS

Present study suggests that second language learners are not influenced by the three different delivery formats in listening comprehension tests. This means that, although the addition of the visual channel, in theory, could add information to the input, this does not seem to affect test takers' performance on the test. It also is not the case that those who were tested on the audio-only version missed any of the information the video groups received, otherwise they would have had significantly lower scores. Test takers were neither hindered nor helped by the input formats. It seems that even though the delivery methods were different, information processing occurred with similar results.

The generalizability of the results is restricted to an academic environment, to mid- to mid-high level ESL proficiency, to a classroom setting and to testing circumstances.

If the results of this study can be further confirmed, test designers could choose to use these listening comprehension inputs without influencing students' performances in the process. This would give test designers the option of using material that might only be available on video/DVD or not to use video in their classrooms at all. Until definitive results a good approach in the classroom would be, to use various media. This would give students the possibility to build new test-taking strategies when confronted with different tests delivery formats.

Since almost all real-life listening experiences are accompanied by dynamic visuals (Feak, Salehzadeh, 2001), video media is a more authentic medium for testing listening comprehension than audio alone. This study suggests that using video for listening comprehension testing purposes will not change performance on such tests.

The present study also has practical significance in the field of computer assisted second language teaching and assessment. On-line lectures, and live video streaming of lectures have become more feasible, and are being widely used. It is possible that the more students are exposed to visual on-line learning the more comfortable they will feel with visual testing materials. Students in the pilot study were asked whether they would prefer audio-only or video assisted input, and 99% would prefer the video format even though only one-third received the video assisted input. It seems that students perceive the video input as more authentic, with possibly more information, or simply they feel more comfortable with information

accompanied with images.

Similarly to Coniam (2001) some students in the study intentionally covered their eyes or turned away from the images not to be distracted from the audio content. This, again, is a perceived efficiency since all groups performed similarly, and could stem from either a learning habit in the second language or from an individual's affinity toward more audio or visual processing tendencies.

Several related issues emerged during the investigation that were not the focus of present study but seem to be closely related and would require further research:

the effects of test takers' note taking habits (is there a correlation between the extent of note taking and performance?)

the implications of not watching the video when it is available (some students in the video assisted tests did not look at the video at all)

the issues of test takers' comfort level with the delivery method (students with more exposure to videos during learning a second language might have an advantage in testing)

More research in second language assessment on how to properly integrate video media will guide the assessment and teaching community in their use of this medium, the demand for which seems to be ever growing.

REFERENCES

- Alderson, J. C., & Urquhart, A. H. (1985). The effect of students' academic discipline on their performance on esp reading tests. *Language Testing* 2, 2, 192-204.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Bostrom, R. N. (1997). The testing of mother tongue listening skills. In c. M. Clapham and d. Corson (eds), *Language testing and assessment: Encyclopedia of language and education*. Dordrecht: Kluwer.
- Brindley, G. (1998). Assessing listening abilities. *Annual Review of Applied Linguistics*, 1998(18), 171-191.
- Brindley, G., & Nunan, D. (1992). *Draft bandscales for listening*. Sidney: National Centre for English Language Teaching and Research, Macquarie University, Australia.
- Buck, G. (1990). The testing of second language listening comprehension (unpublished doctoral dissertation, Lancaster University, United Kingdom).
- Buck, G. (1991). The testing of listening comprehension: An introspective study. *Language Testing*, 8, 67-91.
- Buck, G. (1994). The appropriacy of psychometric measurement models for testing second language listening comprehension. *Language Testing*, 11(2), 145-170.
- Buck, G. (2001). *Assessing listening*. Cambridge: Cambridge University Press.
- Burgoon, J. (1994). *Nonverbal signals*. Londong: Routledge.
- Capek, C. M., et al. (2004). The cortical organization of audio-visual sentence comprehension: An fmri study at 4 tesla. *Cognitive Brain Research* (20), 111-119.
- Chaudron, C., & Richards, J. C. (1986). The effect of discourse markers on the compre-

- hension of lectures. *Applied Linguistics*, 7(2), 113-127.
- Chiang, C. S., & Dunkel, P. (1992). The effect of speech modification, prior knowledge, and listening proficiency on EFL lecture learning. *TESOL Quarterly*, 26(2), 345-374.
- Cho, Y. (2002). Test review of advanced placement international english language (AP-TESTL). *Language Testing*, 19(3), 323-331.
- Carroll, J. B. (1972). Defining language comprehension. In R.O. Freedle and J.B. Carroll (eds.), *Language comprehension and the acquisition of knowledge*, New York: John Wiley and Sons.
- Clark, H. H., & Clark, E. V. (1977). *Psychology and language: An introduction to psycholinguistics*. New York: Harcourt Brace Jovanovich.
- 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), 1-176.
- Dale, A. M., & Halgren, E. (2001). Spatiotemporal mapping of brain activity by integration of multiple imaging modalities. *Neurobiology*, 11(Apr (2)), 202-208.
- Dunkel, P., & Davis, J. N. (1994). *The effects of rhetorical signally cues on recall*. Cambridge: Cambridge University Press.
- Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data*. Cambridge, MA: Bradford Books/MIT Press.
- Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data (revised edition)*. Cambridge, MA: Bradfordbooks/MIT Press.
- Faerch, C., & Kaper, G. (1986). The role of comprehension in second language learning. *Applied Linguistics* (7), 257-274.
- Feak, C. B., Salehzadeh, J. (2001). Challenges and issues in developing an EAP video, listening placement assessment: A view from one program. *English for Specific Purposes*, (20) Supplement 1, 477-493.
- Ginther, A. (2002). Context and content visuals and performance on listening comprehension stimuli. *Language Testing*, 19 (2), 133-167.
- Goetz. (1999). *Textbook of clinical neurology*. Philadelphia: W.B. Saunders Company.
- Gray, R. J., Braver, T. S., and Raichle M. E. (2001). Integration of emotion and cognition in the lateral prefrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America* (99), 4115-4120.
- Green, A. (1998). *Verbal protocol analysis in language testing research: A handbook*. Cambridge: University of Cambridge Local Examinations Syndicate.
- Gruba, P. (1989). A comparison study of audio and video presentation modes in tests of ESL listening comprehension. Unpublished MA thesis, University of California, Los Angeles.
- Gruba, P. A. (1999). *The role of digital video media in second language listening comprehension*. Unpublished doctoral dissertation, University of Melbourne, Melbourne.
- Kellerman, S. (1992). I see what you mean: The role of kinesic behavior in listening and implications for foreign and second language learning. *Applied Linguistics* (13), 239-58.
- Kirk, R. (1995). *Experimental design: Procedures for the behavioral sciences* (Third edition). New York: Brooks/Cole Publishing Company.
- Lado, R. (1961). *Language testing: The construction and use of foreign language tests*. London: Longman.
- Long, D. (1989). Second language listening comprehension: A schema-theoretic perspective. *Modern Language Journal*, 73, 32-40.

- Mayer, E. R., & Anderson, B. R. (1992). The instructive animation: Helping students build connections between words and pictures in multimedia learning. *Journal of Education Psychology*, 84(4), 444-452.
- Meyer, B. J., & Freedle, R. O. (1984). Effects of discourse type on recall. *American Educational Research Journal*, 21(1), 121-143.
- Morrow, K. (1979). Communicative language testing: Revolution or evolution? In C. J. Brumfit and K. Johnson (eds.), *The communicative approach to language testing*, pp. 143-57. Oxford: Oxford University Press.
- Nissan, S., DeVincenzi, F., & Tang, K. L. (1996). *An analysis of factors affecting the difficulty of dialogue items in TOEFL listening comprehension*. (TOEFL Research Report 51 ed.). Princeton, New Jersey: Educational Testing Service.
- Oller, J. W., Jr. (1979). *Language tests at school*. London: Longman.
- Olsen, L. A., & Huckin, T. N. (1990). Point-driven understanding in engineering lecture comprehension. *English for Specific Purposes* (9), 33-47.
- Parshall, C. G., Spray, J. A., Kalohn, J. C., & Davey, T. (2002). *Practical considerations in computer-based testing*. New York: Springer.
- Pressley, M., & Afflerbach, P. (1995). Verbal protocols of reading: The nature of constructively responsive reading. Hillsdale, NJ: Erlbaum.
- Progosh, D. (1996). Using video for listening assessment: Opinions of test-takers. *TESL Canada Journal*, 14(1), 34-44.
- Richards, J. C. (1983). Listening comprehension: Approach, design, procedure. *TESOL Quarterly*, 17(2), 219-240.
- Salomon, G. (1979). *Interaction of media cognition and learning*. San Francisco, CA: Jossey-Bass.
- Salomon, G. (1979). *No distribution without individuals' cognition: A dynamic interactional view*. Cambridge: Cambridge University Press.
- Scardamalia, M., Bereiter, C., McLean, R. S., Swallow, J., & Woodruff, E. (1989). *Distributed cognition*. Cambridge: Cambridge University Press.
- Valette, R. M. (1977). *Modern language testing*. New York: Harcourt Brace Jovanovics.
- Weir, C. J. (1993). *Understanding and developing language tests*. London: Prentice-Hall.
- Wescott, R. W. (1978). Visualizing vision. In B. S. Randhava and W. E. Coffman (eds.), *Visual learning, thinking and communication*, pp. 21-26. New York: Academic Press.

Zsuzsa Cziraky Londe received her BA in Linguistics from UCLA (summa cum laude) and an M.A. and Ph.D. in Applied Linguistics, also from UCLA. Languages, language assessment and teaching are her ongoing academic interests. She is a lecturer and the ITA-testing coordinator at University of Southern California. Working memory and its relationship to listening comprehension in a second language was the topic of her most recent research. Publications and research with other colleagues have taken her to areas of generative linguistics, psycholinguistics and language assessment.