McCrocklin, S. (2012). Effect of audio vs. video listening exercises on aural discrimination of vowels. In. J. Levis & K. LeVelle (Eds.). *Proceedings of the 3rd Pronunciation in Second Language Learning and Teaching Conference,* Sept. 2011. (pp. 178-186). Ames, IA: Iowa State University.

EFFECT OF AUDIO VS. VIDEO LISTENING EXERCISES ON AURAL DISCRIMINATION OF VOWELS

Shannon McCrocklin, Iowa State University

Despite the growing use of media in the classroom, one critical aspect of digital instruction has been largely ignored, the effects of using of audio versus video in aural discrimination tasks. To analyze the impact of the use of audio or video training on aural discrimination of vowels, 61 participants (all students in a large American university) took a pre-test followed by two training sessions on a vowel contrast (/i/-/ɪ/). One group received audio training and the other group received video training. The groups then took a post-test and delayed post-test to determine the impact of the training. For the 40 that met the requirements for data analysis (based on pre-test and completion of all training), results showed that while both groups improved significantly from the pre-test to both post-tests, the video and audio groups performed similarly (no statistically significant difference). The student reactions to the two training types were also obtained through a questionnaire. Results showed that reactions were more favorable to the video training.

Due to increased use of media in language classrooms, it is important to consider what effects these technologies might have in terms of performance, motivation, and attitude. While technology has been recognized as an important teaching tool for pronunciation, most of the emphasis has been on developing automatic speech recognition, visual feedback, and software use (for examples, see Levis & Pickering, 2004 and Wang & Munro, 2004). The effect of incorporating audio versus video in the classroom has largely been ignored. This is a problem because audio and video materials are commonly used by teachers in their classes, whether through teacher-recorded materials, websites, podcasts, or even software programs, without knowing whether the two have different impacts. This research study aims to investigate the effects of using audio and video pronunciation training on perception of English vowels.

Review of the Literature

Learning pronunciation in a second language (L2) involves more than just production; it also involves perception, learning to listen in the L2. For vowels in particular, learning to listen to English will often require that the subject develop vowel categorizations appropriate for English. English vowels are distinguished by the following characteristics: tongue position and height within the oral cavity, lip rounding or spreading, tension, and gliding (vs. simple vowels). Length is not a distinguishing feature, but instead is influenced by the vowel's phonetic environment (Celce-Murcia, Brinton, Goodwin & Griner, 2010).

Research has supported the idea that sounds in an L2 are filtered through the phonological system of the first language (L1) (Beddor & Strange, 1982; Blankenship, 1991; Flege, Munro, & Fox, 1993). Filtering through the L1 can lead an L2 learner to make distinctions that are inappropriate for the L2. English vowel pairings such as /i/ and /ɪ/, /e/ and /ɛ/, and /ɛ/ and /æ/ are likely to be problematic for learners, because vowels with similar articulatory positions are often difficult to discriminate. Learning to listen to the L2 then will entail a redefining of the vowel space to better reflect the vowel distinctions of English. However, filtering through the L1 does not entirely explain the use of vowel errors for L2 learners. Bohn and Flege (1990) show that non-native speakers often rely on vowel length, even if they do not do so in their L1. This means that learning appropriate vowel categories will also entail a shift from focusing on length of the vowel to quality of the vowel (spectral cues).

Learning to create aural discrimination categories based on spectral cues is not only important for comprehension, but is likely to play a role in production. Researchers have found evidence that practice in perception can improve production (Bradlow, 1997; Rochet, 1995; Rvachew, 1994). Thus, instruction should help students listen for and produce the articulatory differences of vowels and lead students away from a reliance on vowel length alone. One way of doing this is through the use of listening exercises that utilize minimal pairs. Research into the effectiveness of minimal pair listening training has shown that it can lead to significant improvement in perception (Bradlow, Pisoni, Yamada, & Tohkura, 1997; Pisoni, Aslin, Perey, & Hennessey, 1982; Strange & Dittmann, 1984).

Although it seems clear that these activities can be useful in increasing ability to discriminate vowels, there is still the question of whether presenting activities through video or audio (video here refers to video with audio) is more effective. Research on the effects of audio vs. video training with minimal pairs has shown that video can promote increased acquisition. For example, research on the /r/ and /l/ contrast in English shows that video training improves perception more than audio training alone (Bradlow et al., 1997; Hardison, 2003; Hardison, 2005). In a study on vowel contrasts, Hirata and Kelly (2010) found similar results for 60 L1 English speaking participants receiving training in Japanese vowels, which, unlike English, are contrasted through length differences. Results show that the added visual of the person saying the words improved perception more than audio training alone.

The increased improvement from the video training groups may be explained through information processing theory which accounts for this benefit by explaining that by using both auditory and visual information a student is able to use dual-coding and access information through multiple routes (Bagui, 1998).

Research is needed, however, to know whether utilization of visual cues will aid or hinder the development of aural discrimination categories for vowels in English. Based on previous research, it is hypothesized that extra modeling and visual cues will aid in the development of vowel categories. It is possible, however, that while students watch videos for training in pronunciation they rely on the facial movements to help them determine the vowel. This, in turn, could allow students to excel in the training activities without developing the ability to listen and use spectral cues in determining vowels. Thus, this research study aims to investigate the impact of audio vs. video training on subjects’ ability to aurally discriminate English vowels.

In addition, this research aims to examine student reactions to the different training delivery methods, which may affect the appeal of the exercises. This, in turn, could affect the students’ motivation to learn and the effectiveness of the training. Bagui (1998), for example, found that the introduction of animation, sound, and interactivity in lessons increased student motivation. Bagui, however, was examining interactive multimedia. It is not clear whether a switch from audio to video would also affect reactions to the training.

Research Questions

This research study thus aims to evaluate the effect of training on the discrimination of vowels, specifically, /i/ and /ɪ/, because these vowels do not contrast phonemically in many languages (Nilsen & Nilsen, 2002), are both frequent in English (Edwards, 1992), and have visible differences in facial movements when pronounced. The research study aims to answer two questions:

1. Will the group receiving audio pronunciation training differ from the group receiving video pronunciation training in their aural discrimination of /i/ and /ɪ/ in the post-test and delayed post-test?
2. Will students find video training more appealing than audio training?

METHODS

Participants

The participants were advanced ESL students enrolled in a college level writing class for ESL students at a large university in the United States. They were assigned to one of two groups: 30 to the video training group and 31 to the audio training group. The formation of groups attempted to control for factors such as native language, age, gender, length of time in the US, and length of English study overall, as well as for the pre-test scores to equalize for proficiency. Table 1 shows the make-up of each group.

Table 1

Group Formation Data

	Group 1-	**Group 2-**
	Video Training	**Audio Training**
N=	30	31
Native Language	80% Chinese 20% Other	87% Chinese 13% Other

Gender	M= 18
	F= 9
	Non-report= 3
		M= 23
		F= 7
		Non-report= 1

Years studying English	8.82	8.05
Months studying in U.S.	11.58	10.44
Pre-test score	17.53	17.6
SD of pre-test score	2.43	2.26

.

Materials and Procedures

The materials used in this study comprised of a pre-test, post-test (which was also used for the delayed post-test), audio and video training materials, a biographical data questionnaire, and a student reaction questionnaire administered after the rest of the study was completed. In the first session, after signing an informed consent form, subjects filled out the biographical data survey and took the pre-test. Both the pre-and post-test each contained twenty listening items. Participants were asked to mark on a sheet decisions about words such as, “Are these two words the same words or different words?” and “Does this single word have the “e” sound like in “feet”?”. The researcher, whose voice was used for all materials, carefully recoded each exercise to control for possible length differences, checking words through Audacity to ensure similar lengths (within .02 seconds of each other).

To ensure equivalence of forms, the items from the pre-and post-test used single, closed syllables for all words and controlled for the number of vowels that would be colored by nasalization, postvocalic [r] or [l]. To check the overall equivalency of forms, the items of the pre- and post-tests were mixed together into a single test taken by three ESL students not participating in the research study. The results indicated that the items were of similar difficulty.

For each of the two training sessions (Sessions 2 & 3), participants watched a video (group 1) or listened to an audio file (group 2). Both videos were a little over 13 minutes. In order to create audio files that were exactly the same (in sound and content), the audio was stripped from the video files by a program called Video MP3 extractor provided by geovid.com. These audio and video materials were provided to students for download through a website. In Session 3, participants also took an immediate post-test.

For Session 4, which occurred a week after Session 3, subjects took the delayed post-test and filled out the questionnaire, which included 5 Likert scale items about the appeal of the training materials. The questionnaire also included two open-ended questions to allow for individual comments on the training.

Analysis

Research question 1.

The pre-, post-, and delayed posttests were used to answer research question 1, whether the groups would differ in performance due to different training. The pre- and post-tests were scored for correct and incorrect answers. Because all subjects took all three tests it was possible to analyze the data using a mixed ANOVA.

Exclusion of subjects from analyzed data.

Some participants had to be excluded from the analyzed data for the pre-test, post-test, and delayed post-test comparisons. There were two possible reasons for exclusion; a participant not completing all four sessions of the research study, which eliminated 12 participants, or a participant receiving a perfect score on the pre-test. The rationale for the second exclusion possibility is that for these participants improvement due to training would not be visible in either post-test. This occurred in 8 cases. This resulted in 21 participants in the audio group and 19 in the video group.

Research question 2.

The student feedback questionnaire was used to answer research question 2, student reactions to the appeal of the training. All subjects that completed the questionnaire and both sessions of training were included in the analyzed questionnaire data (n=54). Reactions to the Likert scale items were scored on a 1-5 range with 5 representing strong agreement with the claim and a 1 representing strong disagreement. The average score for each item was calculated for comparison. Responses to the open-ended questions were coded by the researcher and explored for common themes and types of responses.

RESULTS

Participant Improvement from Pre- to Post- Tests

Results showed that both groups responded similarly to training; they both showed significant improvement (p= .000) from the pre-test to the post-test with an effect size of .70. Despite a decline in the average score from the post-test to the delayed post-test, students maintained a significant improvement from the pre-test to the delayed post-test (p= .008, effect size= .39). The decline from the post-test to the delayed post-test was not statistically significant. Table 2 reports the average scores for each group at each testing time.

Table 2

Scores for Pre-, Post-, and Delayed Post-tests by Group

	Video Group		Audio Group	
	Average	SD	Average	SD
Pre-test	16.63	2.52	16.95	2.01
Post-test	18.11	1.85	18.24	1.55
Delayed post-test	17.47	3.04	18.00	2.07

Although the improvement for the audio group was slightly higher than the video group (7.27% versus 6.43%), this difference was not significant. Also, the score decline from the post-test to the delayed post-test for the audio group was slightly lower than for the video group (1.2% versus 3.2%). This, however, was also not significant. Figure 1 shows the average score for each group on each of the three tests.

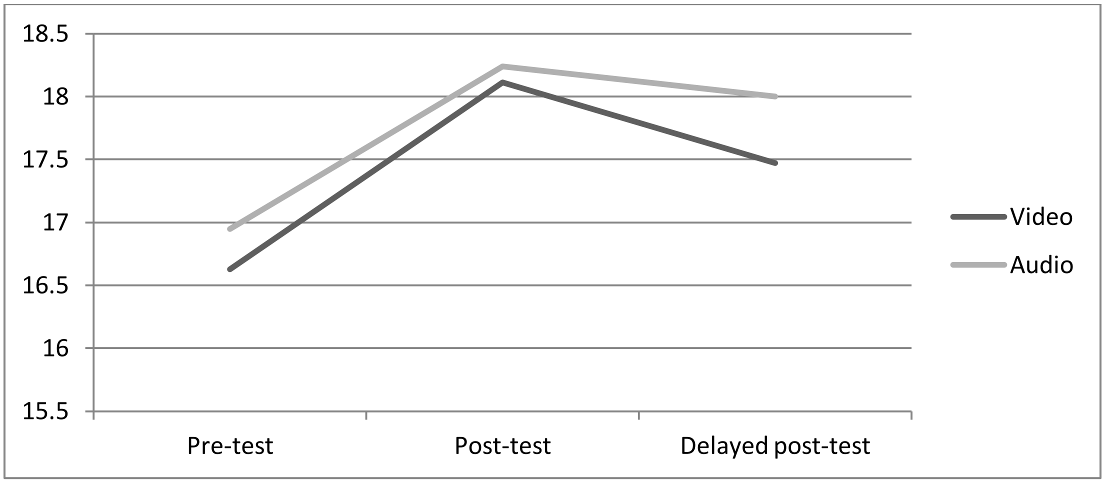

Figure 1 Scores for Pre-, Post-, and Delayed Post-tests by Group

Appeal of the Training

Overall, in response to the questionnaire, the video group gave higher scores for every item. Table 3 shows the five claims presented to students and the average score given for each item. As stated previously, a score of 5 indicates strong agreement while 1 indicates strong disagreement.

Table 3

Scores to Likert Scale Questionnaire Items by Group

	Video Group		Audio Group	
	Average	SD	Average	SD
The instructions for each activity were clear	4.55	0.51	4.43	0.68
The quality of the recordings was high	4.50	0.51	4.00	0.95
I feel that my ability to hear vowel differences has improved	3.75	0.55	3.57	0.75
I feel the training was interesting	3.55	0.76	3.38	0.74
I would like to do more training like this	3.85	0.81	3.05	1.20

This relationship can be better illustrated in Figure 2.

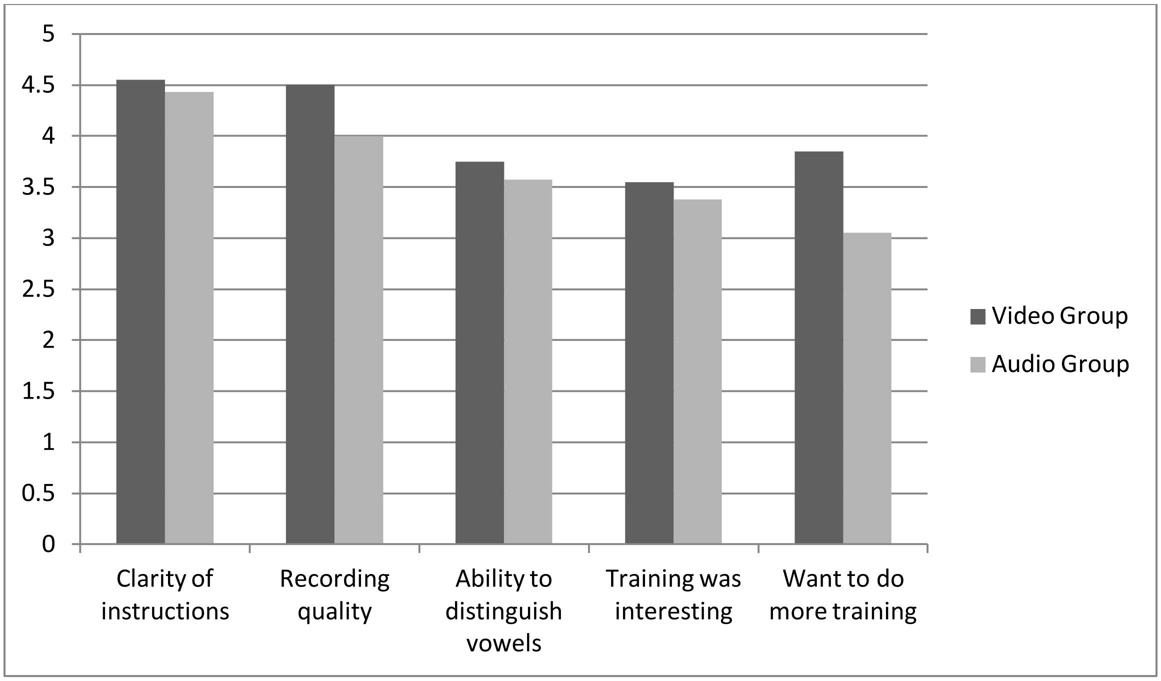

Figure 2 Scores to Likert Scale Questionnaire Items by Group

In response to the open ended question “The other research group received training through [opposite form]. Do you think this would have been better or worse than the training that you received? Please explain your answer" only five of the video group thought that the audio would have been better. In contrast, 12 members of the audio group thought that video would have been better.

Of the participants that thought video was better, 12 indicated that seeing the mouth movements was helpful for understanding how the sounds were made. One participant wrote, "I think [audio] would've been worse. I could learn how lips are different when pronounce different vowels through video." This utilization of visual cues was also mentioned as a negative by the participants who thought audio would be better. Three participants mentioned that because they could get the answers through looking at the different visual cues, audio would be better training. Four other participants answered more generally that because the goal was to improve listening, audio would be better. One participant said, "[Video is] worse. This activity is focus on listening. The audio is focus on listening. The video might make people focus on the screen." Also, two participants specifically mentioned that they thought the video would be distracting.

Although most of the participants responded to the question of which is better in terms of improvement, four participants responded in terms of appeal. They stated that the video would be better because it is more interesting (2 from each group). One participant stated, "I think video must be more interesting and attractive than just audio files."

In response to the question, "How do you think this training could be improved?" the most common response was that the training needed more difficult items and activities (6 for video group and 7 for audio group). Also, the next most common comment for each group was that the training needed more items and questions. Interestingly, although in the previous question, four people indicated that the video would be more interesting, more people in the video group indicated that the training could be improved by making it more interesting. Another interesting finding was that two members of the audio group wanted clearer directions, but none of the video group members indicated this. Finally, one member from each group indicated that they wanted personalized feedback from the training.

DISCUSSION

This study produced two main findings. First, in contrast to previous research (Hardison, 2003; Hardison, 2005; Hirata & Kelly, 2010) the introduction of video versus audio seems to have made little difference. This does not support the information processing theory, which claims that audio plus video would allow for dual coding and better storing and accessing of new information (Bagui, 1998). For teachers, this means that training for English vowels can be done through either method. For most teachers, audio recording, which can be done with free software such as Audacity, would be less time consuming and expensive.

Although the two training types produced similar results in terms of participant improvement, reactions were generally more favorable to the video training. This is in line with previous research that has shown that the use of multimedia can increase student motivation (Bagui, 1998). It seems that the change from audio to video can also produce changes in attitudes and reactions towards training. For teachers, this would suggest that by incorporating video (at least occasionally) teachers may be able to offset feelings of monotony and perhaps increase student interest.

It is important, though, that all findings be considered in light of the limitations of this study. One of the main problems encountered with this study was the ceiling effect caused by numerous high scores on the pre-test. Over 75% of the original 61 participants scored a 17 or higher on the pre-test (out of a possible 20). This left little room for visible improvement. It may be possible that, with a more sensitive pre-test, greater differences could have been found.

Also, although this study began with 61 subjects, data from 20 subjects could not be used for analyzing improvement from the pre-test to the two post-tests. With only 41 subjects, the generalizability of the results is uncertain. Future research with a greater number of subjects or with subjects at lower proficiency levels should be done to check these findings.

Future research should not only look to replicate these findings, but also expand them to include more pronunciation features. Thus far only one English vowel pairing, /i/-/ɪ/, and one English consonant pairing, /l/-/r/, have been investigated. Yet there are many other pairings that have clear differences in visual cues, such as /ʌ/ vs. /a/ or /θ/ vs. /t/ or /s/ that could add to the understanding of the impact of the visual cues.

Also, future research should look more closely at the impact on production for video vs. audio training. It may be that the visual clues, which provide modeling, may be more helpful for improving student production. This line of research could also be extended to include multimedia or software. For example, software could be designed to give answers and personalized feedback to students. This may satisfy the desires of students who want personalized feedback.

Because there has previously been relatively little interest in this area of research, there are many possible directions for future research. As teachers are already employing these modes of delivery in their classrooms and as homework, it is important that further research be conducted to determine the effects of these two methods of training delivery.

ABOUT THE AUTHOR

Shannon McCrocklin is a doctoral student in Applied Linguistics and Technology at Iowa State University. She received her Masters Degree in TESL at University of Illinois at Urbana-Champaign. Her research focuses on issues in the teaching of pronunciation. She has presented at CALICO and PSLLT. She can be contacted at Iowa State University**,** 206 Ross Hall**,** Ames, IA 50011**.** Phone: 217-251-3850. Email: mccrockl@iastate.edu

REFERENCES

Bagui, S. (1998). Reasons for increased learning using multimedia. *Journal of Educational Multimedia and Hypermedia*, *7*, 3-18.

Beddor, P.S. & Strange, W. (1982). Cross-language study of perception of the oral-nasal distinction. *Journal of the Acoustical Society of America*, *71*, 1551-1561.

Blankenship, B. (1991). Second language vowel perception. *Journal of the Acoustical Society of America*, *90*, 2252-2252.

Bohn, O.S. & Flege, J.E. (1990). Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics*, *11*, 303-328.

Bradlow, A. (1997). Training Japanese listeners to identify English /r/ and /l/. *Journal of the Acoustical Society of America*, *101*, 2299-2310

Bradlow, A., Pisoni, D., Yamada, R., & Tohkura, Y. (1997). Effects of audio-visual training on the identification of English /r/ and /l/ by Japanese speakers. *Journal of the Acoustical Society of America*, *102*, 3137-3137.

Celce-Murcia, M., Brinton, D., Goodwin, J. & Griner, B. (2010). *Teaching Pronunciation: a course book and reference guide.* Cambridge: Cambridge University Press.

Edwards, H. (1992) *Applied phonetics: The sounds of American English*. San Diego, CA: Singular Publishing Group Inc.

Flege, J.E., Munro, M.J., & Fox, R.A. (1993). Auditory and categorical affects on cross-language vowel perception. *Journal of the Acoustical Society of America*, *95,* 3623-3641.

Hardison, D. M. (2003). Acquisition of second language speech: Effects of visual cues, context, and talker variability. *Applied Psycholinguistics, 24,* 495-522.

Hardison, D. M. (2005). Second language spoken word identification: Effects of training, visual cues, and phonetic environment. *Applied Psycholinguistics*, *26,* 579-596.

Hirata, Y. & Kelly, S.D. (2010). Effects of lips and hands on auditory learning of second-language speech sounds. *Journal of Speech, Language, and Hearing Research*, *53*, 298-310.

Levis, J. & Pickering, L. (2004). Teaching intonation in discourse using speech visualization technology. *System*, *32*, 505-524.

Nilsen, D. & Nilsen, A.P. (2002). *Pronunciation contrasts in English*. Prospect Heights, IL: Waveland Press, Inc.

Pisoni, D., Aslin, R., Perey, A., & Hennessey, B. (1982). Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology: Human Perception and Performance*, *8*(2), 297-314.

Rochet, B. (1995). Perception and production of second-language speech sounds by adults. In W. Strange (Ed.) *Speech perception and linguistic experience: Issues in cross-language research* (pp. 379-410). Baltimore, MD: York Press.

Rvachew, S. (1994). Speech perception training can facilitate sound production learning. *Journal of Speech and Hearing Research, 37*, 347-357.

Strange, W. & Dittmann, S. (1984). Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception and Psychophysics, 36*(2), 131-145.

Wang, X. & Munro, M. (2004). Computer-based training for learning English vowel contrasts. *System, 32*, 539-552.