Running Head: Foreign Language Learning

Audio-visual learning of foreign language vocabulary

Donna Hopper

Algoma University College

The importance of both visual and auditory stimuli in contributing to the comprehension of spoken language has been the subject of research for many years. The investigations have concluded that both contribute to speech processing. Studies show that visual information greatly increases the likelihood that ambiguous speech will be perceived accurately. The Manner-Place Hypothesis explains this relationship but has not yet been applied to foreign language learning. It then follows that visual stimulus, in the form of articulatory cues, could have positive effects on foreign language learning due to the unfamiliarity and ambiguous nature of foreign language phonemes. Research in this area could have significant implications in the areas of psycholinguistics as well as foreign language teaching styles.

McGurk and MacDonald (1976) reported phonetic fusion when participants were presented with conflicting information in the visual and auditory modalities. Phonetic fusion occurs when information from the two modalities is integrated to produce a new phoneme with qualities of both the seen and heard stimuli. This illusion is known as the McGurk Effect and supports the Manner-Place Hypothesis, which states that the manner in which something is spoken is

detected by ear and the place of articulation is detected by eye (MacDonald and McGurk, 1978). Information from the two modalities is then synthesised into an auditory representation that combines features from both the auditory and visual stimuli (MacDonald and McGurk, 1978).

Three groups of participants (preschoolers, school aged children and adults) were studied under two conditions; audio-visual and audio-only. A video tape of a female speaker repeating the syllables /ba/, /ga/, /ka/ and /pa/ was treated so the voice was contrary to her lip movements. Four combinations were created; (a) /ba/voice-/ga/lips, (b) /ga/voice-/ba/lips, (c) /pa/voice-/ka/lips and (d) /ka/voice-/pa/lips. Those in the audio condition sat with their back to the television screen; participants in both conditions were instructed to listen to and repeat what the model said.

The results exhibited phonetic fusion of the two incongruent modalities (e.g. /ba/voice-/ga/lips is reported to be /da/) in the audio-visual condition, although this occurred more often when a stop consonant (/ba/ or /pa/) was used as the auditory stimulus (McGurk and MacDonald, 1976). This illusion does not habituate over time even if the participant is aware of the existing illusion. Although

fusions occurred in substantial frequency in all three groups, some responses were dominated by a single modality; predominantly auditory for children and visual for adults. This finding may explain why visually impaired children seem to have no problem acquiring language (McGurk and MacDonald, 1976).

Although studies reporting phonetic fusion are the strongest indicators of the link between vision and hearing during speech perception, those using masked auditory signals also result in findings which support the Manner-Place Hypothesis. The masking of audio signals involves dubbing over the existing audio track of a video tape to make the speech ambiguous. Visual dominance is apparent as noise levels increase in the audio-visual condition, in that recognition is superior to that in the audio only condition. At the most intense noise levels, participants in the audio-visual condition were able to correctly identify approximately 50% of the test items, whereas those in the audio condition identified between 0% and 20% of test items (Erber, 1969). In masked conditions, vision strengthens the recognition of consonants and vowels by 72% and 44.5% respectively (O'Neill, 1954).

These findings indicate that the observer utilizes the

visual information provided to make up for what is lacking in auditory stimulus. Vision only and hearing only conditions tend to evoke comparable results. When the two conditions are placed in contention, the visual input takes precedence (Dodd, 1977). Similar results have been found in studies using phonetic decision tasks (Massaro and Cohen, 1983), which also utilize ambiguous speech patterns. It has been demonstrated that the more ambiguous the auditory stimulus becomes, the more the visual stimulus takes precedence in speech processing. This inquiry can be related to foreign language learning because of the diversity and unfamiliarity of language phonemes, which can be compared to ambiguous auditory signals.

The integration of audio-visual information takes place, of course, in the brain. The potential to affect speech processing in the auditory cortex due to visual input has been suggested in past research. In normal conversation, visual information is often detected before the speech is actually heard and processed; the two are then integrated and processed together in the left hemisphere where language is predominantly processed (Sams, Aulanko, Hamalainen, Hari, Lounasmaa, Lu and Simola, 1991).

Recency effects act the same for both auditory and

visual speech (De Gelder and Vroomen, 1992; Campbell and Dodd, 1980). This may be explained by the realization that both lip-read and heard speech share characteristics such as temporal order and phonological components (Campbell and Dodd, 1980) which are processed in similar fashions in the brain. Recall for lip-read material, however, was poorer than that of auditory material. This presents a problem if both types of stimuli use the same linguistic speech processes; why is recall better for auditory material than for visual material? It has been suggested that phonetic decoding of lip-read material requires greater effort which could interfere with cognitive rehearsal strategies (De Gelder and Vroomen, 1992).

Studies show that "lip-read stimuli with the highest similarity ratings are the same as those with the highest auditory ratings (B, C, D, P, T, V)... lip-read stimuli are recoded into auditory representations and these auditory components are the source of these effects" (Manning, 1991). Due to the articulatory similarity, specific syllable combinations (e.g. /da/, /ta/, /ga/, /ka/ and /na/) result in non-illusory perception of the auditory stimulus (MacDonald and McGurk, 1978); inconsistencies between the visual and auditory modalities do not result in phonetic

fusion when these phonemes are placed in contention

Green and Kuhl (1991) agree that both audio and visual stimuli are integrated during speech processing and offer two explanations.

> "Either there are visual analysis mechanisms that operate in parallel and interact with the auditory mechanisms, or the auditory and visual information for a particular dimension are integrated before phonetic analysis, perhaps by mapping them onto underlying dimensions that are amodal in nature" (Green and Kuhl, 1991).

Past research has demonstrated the relationship between vision and audition, which supports the Manner-Place Hypothesis. Because of the ambiguity of foreign language phonemes, this theory may be applied to foreign language learning. When considering the aforementioned research, a logical conclusion can be made that, although one modality may be better than the other, both signals together are better than either of them alone. Rather than competing with one another, vision and hearing compliment each other in speech perception. This realization could have great effects on language teaching methods.

REFERENCES

Campbell, R., Dodd, B. (1980). Hearing by eye. Quarterly Journal of Experimental Psychology, 32, 85-99.

De Gelder, B., Vroomen, J. (1992). Abstract versus modality-specific memory representations in processing auditory and visual speech. Memory and Cognition, 20, 533-538.

Dodd, B. (1977). The role of vision in the perception of speech. Perception, 6, 31-40.

Erber, N. P. (1969). Interaction of audition and vision in the recognition of oral speech stimuli. Journal of Speech and Hearing Research, 12, 423-425.

Green, K. P., Kuhl, P. K. (1991). Integral processing of visual place and auditory voicing information during phonetic perception. Journal of Experimental Psychology: Human Perception and Performance, 17, 278-288.

MacDonald, J. and McGurk, H. (1978). Visual influences on speech perception processes. Perception and Psychophysics, 24, 253-257.

Manning, S. K. (1991). Similarity ratings and confusability of lipread consonants compared with similarity ratings of auditory and orthographic stimuli. American Journal of Psychology, 104, 587-604.

Massaro, D. W., Cohen, M. M. (1983). Evaluation and integration of visual and auditory information in speech perception. Journal of Experimental Psychology: Human Perception and Performance, 9, 753-771.

McGurk, H. and MacDonald, J. (1976). Hearing lips and seeing voices. Nature, 264, 746-748.

O'Neill, J. J. (1954). Contributions of the visual components of oral symbols to speech comprehension. Journal of Speech and Hearing Disorders, 19, 429-439.

Sams, M., Aulanko, R., Hamalainen, M., Hari, R., Lounasmaa, O. V., Lu, S. and Simola, J. (1991). Seeing speech: visualinformation from lip movements modifies activity in the human auditory cortex. Neuroscience Letters, 127, 141-145.

Running Head: Foreign Language Learning

Audio-visual learning of foreign language vocabulary

Donna Hopper

Algoma University College

Final paper submitted to Jack Dunning and Dr. Tom Allaway as partial fulfilment of the Honours Psychology Thesis Program.

ABSTRACT

Previous research has shown that visual cues will aid in the perception of ambiguous speech. The Manner-Place Hypothesis states that articulatory information will greatly increase the likelihood that ambiguous speech will be comprehended accurately. Because of the ambiguity of foreign language sounds, the current study applies this theory to foreign language learning by exploring the relationship between vision and hearing. Thirty-five undergraduate students from Algoma University studied 15 Ojibway words under two conditions; audio-visual or audio-only. When the two conditions were compared on word memory and pronunciation, significance was not found.

Audio-visual learning of foreign language vocabulary

The importance of both visual and auditory stimuli in contributing to the comprehension of spoken language has been the subject of research for many years. The Manner-Place Hypothesis explains the relationship between vision and audition but has not yet been related to foreign language learning. Previous investigations have concluded that vision plays an important role in speech processing. It then follows that visual stimulus in the form of articulatory cues could have positive effects on foreign language learning due to the unfamiliarity and ambiguous nature of foreign language phonemes. Studies reporting phonetic fusion are the most powerful indicators of the interaction between vision and audition during spoken communication.

McGurk and MacDonald (1976) reported phonetic fusion when participants were presented with conflicting information in the visual and auditory modalities. When phonetic fusion occurs, the information from both modalities is integrated to produce a new phoneme which has qualities of both the heard and seen stimuli (e.g. /ba/voice-/ga/lips is reported to be /da/). This illusion does not habituate over time even if the participant is aware of the existing illusion. This indicates that integration of both the visual and auditory stimuli is unconscious and automatic.

The Manner-Place hypothesis accounts for this illusion;

in normal conversation, the manner of articulation of consonants is perceived by ear (whether something is a stop consonant or not) and place of articulation is detected by eye (the position of the tongue and lips during articulation). Information from the two modalities is then synthesised into an auditory representation that combines the features both the auditory and visual stimuli (MacDonald and McGurk, 1978).

The masking of audio signals using white noise or background dialogue is another method used to study the interaction of vision and audition during spoken communication. At the most intense noise levels, participants in audio-visual conditions were able to correctly identify approximately 50% of the test items, whereas those in the audio-only conditions identified between 0% and 20% of test items (Erber, 1969). In masked conditions, visual stimulus strengthens the recognition of consonants and vowels by 72% and 44.5% respectively (O'Neill, 1954). This indicates that the observer utilizes the visual information provided to make up for what is lacking in auditory stimulus. When vision and hearing are placed in contention, the visual input takes precedence (Dodd, 1977; Massaro and Cohen, 1983). These results can be related to foreign language learning due to the diversity and unfamiliarity of foreign language sounds which can be compared to ambiguous auditory signals.

There are many similarities between the strategies used to process visual and auditory stimuli which become apparent when studying brain activity as well as looking at recency effects. The potential to affect speech processing in the auditory cortex due to the input from visual information has been suggested in past research. In normal communication, visual information is often detected before the speech is actually heard and processed; the two are then integrated and processed together in the left hemisphere where language is predominantly processed (Sams, Aulanko, Hamalainen, Hari, Lounasmaa, Lu and Simola, 1991). De Gelder and Vroomen (1992) illustrated that recency effects act the same for both auditory and visual speech. This may be explained by the fact that both lip-read and heard speech share characteristics such as temporal order and phonological components (Campbell and Dodd, 1980). It has also been found that "lip-read stimuli with the highest similarity ratings are the same as those with the highest auditory ratings (B, C, D, P, T, V)... lip-read stimuli are recoded into auditory representations and these auditory components are the source of these effects" (Manning, 1991). Due to the articulatory similarity, specific syllable combinations (e.g. /da/, /ta/, /ga/, /ka/ and /na/) result in non-illusory perception of the auditory stimulus (MacDonald and McGurk, 1978); inconsistencies between the auditory and visual modalities do not result in phonetic fusion when

these phonemes are placed in contention.

The purpose of the current study was to illustrate the relationship between vision and hearing as suggested by the Manner-Place Hypothesis, and to apply this hypothesis to foreign language learning. Participants were not told of the expected results of the study, nor were they given any indication that more than one testing condition existed. It was expected that visual information in the form of articulatory cues would increase the speed and accuracy with which participants learned foreign language vocabulary.

Method

Participants

Thirty-five undergraduate students from Algoma University with no prior knowledge of the Ojibway language and with normal hearing and normal or corrected to normal vision were used. They received credit towards their Introductory Psychology course for participating.

A pretest was conducted in order to determine how many trials were required for a participant to learn approximately 50% of the test items.

Apparatus and Procedure

Participants studied 15 Ojibway words under one of two conditions, (a) audio-only, or (b) audio-visual. A video tape was made of a model who was fluent in the Ojibway language presenting the test words one at a time. Only the head and shoulders of the model were shown. The English

word was provided first followed by the Ojibway translation which was given twice with a four second pause between word presentations. Participants were seated in a classroom in front of a television screen one to four at a time. In the audio condition, the screen was covered with a black card so the speaker on the video tape was not seen.

Each participant was instructed to listen carefully to the speaker and was provided with a list of the English words to refer to. They were not allowed to write anything down or talk to one another during the trials. They were also asked not to discuss the experiment with anyone until after the study was completed. There were four trials in total.

Immediately following the forth trial, each participant was given a five minute distracter task of anagrams to eliminate the possibility of recency effects. The test was then administered to participants one at a time. Participants were provided with the same word list they saw during the trials and were instructed to verbally provide the experimenter with all the Ojibway words, of the 15 they studied, they could remember. All their responses were audio taped using a voice activated recording device. The experimenter also recorded on a test sheet which words responses were given for. When all the tests had been completed, participants were debriefed and the audio tapes were given to a rater who rated each participant's

pronunciation on a scale of 1 to 3, 1 being "very poor" and 3 being "excellent".

Results

A two by two GLM was performed comparing each participant's results on pronunciation and word memory with the condition in which they studied the test items. A GLM was conducted to account for the unequal sample sizes; 19 participants were in the audio-only condition and 16 were in the audio-visual condition. The mean for word memory in the audio-only condition was 4.84 and 5.63 for the audio-visual condition (see figure 1.). For pronunciation, the means for audio-only and audio-visual were 1.86 and 2.27 respectively (see figure 2.). Although there are slight group differences, these results are not statistically significant [$F(1,67)=0.41$, $p=>0.05$].

Discussion

This study was designed to show the relationship between vision and hearing, and how it may influence foreign language learning. Although the results of the current study were not statistically significant, there are indicators that, with larger and equal sample sizes, a wider population representation, and greater environmental control, visual information will increase the speed and accuracy with which a foreign language is learned. The current insignificant results may be due to the small sample sizes, unequal sample sizes or a combination of the two

factors.

Participant motivation may have also contributed to these findings. All participants who took part in the experiment were required to do so in order to complete their Introductory Psychology course. Although completion of the experiment was not mandatory for course credit, participants were encouraged to do so. It is possible that because this was a requirement rather than a choice, many of the participants were not interested in, or concerned about the outcome of the experiment.

Due to time restraints, trials had to be conducted with more than one participant present at a time. This may have caused distractions for some participants, affecting their concentration and reducing their ability to learn the test items. Ideally, an experiment of this nature should be conducted using one participant at a time in order to maintain control over the learning environment.

In order to fully demonstrate the relationship between vision and audition as outlined by the Manner-Place Hypothesis, future studies regarding foreign language learning should consider factors such as participant motivation, testing environment and population sample. With these factors accounted for, the Manner-Place Hypothesis may be applied to foreign language learning.

References

Campbell, R., Dodd, B. (1980). Hearing by eye. Quarterly Journal of Experimental Psychology, 32, 85-99.

De Gelder, B., Vroomen, J. (1992). Abstract versus modality-specific memory representations in processing auditory and visual speech. Memory and Cognition, 20, 533-538.

Dodd, B. (1977). The role of vision in the perception of speech. Perception, 6, 31-40.

Erber, N. P. (1969). Interaction of audition and vision in the recognition of oral speech stimuli. Journal of Speech and Hearing Research, 12, 423-425.

MacDonald, J. and McGurk, H. (1978). Visual influences on speech perception processes. Perception and Psychophysics, 24, 253-257.

Manning, S. K. (1991). Similarity ratings and confusability of lipread consonants compared with similarity ratings of auditory and orthographic stimuli. American Journal of Psychology, 104, 587-604.

Massaro, D. W., Cohen, M. M. (1983). Evaluation and integration of visual and auditory information in speech perception. Journal of Experimental Psychology: Human Perception and Performance, 9, 753-771.

McGurk, H. and MacDonald, J. (1976). Hearing lips and seeing voices. Nature, 264, 746-748.

O'Neill, J. J. (1954). Contributions of the visual

components of oral symbols to speech comprehension. Journal of Speech and Hearing Disorders, 19, 429-439.

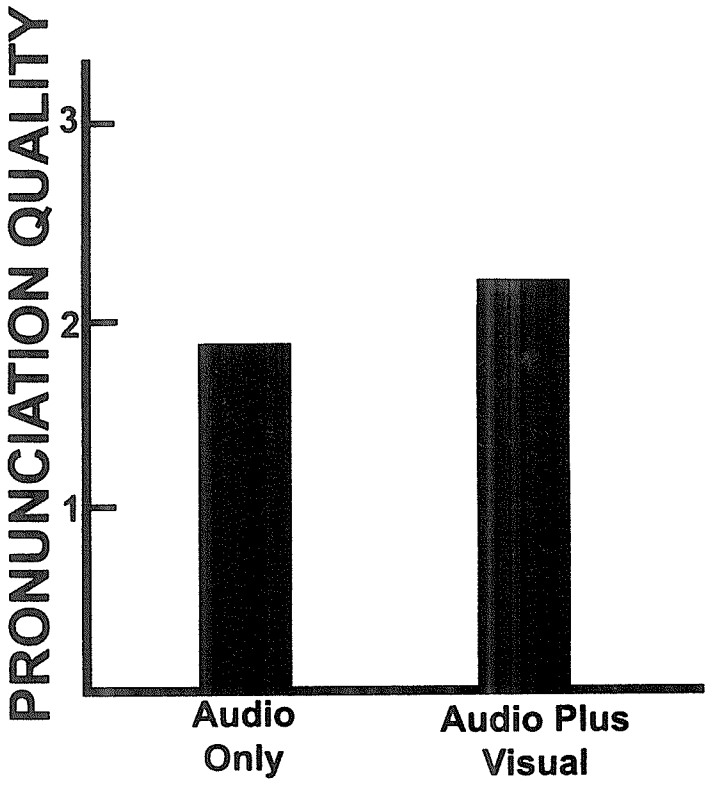
Sams, M., Aulanko, R., Hamalainen, M., Hari, R., Lounasmaa, O. V., Lu, S. and Simola, J. (1991). Seeing speech: visual information from lip movements modifies activity in the human auditory cortex. Neuroscience Letters, 127, 141-145.

Figure Captions

Figure 1. The mean number of words remembered by each group out of a possible 15.

Figure 2. The mean score on pronunciation for each group out of a possible 3.

PRONUNCIATION



WORD MEMORY