RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 27 (2005)
Indiana University

Lip-reading Skills in Bilinguals: Some Effects of L1 on Visual-only Language Identification[1]

Rebecca E. Ronquest and Luis Hernandez

Speech Research Laboratory
Department of Psychological and Brain Sciences
Indiana University
Bloomington, Indiana 47405

[1] This work was supported by NIH T32 Training Grant DC00012 to Indiana University. We would also like to thank David Pisoni and Manuel Diaz-Campos for all of their helpful comments and suggestions.

Lip-reading Skills in Bilinguals: Some effects of L1 on Visual-only Language Identification

Abstract. This study investigated whether observers can identify what language was being spoken in visual-only speech stimuli, and whether or not this ability depends on an observers' prior linguistic experience. Participants watched visual-only speech stimuli and were asked to decide if the talker in the video was speaking English or Spanish. Four groups of participants were studied: monolinguals and bilinguals, who were either native speakers of English or native speakers of Spanish. Results revealed that all subjects were able to identify the language being spoken with about 80% overall accuracy, regardless of their language background. However, the groups of participants differed in terms of their response bias. The L1 English bilinguals were strongly biased toward their native language, whereas the other three groups of participants did not demonstrate a significant bias toward the L1. The results of this experiment and their implications are discussed, and several directions for future research are considered.

Introduction

A large body of research has demonstrated that speech perception is multimodal; that is, the auditory, visual, and tactile properties of speech carry important information that can affect the intelligibility of the speech signal. It is also well established that the visual properties of speech are robust. The pioneering study carried out by Sumby and Pollack (1954) showed that the visual properties of speech carry important information about the linguistic content of the signal. In their study, they found that as the signal-to-noise ratio decreased, the contribution of the visual aspect of speech (i.e. the face of the talker) increased. In other words, when auditory aspects of speech are insufficient to communicate the message, visual information, such as the movements of the talker's face, is often relied upon to fill in the gaps. Both normal-hearing and hearing-impaired listeners take advantage of visual information when perceiving speech.

The findings of Sumby and Pollack have been strengthened by other findings reported by McGurk and MacDonald (1976). They found that when presented with mismatched auditory and visual information, the information carried in those modalities will often become fused together, and the observer will perceive a completely different sound than the one that was presented in either modality. The best example of the McGurk effect occurs when an observer is presented with an auditory /ba/ and a visual /ga/. The perceiver often reports an intermediate version of the two syllables, namely /da/. Thus, the information carried in the visual (gestural) aspects of the signal is great enough to, in a sense, override certain aspects of the auditory signal.

More recently, studies in the field of second language acquisition have shown that the inclusion of visual information, along with auditory information, aids in the acquisition of non-native contrasts. For example, Hazan et al. (2001) reported that visual information facilitates perception of sounds that are constrastive in the L2, but do not contrast in the L1. Another study by Hardison (2003) concluded that facial gestures aid in the perception of L2 targets in difficult phonetic environments and that visual cues to speech can be a source of reliable information for L2 learners.

All of the studies mentioned above suggest that the visual aspects of speech carry information that can contribute substantially to the intelligibility of the signal. However, the amount of information

carried by the visual signal, and whether observers use this information, remain important issues. The goal of the present investigation was to examine these issues by asking several groups of participants to perform a visual-only language identification task. In particular, we asked whether the visual properties of speech are robust enough to allow an observer to extract language-specific information from a visual display.

The present study examined the performance of several groups of native Spanish and English-speaking monolinguals and bilinguals in a visual-only language identification task. The subjects were presented with a series of video clips without sound and were simply asked to decide if the person in the video was speaking English or Spanish. A review of the published literature failed to uncover any other investigations that examined bilingual lip-reading ability. For this reason, it is difficult to make any specific predictions as to how the subjects will perform. Thus, one of the main questions this research addresses is whether the participants can carry out the task successfully. A second question is whether the participant's native language and prior language experience influence their performance in identifying English and Spanish from visual information.

Methods

Participants

A total of 56 participants took part in the present investigation (average age 24.9 years). The participants were from four language groups: Monolingual English speakers (N=16), Monolingual Spanish speakers (N=12), L1 Spanish bilinguals (N=12), and L1 English bilinguals (N=16). The monolingual English speakers were all undergraduate students at Indiana University who reported minimal or no knowledge of Spanish. The monolingual Spanish speakers were all current residents of Caracas, Venezuela, who reported that they did not speak or have knowledge of English.[2] The L1 Spanish bilinguals and L1 English bilinguals were all graduate students in the Department of Spanish and Portuguese at Indiana University. The participants in these two groups reported that they were proficient speakers of both English and Spanish. Age of L2 acquisition ranged from birth to 19 years of age. All participants received $10 for taking part in the study. Each section of the experiment is described in more detail below.

Stimulus Materials and Procedure

The present experiment consisted of three parts: a language history questionnaire, identification of CUNY sentences, and a visual-only language identification task. The stimuli were presented on an Apple Macintosh G4 computer. PsyScript version 5.1 was used for stimulus presentation. Subjects' responses were recorded with the keyboard for the CUNY task, and a button box for the language identification task. The entire experiment took approximately one hour to complete.

Language History Questionnaire. All participants completed a language history questionnaire. The purpose of the questionnaire was to gather demographic information pertaining to the language history of each participant such as the age of L2 acquisition and L2 usage. The monolingual Spanish participants completed a version of the questionnaire that was translated into Spanish. The other three groups of participants completed all paperwork in English.

[2] The data for the monolingual Spanish speakers was collected by Manuel Diaz-Campos in Caracas, Venezuela during July of 2005.

CUNY Sentences. Each participant took part in a CUNY[3] sentences task. The CUNY sentences were presented to each participant in auditory-only, audio-visual, and visual-only modalities. Twelve sentences were presented in each of the three modalities. The participants were asked to type what they thought they heard or saw on each trial. For the visual-only condition, they were told to do their best and guess if they were not able to determine exactly what the person in the video was saying.

Visual-only Language Identification Task. The experimental design of the V-only language identification task consisted of two blocks of 40 V-only video clips of short phrases in Spanish and English. Each block consisted of 20 English phrases and 20 Spanish phrases spoken by either a male or a female talker. One block consisted of 40 phrases presented by the male talker, and the other block consisted of 40 phrases presented by the female talker. The order of the blocks was counterbalanced so that half of the participants were presented with the male speaker first, whereas others saw the female speaker first. After seeing each video clip, the participants were asked to decide if the person in the video was speaking English or Spanish. A button box was used to record the subjects' responses. No feedback was provided. Only the data from the visual-only language identification task will be discussed in this paper.

Results

An initial examination of the data revealed that all subjects were able to successfully complete the visual-only language identification task at accuracy levels that were statistically above chance. The overall mean percent correct score for all subject groups was 78.06 %. A repeated measures ANOVA with stimulus language (English vs. Spanish) and stimulus gender (male vs. female) as within subject variables and participant group (Monolingual English, Monolingual Spanish, L1 English Bilingual, and L1 Spanish Bilingual) as between subject variables revealed a significant main effect for stimulus language ($F(1,49) = 4.107$; $p = .048$) and stimulus gender ($F(1,49) = 4.539$; $p = .038$). The participants performed significantly better on the English stimuli (79.18% English, 76.56% Spanish), and the stimuli spoken by the female talker (79.59 % female, 76.67% male).

The results also revealed a significant interaction between participant group (monolingual, bilingual, L1 English, and L1 Spanish) and stimulus language ($F(3,49) = 5.65$; $p = .002$). No other interactions were significant. Post-hoc paired samples t-tests indicated that, while both groups of monolinguals and the L1 Spanish bilinguals performed no differently on the stimuli presented in English and Spanish, the L1 English bilinguals performed significantly better overall on the English stimuli ($p = .001$). Figure 1 shows the percent correct scores for each group of participants for each presentation condition.

In addition to percent correct scores, non parametric measures of sensitivity (A') and bias (B") were calculated for all subject groups[4]. Both of these measures use the hit and false alarm rates to determine how sensitive the subjects are to the language differences in the signal, and to assess the extent to which they are biased toward one response option over another. A one-way ANOVA of A' score and subject group revealed no significant differences in sensitivity between participant groups. In other words, this indicates that the native language and language experience of the participants did not affect

[3] The group of monolingual Spanish participants did not complete the CUNY sentences because these sentences are in English. At present, there is not a set of CUNY sentences, or equivalent sentences, in Spanish.

[4] Formula for sensitivity (A')= 1/2 + ((P(Hits) - P(FA)) * (1 + P(Hits) - P(FA))/(4 * P(Hits)* (1 - P(FA))) ;
Formula for bias (B") =(P(Hits)*(1-P(Hits)) - P(FA)*(1-P(FA))) / (P(Hits)*(1-P(Hits)) + P(FA)*(1-P(FA)))

their sensitivity to differences in the signal. Figure 2 shows the mean A' scores for each of the four subject groups.

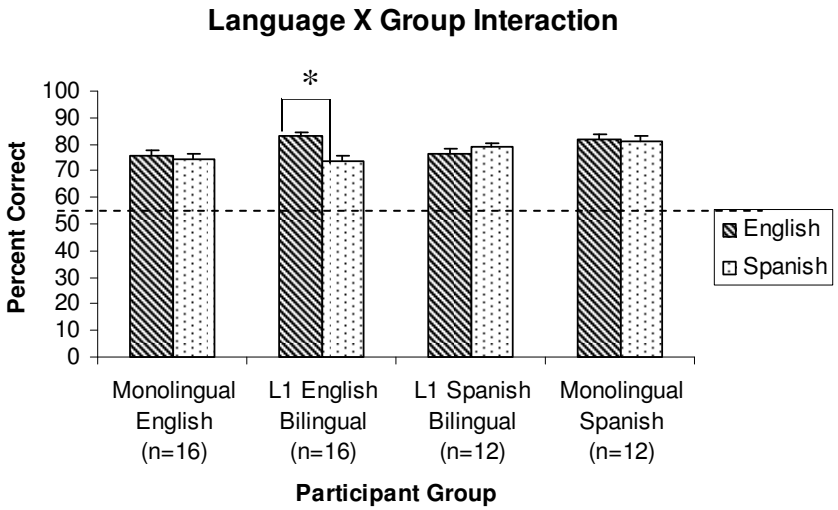

Figure 1. Percent correct scores for four subject groups. The dark bars indicate percent correct score on the English stimuli; the light colored bars represent percent correct scores on Spanish stimuli. Standard error bars are included. The dotted line represents scores significantly above chance using the binomial test.

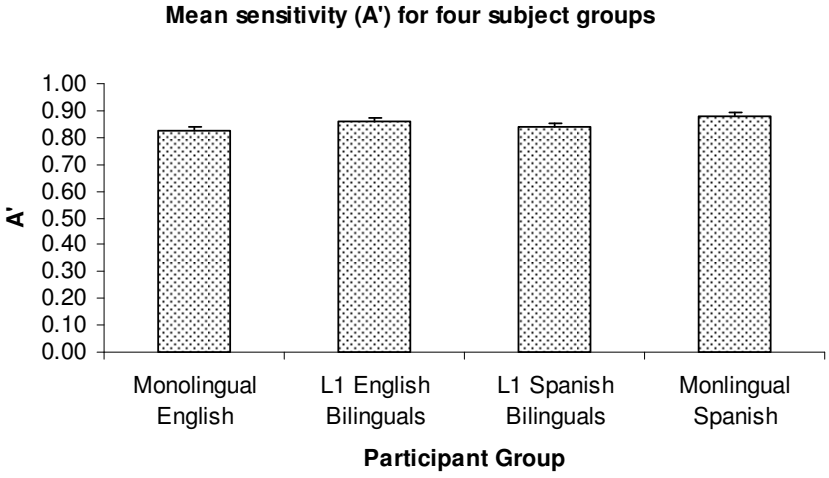

Figure 2. Mean sensitivity A' for all four subject groups.

A one-way ANOVA was also conducted in order to analyze the possible relationship between response bias (B") and participant groups. The results of this analysis revealed that the L1 English bilinguals had a response bias that was significantly different from the other three subject groups. While all participant groups showed at least some kind of response bias towards their native language, this bias was strongest for the group of L1 English bilinguals. This difference is shown in Figure 3.

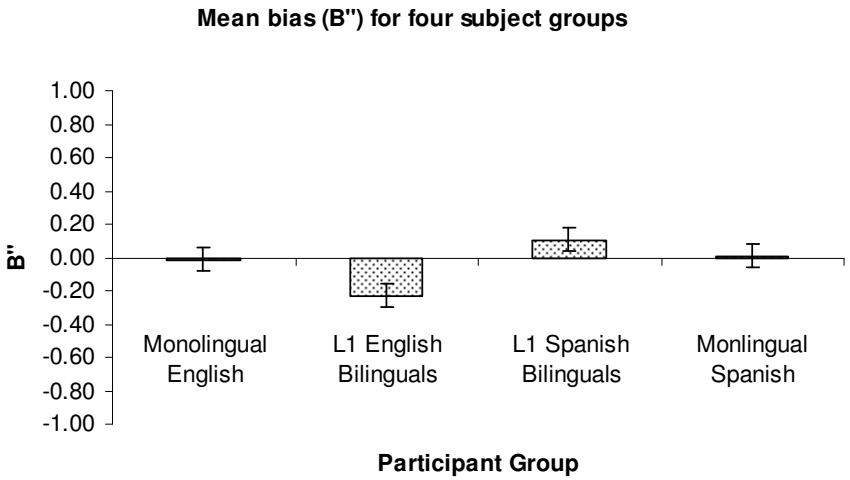

Figure 3. Mean bias (B") for all four subject groups. Negative values indicate a bias to respond more often as English; positive values indicate a bias to respond more as Spanish.

As shown in Figure 3, L1 English bilinguals displayed a strong response bias toward their L1, whereas the other three groups of participants did not appear to favor selecting their native language over their non-native language. The strong response bias toward the L1 explains why the L1 English bilinguals performed much better on the English stimuli.

Conclusions and Directions for Future Research

The results of this preliminary investigation provided some new insights into the robustness of the visual properties of speech. All participants were able to identify the language of a talker in a visual-only stimulus at levels well above chance. This result suggests that the visual signal provides enough information for an observer to correctly select the language being spoken from visual-only displays of speech. We also found that overall, the participants performed significantly better on the English stimuli than on the Spanish stimuli, and that performance was also better with the female talker.

One of the most interesting results was the interaction observed between language group and stimulus language. Our analysis revealed that native English-speaking bilinguals showed an increased level of performance when they were presented with English stimuli. However, the same effect was not found for the native Spanish-speaking bilinguals, or either group of monolinguals. Calculations of sensitivity (A') and bias (B") revealed that, although all four groups had comparable A' scores, the L1 English bilinguals displayed a strong response bias towards their native language. It is possible that the L1 English bilinguals were using a different strategy than the other three groups, which yielded a different result. It is interesting to note, however, that in the preliminary stages of this experiment, the native Spanish-speaking bilinguals showed a similar effect; the first eight subjects showed a higher

percentage correct score on the stimuli in Spanish. However, this effect was attenuated with the addition of more subjects.

One explanation for the lack of bias found in the L1 Spanish bilingual group could be that these participants were "set" in English mode, and thus failed to show the same kind of native language bias as the other group of bilinguals. All paperwork, instructions, and the CUNY sentences were presented to the L1 Spanish bilinguals in their non-native language, whereas the English monolinguals and L1 English bilinguals received instructions and task instructions in their native language. We are planning to present the L1 Spanish bilingual subjects with instructions and materials only in Spanish, as we did with the group of monolingual Spanish speakers from Caracas, Venezuela. Using the native language as the main mode of presentation may produce a native-language bias that is similar to that displayed by the L1 English bilinguals. In order to "set" the L1 Spanish bilinguals in Spanish mode, however, we will need to create a set of CUNY-like sentences in Spanish.

The present experiment measured participants' ability to identify the language of a talker using a visual-only phrase in English or Spanish. In future investigations we plan to examine participants' ability to identify English and Spanish using single words. The stimuli used in the present study varied in length, and on average, the Spanish stimuli were slightly longer and contained more syllables than the English stimuli. Thus, the participants may have used temporal cues to correctly identify the language being spoken. Isolated words, however, are much shorter in length, and may provide the participants with less useful duration information.

In another study we plan to reverse the video clips and present the information backwards in time. As previously mentioned, it is possible that the participants were able to make accurate language identifications based on utterance length or number of syllables. If duration or number of syllables were the major cues to language identity used by the subjects in this task, then temporally reversing the stimulus materials should not have any effect on overall performance. If, however, the participants were making their selections based on articulatory and gestural cues, temporal reversal should produce a decrease in performance on this task.

In conclusion, the present experiment has shown that participants are able to correctly identify the language of a talker when presented with a visual-only stimulus, suggesting that the visual properties of the speech signal are robust even in a language identification task. Future investigations will focus on identifying the particular visual cues that allow subjects to make such reliable judgments.

References

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

Hardison, D. (2005). Variability in bimodal spoken language processing by native and nonnative speakers of English: A closer look at effects of speech style. *Speech Communication, 46,* 73-93.

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

Hazan, V., Sennema, A., & Faulkner, A. (2001). Audiovisual perception in L2 learners. *Proceedings from the AVSP 2001 Conference on Auditory-Visual Speech Processing,*149-154.

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

Sumby, W.H. & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America, 26,* 212-215.