

RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 28 (2007)
Indiana University

Language Identification from Visual-Only Speech¹

Rebecca E. Ronquest, Susannah V. Levi, and David B. Pisoni

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

¹ This work was supported by grants from the National Institutes of Health to Indiana University (NIH-NIDCD T32 Training Grant DC-00012 and NIH-NIDCD Research Grant R01 DC-00111). We would also like to thank Luis Hernandez for assistance with programming, and Manuel Díaz-Campos, Althea Bauernschmidt, and Vidhi Sanghavi for their help in running subjects.

Language Identification from Visual-Only Speech

Abstract. The goal of the present investigation was to examine how observers identify English and Spanish from visual-only displays of speech. First, we replicated the recent findings of Soto-Faraco et al. (2007) with Spanish and English bilingual and monolingual observers using a different methodology. We found that prior linguistic experience affected response bias, but not sensitivity (Experiment 1). Additional experiments investigated the cues that observers used to carry out the language identification task. Participants were able to reliably identify languages when video clips were temporally-reversed, suggesting that prosody provides cues to language identity (Experiment 2). The contribution of lexical information to language identification was also investigated in Experiment 3. Participants' ability to identify stimulus direction (i.e., forwards vs. backwards) confirmed their sensitivity to differences in naturalness (Experiment 4). Taken together, the results of these four experiments indicate that prior linguistic experience, prosody, and perceived naturalness influence visual-only language identification

Introduction

A large body of research has demonstrated that speech perception is multimodal in nature. In addition to the auditory properties of speech, the visual signal carries important information about the phonetic structure of the message that affect the perception of the speech signal (c.f. Summerfield, 1987; Massaro, 1987). The visual aspects of speech have been shown to enhance or alter the perception of the auditory speech signal not only for listeners with hearing impairment, but for normal-hearing listeners as well (c.f., Campbell & Dodd, 1980; Summerfield, 1987; Lachs, 1999; Lachs, Weiss, & Pisoni, 2002; Kaiser, Kirk, Lachs, & Pisoni, 2003). In their seminal study of audio-visual speech perception, Sumbly and Pollack (1954) demonstrated that the visual properties of speech carry important information about the linguistic content of the signal. They found that including the visual signal along with the auditory signal allowed listeners to better understand speech at less favorable signal-to-noise ratios. When the auditory signal became more degraded, the visual aspects of speech were more important, and increased the intelligibility of the speech signal.

The contribution of visual information to speech perception is also illustrated by the McGurk Effect, in which visual information alters the perception of the speech signal. McGurk and MacDonald (1976) found that when observers were presented with mismatched auditory and visual information, they perceived a sound that was not present in either sensory modality. For example, a visual velar stop /g/ paired with an auditory bilabial stop /b/ was perceived as /d/. Thus, the information carried by the visual signal not only enhances speech perception, as found by Sumbly and Pollack (1954), but can override and alter the perception of auditory information, yielding a novel percept, as in the McGurk effect.

More recently, studies in the field of L2 acquisition have shown that the inclusion of visual information, along with the auditory signal aids in the acquisition of non-native contrasts. Hardison (2003) examined the acquisition of the English /l/-/ɹ/ contrast by native Japanese and Korean speakers. Participants were trained to identify these sounds under either auditory-only or auditory-visual presentation conditions. Learners who were trained in the auditory-visual condition showed better identification of /l/ and /ɹ/ in the post-test than those participants who were trained in auditory-only conditions. Hardison (2003) concluded that facial gestures enhance the discrimination of L2 targets in

difficult phonetic environments, and that visual cues to speech can be an additional source of information for L2 learners.

Similar studies have found that the contribution of visual information to speech perception, and the manner in which it is utilized, is also affected by an observer's native language and past experience with a second language. Hazan, Sennema, & Faulkner (2002) reported that visual information can facilitate L2 learners' perception of sounds that are contrastive in the L2, but do not contrast in the native language. For example, English contrasts the bilabial stop /b/ with the labiodental fricative /v/, whereas Spanish does not contain the latter phoneme. Hazan et al. (2002) found that Spanish learners of English who could perceive the contrast in the auditory-only condition also perceived the difference in the visual-only condition. In contrast, learners at early stages of acquisition who demonstrated higher rates of confusion between /b/ and /v/ auditorily did not benefit from the addition of the visual presentation. Hazan et al. (2002) concluded that learners at later stages of acquisition are sensitive to both the acoustic and visual cues associated with the non-native /b-/v/ contrast, whereas less experienced learners do not gain any significant benefits from visual cues until the contrast has been acquired auditorily.

In a related study, Werker, Frost, and McGurk (1992) found that the percentage of "visual-capture" (i.e., when the visual signal overrides the auditory signal) responses in a McGurk-type task was affected by the participants' native language and L2 experience. L1 and L2 speakers of French and English were presented with an auditory-visual stimulus that consisted of conflicting auditory and visual information; auditory /ba/ was paired with visual /ba, va, ða, da, ʒa, and ga/. Werker and colleagues found that beginning and intermediate L2 learners of English demonstrated significantly less visual capture of the interdental place of articulation /ð/ than did more proficient speakers of English. The beginning and intermediate learners of English generally reported "hearing" /ta/ or /da/, thus assimilating the interdental place of articulation with the closest French phoneme (/t/ or /d/). In contrast, the native English speakers, bilinguals, and advanced English learners were more influenced by the visual stimulus, and demonstrated a higher percentage of /ða/ responses. Werker et al. (1992) concluded that the ability to lip-read in a language is highly dependent upon experience with that language.

The studies reviewed above indicate that the visual information carried in the speech signal contributes substantially to speech intelligibility and that linguistic experience affects the manner in which the visual information is processed. Although previous research on visual speech perception and speech-reading has focused primarily on examining participants' ability to identify specific segments or words in a particular language, whether languages can be discriminated or identified based on the information in the visual signal alone has not been directly examined until recently. Two recent studies by Soto-Faraco and colleagues (2007) and Weikum and colleagues (2007) investigated visual-only language discrimination in both adult and infant observers, respectively. Soto-Faraco et al. assessed the ability of monolingual and bilingual observers to discriminate Spanish and Catalan from visual-only displays of speech. Two groups of bilinguals (Spanish dominant, Catalan dominant) and three groups of monolinguals (Spanish, Italian, and English) took part in the task. Bilingual participants exhibited higher rates of discrimination than monolingual Spanish speakers. The English and Italian monolingual speakers were not successful at the task, suggesting that knowledge of at least one of the languages is necessary for visual-only discrimination. Soto-Faraco et al. concluded that prior experience with the specific languages is one of the primary factors contributing to successful discrimination. They suggested that a number of different aspects of the stimuli facilitated discrimination, such as the length of the utterance, and the number of distinctive segments or words present in the stimulus. A similar study with infants showed that 4 to 6 month olds can discriminate between French and English in visual-only displays, but that by 8 months, this ability is limited to bilingual infants (Weikum et al., 2007).

Soto-Faraco et al. suggested that future investigations should examine observers' ability to discriminate or identify languages that are less closely related than Spanish and Catalan. In the present study, we sought to corroborate Soto-Faraco et al.'s earlier findings with a pair of languages that differ in prosody using a different task. Spanish and English were chosen in this study because they differ in terms of prosody, or rhythmic structure (e.g., Pike, 1946; Grabe & Low, 2002); Spanish is considered a syllable-timed language, whereas English is considered a stress-timed language. Syllable-timed languages exhibit more even spacing of syllables in an utterance (Pike, 1946), measured by variability of vowel durations (Grabe & Low, 2002). Thus, the duration of vowels is more regular for syllable-timed languages. In contrast, successive vowel durations in stress-timed languages are more variable. For example, English exhibits extensive vowel reduction and shortened duration of unstressed vowels. In terms of visual correlates of speech, the vocalic gestures (i.e., vocal aperture) in Spanish are more regular, while the gestures in English are more varied. Thus, differences in the rhythmic properties of speech should be perceivable from visual information alone.

In Experiment 1, we replicated the initial findings reported by Soto-Faraco et al. with Spanish-English bilingual talkers and both monolingual and bilingual Spanish-English observers using a two-alternative forced-choice identification paradigm. Soto-Faraco et al. concluded that participants attend to a combination of lexical and segmental cues to discriminate languages in visual-only conditions, but they were unable to determine the exact properties that their participants relied on to discriminate the two languages used in their study. A second goal of the present investigation was to examine in more detail the specific types of cues that observers may use to identify a language from visual-only displays of speech. Experiments 2-4 manipulated several aspects of the visual signal to examine participants' reliance on prosodic cues and lexical information in visually-presented displays of speech.

The first experiment demonstrated that observers can reliably identify the language being spoken from a visual-only stimulus. Experiments 2A and 2B investigated the role of prosodic information in visual-only language identification. The third experiment examined whether participants used lexical information from visual-only displays of speech, by asking them to judge the lexicality of a stimulus. The fourth experiment assessed whether observers could reliably identify the direction (forwards or backwards) of video clips presented in both English and Spanish.

Experiment 1: Visual-only Language Identification

Methods

Stimulus Materials. The stimulus materials in Experiment 1 consisted of a series of visual-only video clips of 40 English and 40 Spanish sentences (see Appendix 1). One male and one female talker were recorded using Behringer B1 Studio Condenser microphone and a Panasonic AG-DVX100 video recorder. All recordings were made in a sound attenuated IAC booth in the Speech Research Laboratory at Indiana University. Both talkers were bilingual speakers of Spanish and English. The male talker was a native of Venezuela and the female talker was a native of Puerto Rico. Both talkers acquired English during early adolescence and had lived in the United States for at least 6 years at the time of recording.

Participants. Four groups of participants were recruited for Experiment 1: monolingual English speakers (N=16), monolingual Spanish speakers (N=12), English-dominant bilinguals (N=16), and Spanish-dominant bilinguals (N=12). The monolingual English observers were all undergraduate students at Indiana University who reported minimal knowledge of Spanish. The monolingual Spanish observers were all residents of Caracas, Venezuela, who reported that they did not speak or have knowledge of English. The Spanish-dominant bilinguals and English-dominant bilinguals were all

graduate students at Indiana University who reported that they were proficient speakers of both Spanish and English, and had some experience teaching college-level Spanish. Age of L2 acquisition for these bilinguals ranged from birth to 19 years of age. None of the participants reported a history of a speech or hearing disorder at the time of testing. All participants received \$10 for taking part in the study.

Procedure. The stimuli were presented to the bilingual and monolingual English-speaking participants on an Apple Macintosh G4 computer. The monolingual Spanish speakers completed the experiment on an Apple Macintosh iBook G3 notebook computer in Caracas, Venezuela. PsyScript version 5.1 was used for stimulus presentation. Participants' responses were recorded with a button box for the language identification task. The entire experiment took approximately one hour to complete.

The visual-only language identification task consisted of two blocks of 40 video clips of short meaningful sentences in Spanish and English (see Appendix A). Each block consisted of 20 English sentences and 20 Spanish sentences spoken by both the male and female talkers. The stimuli were blocked by talker gender and counterbalanced across participants. After seeing each video clip, participants were asked to decide if the person in the video was speaking English or Spanish. No feedback was provided.

Data Analysis. In a two alternative forced-choice (2AFC) identification task, percent correct scores are influenced by both sensitivity and bias. For this reason, non-parametric measures of sensitivity (A') and bias (B'') were calculated for each participant to obtain robust measures of performance (Grier, 1971). Both of these measures use the proportion of hits and false alarms to determine how sensitive the participants are to the differences in the signal and to quantify the extent to which they are biased toward one response alternative over another. In Experiments 1 and 2, a response of "English" to English stimuli was considered a "hit"; a response of "English" to Spanish stimuli was considered a "false alarm."

Sensitivity (A') is measured on a scale of 0.0-1.0, with 0 indicating no ability to discriminate differences in the signal and 1.0 indicating perfect discrimination. A value of 0.5 on the sensitivity scale indicates chance performance. Bias (B'') is measured on a scale of -1.0 to 1.0. In Experiments 1 and 2, negative bias scores denote a tendency to respond "English" when presented with a stimulus, and positive values indicate a tendency to respond "Spanish." A score of zero indicates no response bias.

Results

To determine if participants' sensitivity was above chance performance (above 0.5 on the sensitivity scale) a one-sample t-test was conducted. As shown in Figure 1, the sensitivity measures for all four groups of subjects were significantly above chance (monolingual English $t(15) = 17.72, p < .001$; English-dominant bilinguals $t(15) = 28.30, p < .001$; monolingual Spanish $t(11) = 20.93, p < .001$; Spanish-dominant bilinguals $t(11) = 9.03, p < .001$). Thus, all participants were able to reliably identify the visual stimulus materials as English or Spanish. A one-way ANOVA was conducted on the A' scores with participant group as a between-subjects factor. The results of this analysis were not significant, demonstrating that all four groups performed comparably, and that observers' ability to identify a stimulus as Spanish or English did not depend on their native language or prior language experience.

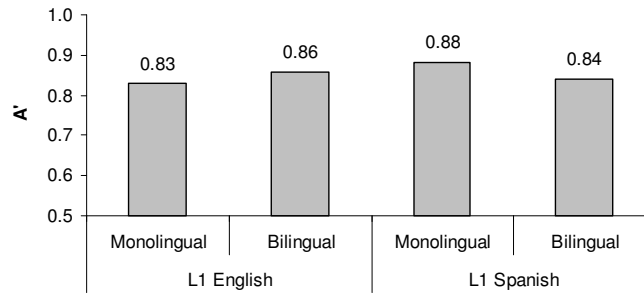


Figure 1. Mean sensitivity (A') for all four participant groups for Experiment 1.

The mean bias (B'') scores for all four participant groups are shown in Figure 2. A one-sample t-test of B'' scores showed that only the group of English-dominant bilinguals showed a response bias that differed significantly from 0.0 ($t(15) = -3.77, p = .002$); the English-dominant bilinguals had a strong tendency to choose the “English” response options, whereas the other three groups of participants did not demonstrate a significant bias. A one-way ANOVA was conducted on the B'' scores in to analyze differences between response bias and participant group. The main effect of participant group was significant ($F(3,52) = 5.95, p = .001$). Post-hoc Tukey tests revealed that the English-dominant bilinguals had a response bias that was significantly different from the other three participant groups (English-dominant bilinguals compared to English monolinguals $p = .03$; Spanish-dominant bilinguals $p = .001$; Spanish monolinguals $p = .03$). While all participant groups showed a tendency to respond with their native language, the bias was strongest for the group of English-dominant bilinguals.



Figure 2. Mean bias (B'') for all four participant groups in Experiment 1. Negative values indicate a bias to respond “English”; positive values indicate a bias to respond “Spanish”

Discussion

Regardless of language background or prior linguistic experience, all four groups of participants were able to complete the language identification task at levels that were significantly above chance. This result suggests that the visual speech signal alone provides sufficient information for an observer to correctly identify the language being spoken. That both monolingual and bilingual observers completed this task successfully replicates the earlier results of Soto-Faraco et al. (2007), who found that knowledge of only one of the test languages was sufficient to allow visual-only discrimination of Spanish and Catalan. The present results demonstrate that monolingual and bilingual participants not only can

discriminate between two different languages in visual-only displays of speech, but that they are able to accurately identify languages in a 2AFC task.

Unlike the results of Soto-Faraco et al. (2007), who found that bilingual observers were more successful in completing the discrimination task, we found no significant differences in sensitivity (A') between any of the four participant groups. Monolingual participants were just as sensitive as bilingual participants at identifying which language was spoken in the video clips. This result suggests that participants may have performed the 2AFC task by considering whether the stimulus was presented in their L1, or not in their L1, as opposed to making an English vs. Spanish judgment.

Measures of response bias (B'') revealed that all four participant groups exhibited some preference to respond with their native language. The bias was particularly strong in the group of English-dominant bilinguals. The monolingual participants showed less response bias than the bilingual participants, although this difference failed to reach statistical significance. Familiarity and naturalness may underlie the patterns of bias observed in Experiment 1. Monolingual English and Spanish speakers who possess knowledge of only one of the two test languages may have responded based on whether they recognized a familiar word or temporal pattern in their L1, reflecting the naturalness of the stimulus. When no familiar words or patterns were present in the video, or when the stimulus looked unnatural, these participants may have indicated that the language was their non-native language. In the case of the bilinguals, all of the video clips had the potential to contain familiar words, segments, or syllable structures, and thus they all appeared to be natural. The bilingual participants, upon finding some degree of familiarity or naturalness in the signal, may have processed the visual signal as belonging to the L1 because of L1 dominance.

The English-dominant bilinguals, who exhibited a significant bias to respond “English”, differed from the other three participant groups; the Spanish-dominant bilinguals failed to show a statistically significant native language bias, suggesting that they may have completed the task in an English mode, and adopted an English perceptual set. All paperwork and instructions were presented to the Spanish-dominant bilinguals in their non-native language (English), whereas the English monolinguals and English-dominant bilinguals received paperwork and task instructions in their native language. Using their non-native language as the primary mode of presentation may have attenuated the native language bias.

The results of Experiment 1 provide new insights into the robustness of the visual properties of speech. Several of the findings first reported in Soto-Faraco et al. were confirmed in the present study. They found that monolingual and bilingual observers could discriminate between Spanish and Catalan in visual-only displays of speech. Our results demonstrate that observers differing in language background and prior linguistic experience are able to identify languages based solely on the visual information. While Soto-Faraco et al. found that bilingual observers were better at completing a discrimination task, we found no significant differences in A' between monolingual and bilingual observers in our identification task. However, the effects of native language and prior linguistic experience were reflected in the differences in response bias (B'') in the present study.

Although we replicated the basic findings reported by Soto-Faraco et al. (2007), neither their study, nor Experiment 1 explained *how* participants carried out the visual-only language identification task. What cues do observers use to identify the language spoken in visual-only speech? The remaining experiments described below examine the contribution of stimulus length, rhythmic properties, and lexical information to visual-only language identification. Unlike Experiment 1 which analyzed differences in language identification between monolingual and bilingual speakers of English and

Spanish, only monolingual English speakers took part in the remaining three experiments. Monolingual English speakers were chosen for two reasons. First, the results of Experiment 1, as well as those of Soto-Faraco et al., suggest that knowledge of one language is sufficient for visual-only language identification and discrimination tasks. Second, the monolingual English speakers in Experiment 1 did not perform differently than the bilingual participants, and showed less response bias.

Experiment 2: Rhythmic Cues to Language Identification

The results of Experiment 1 demonstrated that observers can identify language from visual-only displays of speech. Experiment 2 was designed to assess the contribution of stimulus length and prosodic differences to visual-only language identification. The high level of accuracy obtained in Experiment 1 may have been due in part to the nature of the stimulus set, which consisted of sentence-length utterances. Participants viewed sentences of varying lengths, ranging from 2 to 12 words in both languages. Soto-Faraco et al. (2007) found that language discrimination was better in longer phrases than in shorter phrases. We predict that the same would be true for a visual-only language identification task. Longer utterances provide larger samples of speech and more opportunity for the observer to extract information necessary for accurate language identification. For this reason, both sentences and isolated words were used in Experiment 2 to test whether longer utterances would facilitate language identification. We were also interested in determining whether the limited information from words would provide sufficient information to permit reliable language identification.

In addition to manipulating stimulus length, we also manipulated the direction of the video clips. Temporally-reversed (“backwards”) versions of both the words and sentences were included in the stimulus set to assess whether participants can make accurate judgments about the language once lexical information has been eliminated. One possible way observers might extract language-specific information through visual speech is through rhythmic or prosodic information. Previous studies on visual-only speech perception have reported that observers are able to extract speaking-rate and stress differences from visual-only displays of speech (Green, 1987; Berstein, Eberhardt, Demorest, 1986). Thus, it is possible that observers in our experiments would be able to attend to rhythmic differences in the visual displays. As discussed earlier, Spanish is a syllable-timed language and English is a stress-timed language. Thus in Spanish, the vocalic gestures are more evenly-spaced in terms of duration while in English they are more variable. Temporal reversal of words and sentences preserves these global prosodic differences, but eliminates fine articulatory dynamics. That is, temporal reversal of the sentences and words creates stimuli which maintain overall temporal and rhythmic properties associated with Spanish and English, while at the same time eliminate the more fine-grained gestural-articulatory information necessary for lexical access. If participants use differences in the global rhythmic properties to identify language, we would expect that they should also be able to identify languages in the temporally-reversed stimuli, although they should be more accurate in the forwards condition where both lexical and rhythmic information are preserved. In contrast, if participants are unable to use prosodic cues, performance on the backwards stimuli should be extremely poor.

Experiment 2 examined both length and direction of visual-only stimuli. Manipulating the stimuli in this way allows us to investigate the potential contribution of rhythmic cues to visual-only language identification and to determine if single word utterances contain sufficient information for language identification. The experiment was divided into two parts. In Experiment 2A, participants were not informed that half of the video clips would be temporally-reversed. In Experiment 2B, the stimuli were blocked by direction, and participants were explicitly told that they would be viewing both forwards and backwards video clips.

Methods: Experiment 2A

Stimulus Materials. A total of 320 video clips were utilized in Experiment 2: 20 English and 20 Spanish sentences, and 20 English and 20 Spanish words, each spoken by two talkers, and presented in two directions (forwards and backwards). The 80 forwards sentences utilized in this experiment were the same as those used in Experiment 1 described above. The 80 word stimuli were recorded in the same way as the sentences described in Experiment 1. As with the sentences, each word was produced by the same male and female talker. The word stimuli included days of the week, animals, and the numerals one through ten (see Appendix 2). All video clips were temporally-reversed on an Apple Macintosh computer using Final Cut Pro, resulting in an additional 80 backwards sentences and 80 backwards words.

Participants. Thirty-four students enrolled in an introductory Psychology class at Indiana University participated in Experiment 2A. None of the participants who took part in Experiment 2A had completed Experiment 1. All were monolingual speakers of English who reported little or no knowledge of Spanish, and no history of a speech or hearing disorder at the time of testing. Participants received partial course credit for their participation.

Procedure. The general procedure for Experiment 2A was similar to the procedures used in Experiment 1. Participants were presented with two blocks of 160 stimuli. One block consisted of forwards and backwards words; the other block consisted of forwards and backwards sentences. The presentation of the blocks was counterbalanced across participants. After seeing each video clip, participants were asked to decide if the person in the video was speaking English or Spanish. A button box was used to record the participants' responses. The participants were not informed that half of the video clips in each block were time-reversed. No feedback was provided.

Results: Experiment 2A

As in Experiment 1, non-parametric measures of Sensitivity (A') and Bias (B'') were calculated for each participant. Mean values of A' and B'' are presented in Figures 3 and 4. A one-sample t-test of A' scores for the four conditions revealed that participants were sensitive to differences between the languages at levels statistically above chance (forwards sentences $t(33) = 16.84, p < .001$; backwards sentences $t(33) = 7.69, p < .001$; forwards words $t(33) = 7.96, p < .001$; backwards words $t(33) = 4.025, p < .001$). This finding indicates that participants were able to reliably identify the language from visual-only stimuli in all conditions. Moreover, the languages could be accurately identified when presented in the backwards condition. A repeated-measures ANOVA of A' scores with Stimulus Direction (forwards vs. backwards) and Length (words vs. sentences) as within-subjects variables revealed a significant main effect of Stimulus Direction ($F(1,33) = 4.42, p = .04$) and Length ($F(1,33) = 28.04, p < .001$). Participants identified the language as English or Spanish better when the stimuli were presented forwards ($A' = 0.73$) than backwards ($A' = 0.64$). The length of the stimuli also affected sensitivity. Participants were more accurate when presented with sentences ($A' = 0.71$) than with isolated words ($A' = 0.66$). The Direction by Length interaction approached significance ($F(1,33) = 3.81, p = .059$). Post-hoc analyses of this interaction revealed that participants were better able to identify the language being spoken in forwards sentences than in forwards words ($t(33) = -2.95, p = .006$). In the forwards conditions, observers' sensitivity was increased with increased length of the (words $A' = 0.70$, sentences $A' = 0.78$). In the backwards condition, however, longer utterances did not increase performance (words $A' = 0.62$, sentences $A' = 0.66$; $t(33) = -.942, p = .35$).

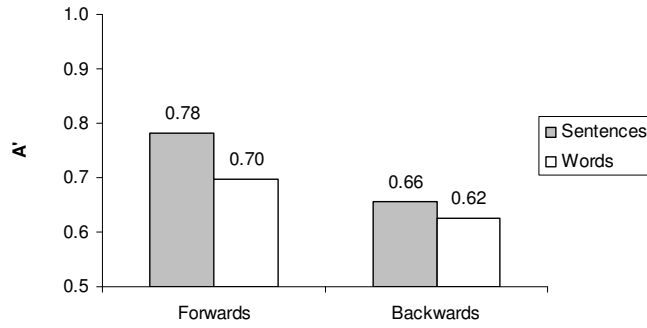


Figure 3. Mean sensitivity (A') in all four stimulus conditions for Experiment 2A.

Bias (B'') scores for each of the participants were also calculated. A repeated-measures ANOVA of B'' scores revealed a significant main effect of direction ($F(1,33)=22.03, p<.001$), indicating that participants were more biased to respond “English” for the forwards stimuli ($B'' = -0.09$), and “Spanish” to the backwards stimuli ($B'' = 0.05$). The main effect of Length was not significant. The Direction by Length interaction also reached significance ($F(1,33) = 8.44, p = .006$). Examination of this interaction revealed that participants displayed a greater bias to respond “English” when presented with forwards sentences than with forwards words (words $B'' = -0.03$, sentences $B'' = -0.15$; $t(33) = 2.29, p = .028$). In the backwards condition, although the overall trend was a greater bias towards Spanish, the B'' scores were not significantly different for words and sentences (words $B'' = 0.04$, sentences $B'' = 0.07$; $t(33) = -1.21, p = 0.23$).

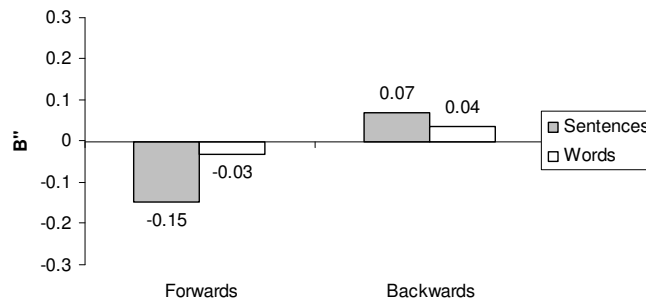


Figure 4. Mean bias (B'') in all four stimulus conditions for experiment 2A. Negative values indicate a bias to respond “English”; positive values indicate a bias to respond “Spanish”.

Methods: Experiment 2B

A modified version of Experiment 2A was conducted to examine the effects of direction when participants were explicitly told that some of the stimuli had been temporally-reversed. Because temporal reversal of the stimuli eliminated fine articulatory details and lexical cues, we hypothesized that awareness of the direction of the stimuli would force participants to rely on the prosodic information present in the video clips.

Stimulus Materials. The stimulus materials used in Experiment 2B were the same as those used in Experiment 2A.

Participants. A total of 33 introductory Psychology students took part in this experiment. A total of 13 participants were eliminated: three participants were eliminated because they had studied Spanish; one because of native Spanish speaking parents; one had undergone speech therapy; four due to computer malfunction; an additional four participants were eliminated so that the number of participants in each block order condition was equivalent. The remaining 20 participants were monolingual speakers of English who reported little or no knowledge of Spanish, and no history of a speech or hearing disorder. Participants received partial course credit for their participation. None of the participants had taken part in the previous experiments.

Procedure. Four blocks of visual-only stimuli (forwards words, forwards sentences, backwards words, and backwards sentences) were presented to participants. Prior to the presentation of each stimulus block, participants were told whether the stimuli would be presented forwards or backwards, and whether they would be viewing single words or whole sentences. Participants were divided into four groups based on the order of block presentation: 1) forwards words, forwards sentences, backwards words, backwards sentences, 2) forwards sentences, forwards words, backwards sentences, backwards words, 3) backwards words, backwards sentences, forwards words, forwards sentences, and 4) backwards sentences, backwards words, forwards sentences, forwards words. After viewing each video clip, participants were asked to decide if the person in the video was speaking English or Spanish. As in Experiment 2A, each block consisted of an equal number of English and Spanish tokens spoken by both the male and female talkers. No feedback was provided.

Results: Experiment 2B

The same statistical analyses carried out on the data from Experiment 2A were performed on the data collected in Experiment 2B. A summary of the A' scores is shown in Figure 5. A one-sample t-test of sensitivity (A') scores revealed that, as in Experiment 2A, participants could identify language at levels above chance (forwards sentences $t(19) = 7.75, p < .001$, backwards sentences $t(19) = 6.52, p < .001$; forwards words $t(19) = 5.73, p < .001$; backwards words $t(19) = 2.11, p = .04$). A repeated-measures ANOVA with Stimulus Direction (forwards vs. backwards) and Length (word vs. sentence) as within-subjects variables revealed a significant main effect of Direction ($F(1,19) = 10.95, p = .004$) and Length ($F(1,19) = 7.23, p = .01$). As observed in Experiment 2A, participants were more sensitive to language differences when the stimuli were presented forwards ($A' = 0.68$) than backwards ($A' = 0.60$), and were also more accurate with sentences ($A' = 0.67$) than words ($A' = 0.61$). The Direction by Length interaction was also significant ($F(1, 19) = 12.62, p = .002$). Post-hoc paired samples t-tests on this interaction revealed that accuracy was affected by length in the backwards condition (words $A' = 0.56$, sentences $A' = 0.64$; $t(19) = -3.65, p = .002$). In contrast to the results of Experiment 2A, no difference in length was found in the forwards direction (words $A' = 0.67$, sentences $A' = 0.69$; $t(19) = -1.33, p = .19$).

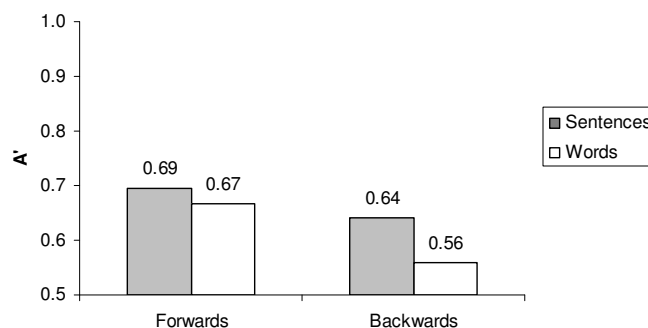


Figure 5. Mean sensitivity (A') in all four stimulus conditions for Experiment 2B.

Measures of response bias (B'') were also calculated. A summary is presented in Figure 6. A repeated-measures ANOVA with Stimulus Direction (forwards vs. backwards) and Length (word vs. sentence) as within-subjects variables revealed a significant main effect of Direction ($F(1,19) = 7.66; p = .012$). Participants were more likely to respond “English” in the forwards condition than in the backwards condition. The general pattern of response bias is similar to that observed in Experiment 2A, but the magnitude of bias was attenuated. The main effect of stimulus Length and the Direction by Length interaction were not significant.

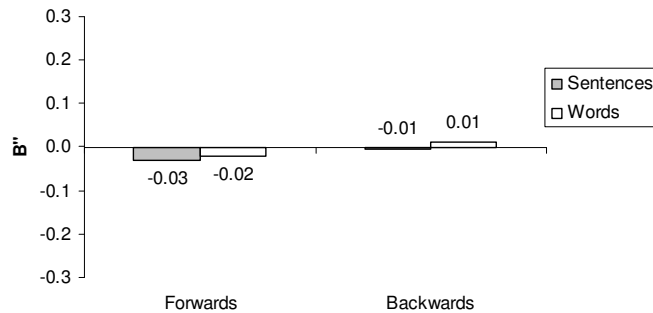


Figure 6. Mean bias (B'') in all stimulus conditions for experiment 2B. Negative values indicate a bias to respond “English”; positive values indicate a bias to respond “Spanish”.

Discussion: Experiments 2A and 2B

Experiments 2A and 2B were designed to examine the contribution of rhythmic information to visual-only language identification. As previously mentioned, global rhythmic differences between English and Spanish are retained in temporally-reversed stimuli. The results of these two experiments demonstrate that observers can identify differences in rhythmic structure from visual-only stimuli, and that they use this information in a language identification task. Participants’ ability to reliably identify the language from the backwards stimuli suggests that even when access to lexical information is eliminated, a sufficient amount of prosodic information is still available to facilitate identification. Moreover, the present results demonstrate that monolingual speakers of English are able to make reliable judgments about language identity based on the visual information alone in the backwards stimuli. The results of this experiment suggest that observers perceive and utilize prosodic information associated with English and Spanish. We conclude that the rhythmic properties of a language are one cue that participants use to determine language identity from visual-only displays of speech.

Sensitivity to the language differences in the signal was greater when the stimuli were presented in the forwards condition as compared to the backwards condition. Greater sensitivity in the forwards condition was attained because forwards stimuli contain all possible cues to language identification; that is, forwards stimuli contain both rhythmic and lexical information, whereas only rhythmic cues are retained in backwards stimuli. The finding that sensitivity to language differences in the forwards condition was greater also suggests that participants use other sources of information in addition to rhythmic information to make their decisions about language identity. If rhythm and timing were the only properties observers attended to, then performance on the forwards and backwards stimuli would have been equivalent. In addition, stimulus length was also found to influence performance; sentence-length stimuli provided more information to language identity than isolated words. The sentence-length utterances contained more information than the words, and also provided participants with more time to make their decisions.

In addition to differences in sensitivity, response bias was also affected by the stimulus condition. Participants showed a greater tendency to respond “English” when they were presented with forwards stimuli and “Spanish” when presented with backwards stimuli. The differences in response bias suggest that in the backwards condition, when a word or a sentence appeared to be less natural and less familiar, or did not contain any recognizable information, participants were more likely to respond that the stimulus was Spanish.

Participants in Experiment 2A were not told that stimuli would be presented to them in two directions. When presented with stimuli, observers may have been making their decisions based on whether the stimulus display appeared natural or familiar. In the backwards condition, stimuli appeared less natural and less familiar, influencing the observers to identify these stimuli more often as Spanish; the forwards stimuli, because they were more familiar and natural, were more likely to be judged as English. In Experiment 2B, participants were explicitly told whether the stimuli were temporally-reversed. Thus, this group of participants was aware that they could no longer rely strictly on naturalness or familiarity to make their decisions, because half of the stimuli would appear unnatural; they also were aware that they would not be able to access lexical information in half of the stimuli. Participants’ knowledge of stimulus direction altered their strategy in this task, and resulted in smaller response bias.

We also found that response bias towards English or Spanish was slightly greater with sentences than with words, although this difference was not statistically significant in all conditions. In the forwards condition, response bias to English was greater with the sentences than with words; observers were slightly more biased to respond “Spanish” when presented with a backwards sentence than with a backwards word. Participants may have exhibited stronger biases when presented with longer utterances because the additional length provided more cues to naturalness. Longer utterances also provided more information about gestures and articulation, which afforded participants more opportunity to decide if the stimulus looked natural or familiar. Sentence-length utterances offered more articulatory and timing information than word-length utterances.

The rhythmic properties of a language, which were maintained in the temporally-reversed versions of the stimuli, provided sufficient cues to language identity. Thus, it is not necessary for lexical information to be present for reliable language identification to occur. In the forwards condition, however, when both the rhythmic and lexical properties of the language were present, overall performance was enhanced. Greater sensitivity to the linguistic differences in the forwards stimuli suggests that a combination of rhythmic cues and lexical information is more beneficial than having only one available set of cues.

Experiment 3: Lexicality Judgments

Greater accuracy in the forwards condition in Experiments 2A and 2B suggests that participants attended to other properties of the stimulus, in addition to rhythm, when completing the language identification task. We hypothesize that observers extract both rhythmic cues and lexical information when making their decisions. Research on lip-reading has shown that both lexical and segmental information can be extracted from isolated words in the visual-only modality (Lachs et al., 2002; Kaiser et al. 2003). The purpose of Experiment 3 was to examine participants’ ability to extract lexical information from visual-only isolated words, using a lexical decision task.

If participants accessed and used lexical information to carry out the language identification tasks in our earlier experiments, they should be more likely to report that forwards English stimuli are “words”

than Spanish stimuli. We also expected participants to be more likely to indicate that backwards video clips were “nonwords” than forwards video clips.

Methods

Stimulus Materials. The stimulus materials used in Experiment 3 consisted of the same forwards and backwards words utilized in Experiments 2A and 2B.

Participants. The participants in Experiment 3 were 32 introductory Psychology students at Indiana University. All participants met the same specifications described for Experiments 2A and 2B above. Partial course credit was awarded to all those who participated in this experiment. None of the participants had taken part in any of the previous experiments.

Procedure. Participants were presented with a single block of 160 trials mixed by talker, language, and stimulus direction. In contrast to the previous three language-identification tasks, participants were instructed to decide if the talker was saying a “word” or a “nonword.” Participants were not informed that the words were spoken in English and Spanish, nor were they told that half of the video clips had been temporally-reversed. No feedback was provided.

Results

The number of “word” and “nonword” responses in each of the four conditions was calculated and these response frequencies were then analyzed using a Chi-square test of independence to determine if the distribution of responses was different across conditions. Collapsing over the direction of the stimuli, the distribution of “word” and “nonword” responses was significantly different for the English and Spanish stimuli ($\chi^2(1, N = 5106) = 32.425, p < .001$). This indicates that participants were more likely to categorize an English stimulus as a word than a Spanish stimulus (60% for English, and 52% for Spanish). The overall differences in frequency distribution reported for the total number of English and Spanish videos were also present when the stimuli were subdivided further. Chi-square analyses comparing forwards English and forwards Spanish words was significant ($\chi^2(1, N = 2552) = 39.507, p < .001$), indicating that there were more “word” responses to the forwards English stimuli (75%) than to the forwards Spanish stimuli (63%). Finally, the backwards Spanish stimuli were labeled as “words” less often than the backwards English stimuli (47% for English, and 42% for Spanish; $\chi^2(1, N = 2554) = 4.673, p < .05$).

Collapsing over language, the distribution of “word” and “nonword” responses for the forwards and backwards stimuli was also significant ($\chi^2(1, N = 5106) = 319.36, p < .001$). The participants categorized the forwards stimuli as “words” more often than the backwards stimuli (69 % for forwards video clips, and 44% for backwards videos). The overall pattern of responses found for direction was also observed within each language. Forwards English videos were labeled as words on 75% of the trials, whereas backwards English videos were labeled as words in only 46% of the trials. The chi-square analyses of this distribution was significant ($\chi^2(1, N = 2555) = 215.45, p < .001$). The distributions of the forwards and backwards Spanish stimuli was also significantly different ($\chi^2(1, N = 2551) = 114.14, p < .001$). Forwards Spanish stimuli were judged to be words more often than backwards Spanish stimuli (63% for forwards Spanish, and 41% for backwards Spanish).

In short, when observers were asked to make word/nonword judgments on isolated visual displays of English and Spanish words, they displayed a highly consistent pattern that differed statistically from chance expectation.

Discussion

The chi-square analyses indicated that observers' responses were not randomly distributed across the different stimulus conditions. "Word" and "nonword" responses varied systematically depending on the experimental conditions. Moreover, observers were more likely to judge an English stimulus as a word than a Spanish stimulus. The same preference for the "word" response was also observed with the forwards stimuli, regardless of the language of the stimulus. The forwards versions of both the English and Spanish stimuli were labeled as words more often than the backwards versions of the same stimuli.

The main goal of this experiment was to examine the extent to which participants access the lexicon when engaging in a visual-only language identification task. Although there is some evidence that lexical information may be accessed due to the higher frequency of "word" responses with the forwards English stimuli as opposed to the forwards Spanish stimuli, it is not possible to describe the extent to which lexical information contributes to visual-only language identification. Only monolingual speakers of English took part in this experiment, and it was thus assumed that these observers did not possess a Spanish lexicon. The fact that participants labeled approximately half of the Spanish stimuli as words suggests that they may have been making "word"/ "nonword" decisions based on whether the stimuli looked as if they could be possible words in English and not as a result of explicitly recognizing a stimulus as a specific lexical item. The forwards English stimuli were judged to be "words" the most frequently, followed by the Spanish words. In the backwards condition, the Spanish stimuli identified as "nonwords" more often than the English stimuli.

The pattern of responses observed in this experiment suggest that as in Experiment 2, the participants were attending to more global properties of the stimuli that are related to naturalness and familiarity, as opposed to making their decisions based on whether they recognized a specific word in their language. In the forwards English condition, the greatest number of cues to identity, both lexical and temporal information, is maintained in a coherent manner, and these stimuli should appear to the most natural-looking of all four stimulus types. The forwards Spanish stimuli are potentially recognizable as language, consisting of a combination of sounds and gestures that are also possible in English, but appear less recognizable than the English words. The backwards English and Spanish stimuli may maintain some of the rhythmic properties associated with each language, but lack the specific details necessary to identify a particular word.

Although the ability to recognize lexical information may contribute to more accurate language identification, the results of this experiment suggest that lexical properties of visual speech may not be as robust as the more global rhythmic and timing information. We conclude that observers may have been basing their decisions on whether the stimulus appeared as if it *could* be a word in English, or whether it looked highly unnatural and was therefore unlikely to be a possible word in English.

Experiment 4: Direction

The previous three experiments investigated participants' ability to identify language in visual-only stimuli and examined the extent to which they utilized prosodic and lexical information when making their decisions. In Experiment 3, participants may have made their word/nonword judgments based on the naturalness of the stimuli. That is, the forwards stimuli are considered natural, since they are actual language productions, whereas the backwards stimuli are unnatural. The goal of Experiment 4 was to investigate the question of articulatory naturalness by examining whether participants can reliably

identify the direction (forwards or backwards) of a silent video clip. We were also interested in determining if performance on this task would be affected by the language of the stimulus.

Methods

Stimulus Materials. The stimulus materials used in Experiment 4 consisted of the same set of video clips used in Experiments 2A and 2B: forwards and backwards visual-only video clips of English and Spanish words and sentences spoken by a male and female talker.

Participants. Twenty-five additional participants took part in this experiment. Three participants were eliminated due to computer malfunction, and two others for not following directions. Of the remaining 20 participants, 14 were introductory Psychology students who received course credit for taking part in this experiment. The other six participants were paid \$10 for participating. None of the participants had completed any of the previous experiments described in this paper.

Procedure. Each participant was presented with one block of 160 words and one block of 160 sentences that were mixed by talker and language, but separated by stimulus length. All participants were presented with the words block first, followed by the sentences block. After viewing each video clip, participants were instructed to decide if the video they had just seen was forwards or backwards. The participants were not told that half of the video clips were in English and that half were in Spanish. No feedback was provided.

Data Analysis. As in Experiments 1 and 2, sensitivity (A') and bias (B'') were the primary means of measuring performance on this task. In contrast to the previous experiments, however, participants were not asked to make language judgments, but instead were asked to identify direction. For this reason, in Experiment 4, a response of “forwards” to a forwards stimulus was considered a “hit”. A false alarm occurred when a participant incorrectly identified a backwards stimulus as being forwards. Negative B'' scores would thus indicate a tendency to respond “forwards,” whereas positive scores would be indicative of a bias to respond “backwards.”

Results

To examine observers' ability to identify the direction of each video clip, sensitivity (A') scores in the four stimulus conditions were calculated. A summary of these scores is presented in Figure 7. A one-sample t -test of A' scores for each condition was significant, indicating that participants were able to reliably discriminate between the forwards and backwards video clips (English sentences $t(20) = 6.23$, $p < .001$; English words $t(20) = 7.77$, $p < .001$; Spanish sentences $t(20) = 5.26$, $p < .001$; Spanish words $t(20) = 8.30$, $p < .001$). Thus, participants were able to reliably determine if the video clip they had just seen had been presented to them forwards or backwards. A repeated-measures ANOVA with Stimulus Language (English vs. Spanish) and Length (word vs. sentence) as within-subjects variables was conducted, and revealed a significant main effect of Length ($F(1,20) = 5.36$, $p = 0.03$). Observers were better able to identify a video clip as forwards or backwards when presented with an isolated word ($A' = 0.73$) than when presented with a sentence ($A' = 0.68$). Thus, in contrast to our earlier findings in Experiment 2, participants' ability to judge the direction of a stimulus was not enhanced when the video was longer in duration. The main effect of Stimulus Language and the Language by Length interaction were not significant. That the main effect of language was not significant indicates that participants were able to determine the direction of the video clip regardless of the language of presentation.

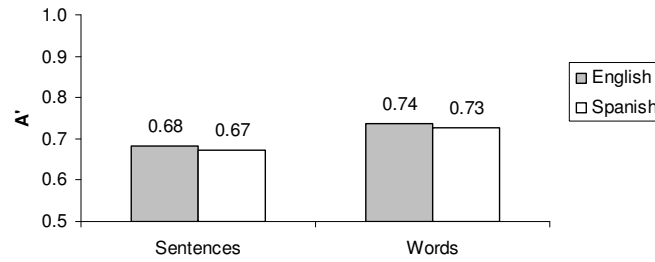


Figure 7. Mean sensitivity in four stimulus conditions in Experiment 4.

Mean bias (B'') scores for each condition are presented in Figure 8. As shown here, all B'' scores were negative, indicating a bias towards the “forwards” response alternative in all conditions. A repeated-measures ANOVA with Stimulus Language (English vs. Spanish) and Length (words vs. sentences) as within-subjects variables revealed a significant main effect of Stimulus Language ($F(1,20) = 8.36, p = .009$). This result indicates that participants had a greater tendency to respond “forwards” when presented with an English video ($B'' = -0.16$) than with a Spanish video ($B'' = -0.08$). The main effect of Length, and the Language by Length interaction did not reach significance.

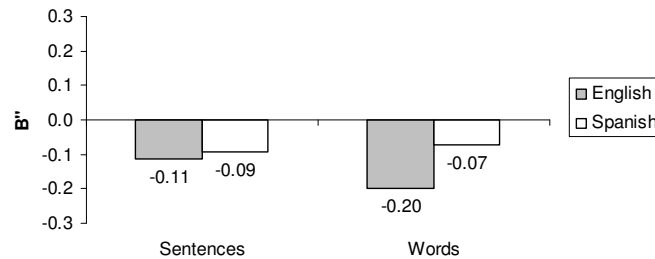


Figure 8. Response bias (B'') in all four stimulus conditions in Experiment 4. Negative values indicate a bias to respond “forwards,” whereas positive values indicate a tendency to respond “backwards.”

Discussion

Overall, the results of Experiment 4 show that observers can successfully identify the direction (forwards or backwards) of a visual-only video clip. This finding suggests that participants are able to reliably identify differences in naturalness in the visual-only modality. Analysis of sensitivity (A') revealed no significant effects of presentation language. Participants were able to reliably identify the direction of a video clip regardless of the language in which it was spoken. We conclude that natural productions of speech (i.e. forwards utterances in both languages) were more natural-looking to observers because they specify gestural properties that are identifiable as being possible in language.

An examination of response bias (B''), however, did reveal an effect of language; participants were biased to respond “forwards” across all conditions, but this bias was strongest when the language of

presentation was English. Differences in naturalness and familiarity may be able to explain this trend. Because the participants were all monolingual English speakers, English sentences and English words would appear to be the most natural utterances to observers, and would also be the most familiar. Experiment 2 showed that the temporally-reversed versions of the stimuli maintained some of the global prosodic characteristics associated with English, and that this information could be reliably perceived from silent video clips. Thus, all English stimuli, including those that were presented backwards, contained some familiar properties of the native language which may have influenced observers to respond “forwards.”

On the other hand, the Spanish stimuli would have appeared less natural and less familiar to participants. In the forwards condition, Spanish words and sentences may have looked as though they contained possible consonant and vowel gestures, but were less familiar in terms of their global prosodic characteristics. In the backwards condition, the Spanish words and sentences contained little, if any, familiar information to which the observers’ could attend. Thus, participants were less likely to respond “forwards” when presented with a Spanish stimulus because of their lack of experience and familiarity with the rhythmic properties of the language. With the exception of the backwards Spanish stimuli, the video clips contained information that was in some way familiar to participants, influencing them to respond “forwards” more often than “backwards.” The number of cues and the degree of familiarity and naturalness was greater for the English stimuli, reflecting why observers were more biased to identify these stimuli as being presented forwards than with the Spanish stimuli.

The second finding of this experiment is that participants exhibited greater ability to identify direction when the stimuli were word-length utterances than when they were sentences. This finding contrasts with the results obtained in experiments 2A and 2B, in which sensitivity to language differences was greater with longer utterances. It is possible that in this experiment, longer stimuli provided more opportunity for participants to believe that they had seen a familiar structure, resulting in greater confusion and lower accuracy with longer utterances.

General Discussion and Conclusions

In this paper we investigated how observers identify language from visual-only displays of speech. Our main goals were to replicate and extend the earlier findings of Soto-Faraco et al. (2007) using a different methodology. Overall, the results of Experiment 1 confirmed their earlier findings that language identification is possible from information in visual-only displays of speech. Although we found no differences in measures of sensitivity between monolingual and bilingual speakers of Spanish and English, the effect of prior linguistic experience was observed in measures of response bias; bilingual English speakers differed from all other participant groups, showing a bias for their native language.

A second goal of our investigation was to determine *how* observers identify languages when provided with visual-only information, by examining their reliance on prosodic information, lexicality, and naturalness. In Experiment 2, we directly examined the contribution of prosody to language identification by temporally reversing the video clips. We found that even when the visual stimuli were presented backwards, participants were still able to reliably identify stimuli as English or Spanish, although performance was significantly better in the forwards condition. We also found that observers were able to identify languages from short, isolated words, as well as sentences, but that sensitivity to language differences was greater in longer utterances. To examine if observers were accessing and using lexical information in the previous experiments, a lexical decision task was conducted in Experiment 3. The differential pattern of response frequencies suggested that observers may have been accessing some lexical information, but we concluded that “word/nonword” decisions were more likely influenced by the

perceived naturalness of the stimuli. Observers' attention to naturalness of the stimuli was investigated further in Experiment 4, in which participants were asked to decide if a video had been presented to them forwards or backwards. Participants were able to reliably identify isolated words and sentences as forwards or backwards, indicating that they were able to detect whether a stimulus looked like a natural language production. Although observers demonstrated a bias to respond "forwards" to all stimuli, the bias to respond "forwards" was stronger when the observers were presented with English stimuli than with Spanish stimuli. Based on the findings described above, we conclude that observers' prior linguistic experience influenced the way they performed the visual-only identification tasks, and that they were able to identify languages from visual-only stimuli using prosody and naturalness.

In their recent study, Soto-Faraco et al. (2007) found that linguistic experience affected observers' ability to discriminate languages when provided only with visual information. Bilingual speakers of Spanish and Catalan exhibited the highest discrimination scores, followed by the Spanish monolinguals. Monolingual speakers of English and Italian were unable to complete the discrimination task successfully, leading the authors to conclude that knowledge of at least one of the languages presented in the visual-only displays was necessary for reliable discrimination. Although observers in Experiment 1 did not exhibit sensitivity (A') differences, the effects of linguistic experience were revealed in differences in response bias. Analysis of response bias (B'') revealed significant differences between the English-dominant bilinguals and the other three groups of observers. English dominant bilinguals exhibited a strong response bias toward their native language, whereas response bias for the other three groups did not differ significantly from each other. Although all four groups of observers showed some tendency to respond more with their native language, bias was stronger with the bilinguals than the monolinguals.

The bias exhibited by the English-dominant bilinguals can be attributed both to linguistic experience and methodological factors. All stimulus materials contained sentences that were potentially recognizable to this group of bilinguals. Upon viewing a video clip, the English-dominant bilinguals were more likely to indicate that the stimulus was English based on their L1 dominance. The information presented in the video clips may have been processed through their L1, influencing participants to indicate that the stimulus was English more often than it was Spanish. The Spanish-dominant bilinguals also showed a bias to respond more with their native language, although this tendency did not reach significance. All paperwork and instructions were presented in English. Thus, the Spanish-dominant bilinguals did not show the same native-language effects because they were perceptually set in an English mode.

Effects of prior linguistic experience were also observed in the B'' scores obtained in Experiments 2 and 4. In Experiment 2, monolingual English observers displayed a bias toward responding "English" when presented with forwards stimuli, and "Spanish" when presented with backwards stimuli. Because the monolingual English participants had more experience with English, the forwards stimuli were more familiar and natural to the observers. This familiarity influenced them to respond "English" more often when they viewed a forwards video clip. In the backwards condition, the stimuli appeared less familiar, resulting in a greater tendency to respond "Spanish." Thus, when observers were able to recognize a stimulus as a natural articulatory pattern, they showed a greater likelihood of indicating that the video clip was English.

The B'' scores obtained in Experiment 4 suggested similar effects of prior linguistic experience. In this experiment, observers were presented with forwards and backwards English and Spanish words and sentences, and were asked to decide if the video clip had been presented "forwards" or "backwards." The general tendency observed here was to judge all stimuli as "forwards," but the bias to respond

“forwards,” was greater for the English videos than for the Spanish videos, again revealing effects of prior linguistic experience. When presented with the English stimuli – regardless of direction – observers attended to both prosodic characteristics and naturalness. Because all of the English video clips maintained the basic temporal patterns of the observers’ L1, participants had a tendency to indicate that the stimuli were forwards because in some respects, they all appeared to be possible and natural. As monolingual English-speaking participants have more experience with English utterances, English stimuli looked more natural, which may account for why more English video clips were categorized as “forwards” than Spanish video clips.

The effects of prior linguistic experience were also observed in the differential pattern of response frequencies obtained Experiment 3. The greater frequency of “word” responses to English stimuli than Spanish stimuli is clearly a consequence of participants’ being monolingual English speakers. The increased number of “word” responses to all English stimuli – regardless of direction – is likely due to the prosodic cues preserved in both directions. Because the English stimuli contained some degree of naturalness or familiarity, they were categorized more often as “words” than the Spanish stimuli, which exhibited a different prosodic pattern.

We determined that prosodic information and naturalness of the stimuli were two sources of information that observers used when identifying the language spoken in a visual-only video clip. As previously mentioned, the contribution of prosodic information to visual-only language identification was examined in Experiment 2. Temporally-reversed video clips of words and sentences in English and Spanish were presented to observers, who were then asked to decide the language of the video clip. We found that observers were able to reliably identify the language even from backwards stimuli, suggesting that gross differences in prosody are sufficient to support language identification. Lexical information does not need to be present in order for observers to identify languages; prosodic cues alone provide sufficient information for language identification in this task. That sensitivity to language differences was greater in the forwards condition, however, indicates that the presence of additional information available in the forwards stimuli improves identification accuracy.

The objective of Experiment 3 was to determine the extent to which observers were able to access lexical information when provided with visual-only video clips of isolated English and Spanish words. Response frequencies in all stimulus conditions revealed a systematic pattern; “word” responses were more frequent for English stimuli versus Spanish stimuli, and also for forwards videos versus backwards videos. We hypothesized that monolingual English participants would judge English words as “words,” and all other stimuli as “nonwords” based on observers’ lack of experience with Spanish. However, many of the Spanish video clips were also judged as “words,” suggesting that participants were not accessing specific lexical items, but instead may have been making their decisions based on the naturalness of the articulatory gestures and visual trajectories. In the forwards condition, both the English and Spanish video clips appeared natural because they contained temporal and gestural patterns that naturally occur in language. The backwards video clips, however, only maintain gross rhythmic information, and only the temporal patterns of English would have seemed familiar to this group of observers. Thus, backwards Spanish stimuli were judged as “nonwords” more often than backwards English stimuli because they lacked cues to naturalness and familiarity.

The results of Experiment 4 provided additional support for our hypothesis that differences in naturalness were detectable from visual-only displays of speech. Participants were able to reliably identify a stimulus as “forwards” or “backwards” regardless of the language of presentation. This result suggests that visual displays encode a number of highly salient properties (i.e. prosodic, articulatory, and perhaps lexical) that make them appear natural to observers.

Taken together, the results of the four experiments reported here demonstrate that visual displays of speech contain highly detailed information about the speech signal, and that observers' prior linguistic experience affects the way in which these sources of information are processed. We found that prosodic and lexical information, as well as cues to naturalness, are present in the visual signal. Observers are able to attend to and reliably use these sources of information in order to identify English and Spanish in silent video clips. Future investigations of visual-only language identification and discrimination will provide additional insights into how observers complete these tasks, and assess the extent to which lexical, segmental, and suprasegmental (prosodic) information is accessed during visual-only perception.

References

- Abercrombie, D. (1965). *Studies in phonetics and linguistics*. London: Oxford University Press.
- Bernstein, L.E., Eberhardt, S., & Demorest, M. (1986). Judgments of intonation and contrastive stress during lipreading. *Journal of the Acoustical Society of America*, 80, S78.
- Campbell, R., & Dodd, B. (1980). Hearing by eye. *Quarterly Journal of Experimental Psychology*, 32, 85-99.
- Davis, H., & Silverman, S. R (Eds.). (1970). *Hearing and deafness* (3rd ed.). New York: Holt, Rinehart, & Winston.
- Grabe, E., & Low, E.L. (2002). Durational variability in speech and the rhythm class hypothesis. In C. Gussenhoven & N. Warner (Eds.), *Papers in Laboratory Phonology 7* (pp. 515-546). Berlin: Mouton de Gruyter.
- Green, K. (1987). The perception of speaking rate using visual information from a talker's face. *Perception & Psychophysics*, 42(6), 587-593.
- Grier, J. (1971). Nonparametric indexes for sensitivity and bias: computing formulas. *Psychological Bulletin*, 75(6), 424-429.
- Hardison, D. (2003). Acquisition of second language speech: Effects of visual cues, context, and talker variability. *Applied Psycholinguistics*, 24, 495-522.
- Hazan, V., Senemma, A., & Faulkner, A. (2002). Audiovisual perception in L2 learners. In H. L. Hansen and B. Pellom (Eds.), *Proceedings of the 7th International Conference on Spoken Language Processing (ICSLP) 2002*, September 16-20 (pp. 1685-1688).
- Huarte, A., Molina, M., Manrique, M., Olleta, I., & García-Tapia, R. (1996). *Protocolo para la valoraciones de la audicion y el lenguaje, en lengua española, en un programa de implantes cochleares*. Editorial Garsi, Grupo Masson: Madrid.
- Kaiser, A.R., Kirk, K., Lachs, L., & Pisoni, D.B. (2003). Talker and lexical effects on audiovisual word recognition by adults with cochlear implants. *Journal of Speech, Language, and Hearing Research*, 46, 390-404.
- Lachs, L. (1999). Use of partial stimulus information in spoken word recognition without auditory stimulation. In *Research on Spoken Language Processing No. 23* (pp. 81-118). Bloomington, IN: Speech Research Laboratory, Indiana University Bloomington.
- Lachs, L., Weiss, J., & Pisoni, D.B. (2002). Use of partial stimulus information by cochlear implant patients and normal hearing listeners in identifying spoken words: Some preliminary analyses. *The Volta Review*, 102(4), 303-320.
- Massaro, D. (1987). Speech perception by ear and eye. In B. Dodd & R. Campbell (Eds.), *Hearing by eye: The psychology of lip-reading* (pp. 53-83). Hillsdale: Lawrence Erlbaum Associates.
- McGurk, H. & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264, 746-748.
- Pike, K. (1946). *The intonation of American English*. (2nd Edition). Ann Arbor: University of Michigan.

- Soto-Faraco, S., Navarra, J., Weikum, W.M., Vouloumanos, A., Sebastián-Gallés, N., & Werker, J.F. (in press, 2007). Discriminating languages by speech reading. *Perception and Psychophysics*, 69, 218-237.
- Sumby, W.H. & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, 26, 212-215.
- Summerfield, Q. (1987). Some preliminaries to a comprehensive account of audio-visual speech perception. In B. Dodd & R. Campbell (Eds.), *Hearing by eye: The psychology of lip-reading* (pp. 3-51). Hillsdale: Lawrence Erlbaum Associates.
- Weikum, W., Vouloumanos, A., Navarro, J., Soto-Faraco, S., Sebastian-Galles, N., & Werker, J.F. (2007). Visual language discrimination in infancy. *Science*, 316(5828), 1159.
- Werker, J., Frost, P., & McGurk, H. (1992). La langue et les lèvres: Cross-language influences on bimodal speech perception. *Canadian Journal of Psychology*, 46, 551-568.

Appendix A: List of sentences used in Experiments 1, 2A, 2B, and 4

English CID sentences list #9-10 (Davis & Silverman, 1970).

1. Where can I find a place to park?
2. I like those big red apples we always get in the fall.
3. You'll get fat eating candy.
4. The show's over.
5. Why don't they paint their walls some other color?
6. What's new?
7. What are you hiding under your coat?
8. How come I should always be the one to go first?
9. I'll take sugar and cream in my coffee.
10. Wait just a minute!
11. Breakfast is ready.
12. I don't know what's wrong with the car, but it won't start.
13. It sure takes a sharp knife to cut this meat.
14. I haven't read a newspaper since we bought a television set.
15. Weeds are spoiling the yard.
16. Call me a little later!
17. Do you have change for a five-dollar bill?
18. How are you?
19. I'd like some ice cream with my pie.
20. I don't think I'll have any dessert.

Spanish sentences, adaptation of CID list #9-10 (Huarte et al., 1996)

1. El desayuno está preparado en la mesa.
2. Qué le pasará al coche, que no funciona.
3. ¿Crees que el cuchillo cortará bien la carne?
4. No he leído un periódico desde que compré la televisión.
5. Las malas hierbas están estropeando el jardín de mi casa.
6. Llámame si puedes un poco más tarde, por favor.
7. ¿Tienes cambios de mil pesetas en la cartera?
8. ¿Qué tal estás?
9. Me gustaría tomar un poco de helado de chocolate con la tarta.
10. Creo que no tomaré ningún postre.
11. ¿Dónde puedo encontrar un sitio para aparcar?
12. Me gustan las manzanas grandes y rojas que hay en los árboles.
13. Si comes muchos dulces, vas a engordar
14. La película ha terminado tarde.
15. ¿Por qué no pintas las paredes de otro color?
16. ¿Cuál es la noticia mas importante hoy?
17. ¿Qué escondes debajo del abrigo azul?
18. Espera un minuto en la puerta del cine.
19. Pondré azúcar y leche en mi café.
20. ¿Cómo puedo ser siempre el primero en llegar?

Appendix B: List of words used in Experiments 2A, 2B, 3, and 4

List of common English words

1. Monday
2. Wednesday
3. Friday
4. Saturday
5. Sunday
6. One
7. Three
8. Four
9. Five
10. Seven
11. Eight
12. Nine
13. Ten
14. Bird
15. Fish
16. Chicken
17. Duck
18. Dog
19. Donkey
20. Giraffe

List of common Spanish words

1. Lunes
2. Miércoles
3. Viernes
4. Sábado
5. Domingo
6. Uno
7. Tres
8. Cuatro
9. Cinco
10. Siete
11. Ocho
12. Nueve
13. Diez
14. Pájaro
15. Pez
16. Gallina
17. Pato
18. Perro
19. Burro
20. Jirafa