Perception & Psychophysics
2007, 69 (2), 218-231

Discriminating languages by speech-reading

Salvador Soto-Faraco
ICREA and Universitat de Barcelona, Barcelona, Spain

Jordi Navarra
Universitat de Barcelona, Barcelona, Spain

Whitney M. Weikum
University of British Columbia, Vancouver, British Columbia, Canada

Athena Vouloumanos
McGill University, Montreal, Quebec, Canada

Núria Sebastián-Gallés
Universitat de Barcelona, Barcelona, Spain

and

Janet F. Werker
University of British Columbia, Vancouver, British Columbia, Canada

The goal of this study was to explore the ability to discriminate languages using the visual correlates of speech (i.e., speech-reading). Participants were presented with silent video clips of an actor pronouncing two sentences (in Catalan and/or Spanish) and were asked to judge whether the sentences were in the same language or in different languages. Our results established that Spanish–Catalan bilingual speakers could discriminate running speech from their two languages on the basis of visual cues alone (Experiment 1). However, we found that this ability was critically restricted by linguistic experience, since Italian and English speakers who were unfamiliar with the test languages could not successfully discriminate the stimuli (Experiment 2). A test of Spanish monolingual speakers revealed that knowledge of only one of the two test languages was sufficient to achieve the discrimination, although at a lower level of accuracy than that seen in bilingual speakers (Experiment 3). Finally, we evaluated the ability to identify the language by speech-reading particularly distinctive words (Experiment 4). The results obtained are in accord with recent proposals arguing that the visual speech signal is rich in informational content, above and beyond what traditional accounts based solely on visemic confusion matrices would predict.

Speech is a multisensory stimulus because it is simultaneously available to the ear and to the eye (Calvert, Spence, & Stein, 2004; Campbell, Dodd, & Burnham, 1998; Dodd & Campbell, 1987; Massaro, 1998). Although the acoustic aspect of the speech signal is in principle sufficient for understanding the message, the visual correlates of speech carry an important source of information. The brain exploits these associated visual cues, when they are available, in order to decode spoken language more reliably. The importance of audio–visual integration processes is not only highlighted by some theories of speech perception (Fowler, 1996; Liberman & Mattingley, 1985; Massaro, 1998) but is now starting to be revealed by current investigations using brain imaging techniques (see, e.g., Callan et al., 2003; Calvert, Brammer, et al., 1999; Calvert, Bullmore, et al., 1997; Sams et al., 1991; Zatorre, 2001). The behavioral consequences of this audio–visual link in speech processing have been known for a long time. For example, many studies show that correlated visual speech information can greatly enhance the comprehension of spoken messages in noisy conditions (e.g., Sumby & Pollack, 1954), if the message is in a second language (Navarra & Soto-Faraco, 2007; Reisberg, McLean, & Goldfield, 1987), or if it is conceptually difficult to understand (Reisberg et al., 1987). A classic example of the intimate link between linguistic visual and acoustic cues is illustrated by the McGurk illusion (e.g., McGurk & MacDonald, 1976; Soto-Faraco, Navarra, & Alsius, 2004), in which the final auditory percept results from a combination of the visually and acoustically specified speech cues. For example, listening to the spoken syllable /ba/ while watching the lip movements corresponding to the syllable /ga/ often results in the perception of /da/.

The examples above illustrate the contribution of visual information to speech perception under normal hearing con-

S. Soto-Faraco, salvador.soto@icrea.es

ditions, when lip movements and sounds can be combined. However, current knowledge of the specific contribution of visual information to speech processing in the absence of sound (also called *speech-reading* or *lip-reading*) is quite limited in comparison with what is now known about auditory and audiovisual speech perception (see, e.g., Bernstein & Benoît, 1996; Bernstein, Demorest, & Tucker, 1998).

How much information can be extracted from the visual signal alone? The ability to speech-read (i.e., to identify the syllables and words on the basis of visual cues alone) rarely affords robust comprehension (see, e.g., Bernstein et al., 1998; Samuelsson & Rönnberg, 1993) except, perhaps, in a few gifted individuals (see Dodd & Murphy, 1992; Sacks, 1990) or in situations in which the content of the message is strongly constrained by context (e.g., Rönnberg, Samuelsson, & Lyxell, 1998). Perhaps surprisingly, the capacity for speech-reading is somewhat resistant to training, and only modest success has been achieved so far (e.g., Bernstein, Auer, & Tucker, 2001; Heider & Heider, 1940; but see Massaro, Cohen, & Gesi, 1993; Walden, Erdman, Montgomery, Schwartz, & Prosek, 1981; Walden, Prosek, Montgomery, Scherr, & Jones, 1977, for more successful results with the use of isolated syllables). Indeed, past research suggests that even deaf individuals who have been long deprived of hearing and therefore forced to rely on visual cues alone for communicative purposes are not necessarily better speech-readers than hearing individuals (see Conrad, 1977; Lyxell & Rönnberg, 1991; Mogford, 1987; Owens & Blazek, 1985; but see Bernstein et al., 1998, 2000, for a different point of view).

As many researchers have pointed out, the visual correlates of speech are often ambiguous as to which speech sound has actually been produced. For instance, although place of articulation may provide visually accessible information, other phonetic features, such as manner of articulation and voicing, are realized by articulators usually hidden from view in face-to-face situations (see, e.g., Summerfield, 1987; Walden et al., 1977). Consequently, a given visible speech gesture (*viseme*) can correlate with several potential speech sounds (phonemes). The typical example of this involves the phonemes /b/, /p/, and /m/, which correspond to very similar visemes (arguably even the same viseme; Auer & Bernstein, 1997; Fisher, 1968; Massaro, 1998; Owens & Blazek, 1985; Summerfield, 1987). Because of this high degree of visual confusability, it has been argued that many words are virtually indistinguishable from each other on the basis of lipreading alone (a phenomenon called *homopheny*—see Berger, 1972; Nitchie, 1916).

Other studies, however, suggest that there is more detail to be retrieved from the visual speech signal than was originally thought, and predict that finer distinctions are possible even within gestures belonging to the same viseme or between certain *homophenous* words (i.e., words having different sounds but resulting in the same mouth shape; see, e.g., Bernstein et al., 2000). For example, the movement patterns (kinematics) associated with speech (involving the jaw, the cheeks, and the mouth) can be useful for extracting information about the acoustic properties of the signal (Vatikiotis-Bateson, Munhall, Kasahara, Garcia, & Yehia, 1996; Yehia, Kuratate, & Vatikiotis-Bateson, 2002). Head movements carry information regarding the fundamental frequency (Yehia et al., 2002) as well as information regarding suprasegmental features of speech that convey lexical stress, syntactic boundaries, and pragmatics (Hadar, Steiner, Grant, & Rose, 1983, 1984; Munhall, Jones, Callan, Kuratate, & Vatikiotis-Bateson, 2004). There is evidence that even some information regarding the voice of the speaker can be retrieved from the visual speech signal (see, e.g., Kamachi, Hill, Lander, & Vatikiotis-Bateson, 2003). Another possibly important source of information is that provided by the lexical and phonotactic constraints of the language (see Auer & Bernstein, 1997). Thus, even if there is a many-to-one mapping from phonemes to visemes, not all possible phonological interpretations of a given sequence of facial gestures lead to real words or allowable phonotactic combinations. Accordingly, recent studies have found that lexical distinctiveness in visual speech perception is much better than what would be predicted on the basis of the visemic repertoire alone (Auer, 2002; Bernstein, Iverson, & Auer, 1997; Mattys, Bernstein, & Auer, 2002).

Scope of the Present Study

There are few ways to assess the role played by visual information alone, especially if one wants to avoid methodologies based on directly (i.e., explicitly) asking the observer about linguistic properties. Language discrimination is a natural and simple task that can help determine, in an indirect way, the key linguistic information available to the observer that is representative of one or both languages (see Navarra, Sebastián-Gallés, & Soto-Faraco, 2005, on the importance of using indirect measures to test linguistic abilities across languages). Thus, one strategy that would help to clarify the degree of detail provided by visual speech information is to test whether or not it affords discrimination between different languages. For instance, it is almost trivial that adult humans are able to distinguish between their own language and another language, or even between certain foreign (i.e., unfamiliar) languages (see, e.g., Navarra, Spence, & Soto-Faraco, 2007; Ramus & Mehler, 1999), on the basis of auditory information alone. Perhaps less trivial is the fact that discrimination between some language pairs is possible even when the speech signal has been resynthesized so that barely any segmental information is available (i.e., by replacing all vowels with /a/ and all consonants with /s/; see Navarra, Spence, & Soto-Faraco, 2007; Ramus, 2002). Interestingly for the purposes of the present study, these findings indicate that listeners can discriminate languages on the basis of some phonotactic and rhythmic information. Similarly, prelinguistic infants as young as 4.5 to 5 months can discriminate acoustically between the language usually spoken in their environment and unrelated, unfamiliar languages (see, e.g., Bosch & Sebastián-Gallés, 1997; Mehler et al., 1988; Nazzi, Jusczyk, & Johnson, 2000). Prelinguistic infants are even sensitive to the differences between certain pairs of unfamiliar languages as long as these languages belong to distinct rhythmic groups (see, e.g., Nazzi, Bertoncini, & Mehler, 1998). Because infants can make these discriminations not only with naturally spoken language but also when using resynthesized and low-pass filtered speech, researchers have hypothesized

that their (acoustic) discrimination abilities rely on rhythmic patterns (see, e.g., Ramus, Hauser, Miller, Morris, & Mehler, 2000; Ramus, Nespor, & Mehler, 1999). What is not known, however, is whether there is enough information in the visual signal alone to allow successful discrimination between languages. Since the repertoire of visemic units is more limited than the repertoire of phonological units, one would suspect that a distinction between two languages with similar visemic repertoires, highly overlapping lexicons, and similar rhythmic properties would become virtually impossible.

We addressed this question by testing groups of speakers from different linguistic backgrounds using Spanish and Catalan sentences. These two Romance languages are certainly similar in terms of phonological properties, rhythmic pattern, and lexicon, and therefore discrimination should be very difficult. In terms of phonological segments, the Spanish repertoire contains only two sounds—unvoiced fricatives—that are not present in Catalan: /θ/ (as in ***z****apato*, "shoe") and /x/ (as in ***j****amón*, "ham"). It is unclear whether /x/ has a visual correlate that is different from the visible cues provided by other sounds that exist in Catalan (i.e., /g/, /k/). It may be argued, however, that there is a subtle difference between the visual correlates of /θ/ (present in Spanish only) and /d/ (present in both Catalan and Spanish), since the former involves a somewhat larger tongue protrusion. Indeed, these two speech gestures are usually classified into different viseme categories (see, e.g., Auer, 2002; Mattys et al., 2002; Walden et al., 1977). On the other hand, Catalan phonology contains several sounds that do not belong to the Spanish repertoire: The vowels /ɛ/, /ɔ/, and /ə/, as well as the unvoiced fricative sound /ʃ/ (as in ***x****ocolata*, "chocolate") and the voiced fricative sounds /ʒ/ (as in ***g****erro*, "flowerpot") and /z/ (as in *ca****s****a*, "house"). Voicing is not a visually distinctive feature, and it is unclear that these voiced consonants would provide any differential visual correlates with respect to some of the unvoiced consonant sounds existing in the Spanish repertoire (/tʃ/, /x/, and /s/). As for the vowels, the visual correlates of the three Catalan-only vocalic sounds (/ɛ/, /ɔ/, and /ə/) have, again, very close counterparts in the Spanish vowel space (/e/, /o/, and /a/), with the mouth opening being perhaps the most important correlate of the Catalan-only vowels (see MacEachern, 2000, and Summerfield, MacLeod, McGrath, & Brooke, 1989, for visually based vowel distinctions in English). At the suprasegmental level, Spanish and Catalan are both syllable-timed languages and should therefore have similar prosodic properties. There are differences, however. The most relevant ones are that (1) Catalan has vowel reduction whereas Spanish does not, (2) there are more monosyllabic words in Catalan than in Spanish, and (3) Catalan allows consonant clusters in syllabic coda position whereas Spanish does not (Solà, Lloret, Mascaró, & Pérez-Saldanya, 2000).

For this initial investigation, we used natural speech (a silent video clip of the face of a female speaker who was producing sentences) to allow participants to have access to all potential information that the visual signal offers, be it segmental, suprasegmental, or even possibly lexical.

EXPERIMENT 1
Spanish–Catalan Discrimination by Spanish–Catalan Bilinguals

The goal of this experiment was to ascertain whether or not observers can visually discriminate between two very similar languages with which they are familiar. To this end, we selected a group of Spanish–Catalan bilinguals. We divided the sample between Catalan-dominant bilinguals, Spanish-dominant bilinguals, and simultaneous bilinguals in order to detect any asymmetry among these linguistic groups.

Method

Participants. We recruited 48 Spanish–Catalan bilingual undergraduate students from the Universitat de Barcelona for this experiment. We divided the sample into three different groups as a function of each participant's linguistic background[1]: Spanish dominant (both parents spoke Spanish at home), Catalan dominant (both parents spoke Catalan at home), and simultaneous bilinguals (one parent mainly spoke Catalan and the other parent mainly spoke Spanish with the participant). None of the participants reported any hearing problems or any visual problems other than those corrected by wearing lenses. They received course credit in exchange for their participation.

Apparatus and Materials. We recorded the experimental sentences in a sound-attenuated room using a VHS camera recorder. The speaker was a female Spanish–Catalan simultaneous bilingual for whom current use and dominance of the two languages was balanced. She was asked to read the sentences at a normal rate. The sentences were extracted from several poetry texts in order to control metrics (in number of syllabic units) and duration, although we made an effort to avoid very low-frequency words. The sentences were equalized across languages in terms of duration and length (16, 22, and 32 syllables long, 16 sentences each, per language).

The video recordings consisted of a fixed take showing a full view of the speaker's face, from the shoulders to the top of the head. The content of the videotape was digitized into separate computer files (Audio Video Interleave, 25 frames/sec; frame size, 704 × 576 pixels; depth, 24 bits; compressor, DIV3) corresponding to each sentence. Each of these video clips was edited using Adobe Premiere software in order to equalize as much as possible the time between the beginning of the video clip and the onset of the lip movements. All clips started with a four-frame, 160-msec transition from black, videotaped as the speaker was starting the utterance, and ended with a 160-msec transition into black as the speaker was ending the utterance. The image, as shown to the participant, was surrounded by a 6-pixel-wide frame (red or green; see below). The video clips were presented on a 17-in. CRT monitor screen using a Pentium III PC running a custom-made software program.

Procedure. The participants sat at a distance of about 75 cm from the computer screen. Each participant received a total of 48 trials, each following the same sequence (see Figure 1): A white fixation dot appeared in the center of a black screen for 500 msec. Then, a silent video clip of the speaker pronouncing one sentence in Catalan or in Spanish was played inside a red frame. After a 1-sec interstimulus period (black screen), a second video clip of the speaker pronouncing a different sentence (with the same number of syllables) in one of the two languages was played inside a green frame. The participants were asked to press the right button of the mouse (labeled with the word "yes") if they thought the language of the second sentence was the same as that of the first sentence, or to press the left button (labeled with the word "no") if they thought the second sentence was in a different language. The participants were instructed to respond as soon as they were sure of their judgment, during the second sentence (the green frame served as the response cue). A white question mark appeared in the center of the

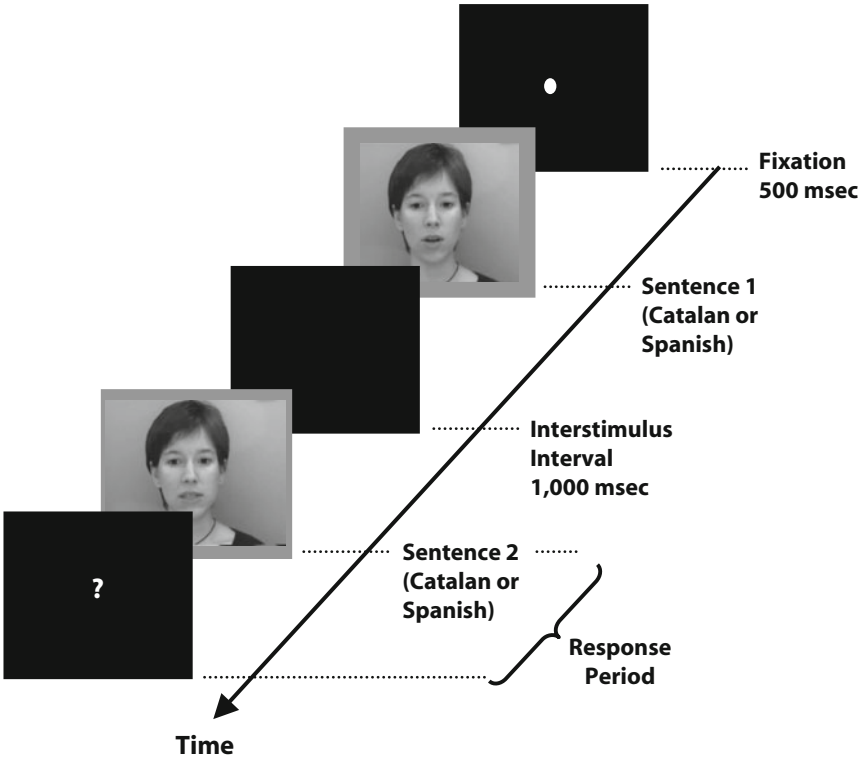

Figure 1. Illustration of the temporal sequence of a trial in Experiments 1–3. The length of Sentence 1 and Sentence 2 was equalized in number of syllables (16, 22, or 32) within each trial and varied randomly across the experiment. The two sentences within a trial were always unrelated in meaning. Responses were allowed from the onset of Sentence 2 (green frame) until 2,000 msec after the end of Sentence 2 (during which time a question mark was presented centrally on a black screen).

screen after the second sentence and stayed on for a maximum of 2,000 msec or until a response had been made. The languages of the first and second sentences were pseudorandomly chosen on a trial-by-trial basis for each participant, with order and the total number of sentences in each language being equiprobable. Each sentence appeared only once in the experiment, and all pairs of sentences in a trial were of the same length in syllables.

Results

We conducted two ANOVAs on the discrimination scores (see Table 1). In one ANOVA, we analyzed the percentage of correct responses averaged by participants, with linguistic group of the participant (Spanish dominant vs. Catalan dominant vs. simultaneous bilinguals) as a between-subjects factor and length in number of syllables (16 vs. 22 vs. 32) as a within-subjects factor. In the other ANOVA, we included percentage of correct responses averaged by items as the dependent variable, with length in number of syllables as a between-items factor and linguistic group as a within-items factor. The analyses did not reveal any significant effect of linguistic group ($F_1 \leq 1$, $F_2 \leq 1$). The effect of number of syllables was significant [$F_1(2,90) = 4.1, p = .019$; $F_2(2,93) = 3.1, p = .05$], with the highest discrimination scores for the longest sentences (57%, 57%, and 64% for the 16-, 22-, and 32-syllable utterances, respectively), although sentences from each of the three syllable lengths were discriminated at above-chance levels as determined by one-sample t tests (all ps $< .001$). The interaction between linguistic group and number of syllables was far from significant in both the participants and the items analyses ($F_1 < 1, F_2 < 1$). Separate t test analyses on each group's average performance revealed that participants from all three groups performed above chance level [Spanish dominant, 57%, $t(16) = 5.3$, $p < .001$; Catalan dominant, 61%, $t(14) = 5.2, p < .001$; simultaneous bilinguals, 60%, $t(15) = 4.5, p < .001$].

Although the two-interval response task used in the experiment should be robust to potential biases arising from the adoption of different response criteria by different groups, we used the signal detection theory method (see Macmillan & Creelman, 1991) to assess the sensitivity and possible differences in criterion. From the hit rate (percentage of "yes" responses given on same-language trials) and the false alarm rate (percentage of "yes" responses given on different-languages trials), we obtained independent estimates of the sensitivity (d') and criterion (c) parameters. Since none of the participants' individual proportions of hits or false alarms was either 0 or 1, we did not have to replace any values in the signal detection analyses (this is also true for Experiments 2 and 3). On average, sensitivity was low but different from zero in all three groups ($d'_{\text{Spanish dominant}} = 0.39$, $d'_{\text{Bilinguals}} = 0.56$, $d'_{\text{Catalan dominant}} = 0.63$). There were no significant differences across the three linguistic groups ($p = .293$). The data also revealed no significant differences in terms

Table 1
Discrimination Performance Scores in Experiments 1–3

Property	% Correct *M*	% Correct *SD*	Hits	FAs	*d′*	Criterion
Experiment 1						
Spanish dominant	57.4*	11.4	.591	.445	0.39	0.14
Catalan dominant	60.9*	14.1	.621	.402	0.63	0.28
Simultaneous bilinguals	59.8*	13.4	.588	.391	0.56	0.31
Experiment 2						
English speakers	53.1	15.8	.528	.466	0.17	0.09
Italian speakers	52.2	14.5	.536	.487	0.12	0.03
Experiment 3 (Spanish monolingual speakers)	55.8*	11.7	.542	.425	0.32	0.20

Note—Correct discrimination (% correct), hit and false alarm (FA) rates, and *d′* and criterion parameters are presented as a function of language background. *Scores significantly above chance.

of criterion ($c_{\text{Spanish dominant}} = 0.14$, $c_{\text{Bilinguals}} = 0.31$, $c_{\text{Catalan dominant}} = 0.28$; $p = .294$).

Discussion

The main result to emerge from Experiment 1 is that the observers were able to visually distinguish between their two languages despite the fact that the languages are phonologically very similar. This result shows, for the first time, that adults can use the information contained in visual speech to distinguish two very similar languages from one another. This demonstration of language discrimination adds significantly to our understanding of visual speech perception. It substantiates previous claims, based on performance in word or phoneme identification tasks, that the information available from visual speech may be richer and more detailed than was initially thought (see Bernstein et al., 2000, for a similar argument). The three different groups of bilinguals tested here showed a consistent pattern of results, and no differences between them were detected.

As would have been expected, we found that the longer the utterances, the higher the success rate in the task. This likely reflects the fact that longer sentences (and/or words) result in more accurate speech-reading, presumably because more information can be gathered and, therefore, contextual, lexical, and/or phonotactic constraints are stronger. What is particularly noteworthy, however, is that our participants were able to discriminate the two languages at above chance levels even in the shortest (16-syllable) utterances.

The fact that Spanish–Catalan bilinguals succeed in the visual discrimination of their two languages raises an important question: What type of information do observers use to discriminate between these two languages? In general, linguistic cues that can help distinguish languages are present in the speech signal, but in the case of the Spanish–Catalan distinction it is unclear how informative these cues can be. As discussed above, there are very few visemes that distinguish between Spanish and Catalan. The high degree of lexical overlap is also important, although our participants may have been able to recognize some specific words, a few of which were perhaps frequent and distinctive in one of the two languages.

However, an alternative explanation of the result obtained in Experiment 1 must be considered first. In particular, it could be that there was some kind of subtle, nonlinguistic cue in the stimuli that allowed the participants to classify the sentences regardless of any other consideration (e.g., the speaker may have looked slightly happier when speaking in one language than in the other). We attempted to minimize the presence of these kinds of cues by using only 1 very well-balanced bilingual speaker for all our recordings and videotaping all the materials during a single session. Yet it is difficult to be completely certain that our speaker did not inadvertently display subtle extralinguistic cues that differed across the two languages. If this were the case, even a group of observers unfamiliar with the two languages should be able to detect these cues and use them to succeed in the discrimination task. Experiment 2 addressed this question directly.

Speakers who are unfamiliar with the languages being tested are less likely to be able to decode and make use of the very subtle linguistic signals that distinguish these two languages. Nor will they be able to use lexical or phonotactic constraints that specify the linguistic distinction at stake. Consequently, their ability to classify Spanish and Catalan on the basis of visual input alone should be poorer than that observed in Experiment 1. If, on the other hand, the materials contain any type of extralinguistic cue that permits discrimination of the stimuli, then even linguistically unfamiliar observers should be able to succeed at this task.

EXPERIMENT 2
Spanish–Catalan Discrimination by Speakers Unfamiliar With These Languages

The next step in this study was to assess whether people can discriminate between the two languages tested in Experiment 1 when these languages are unfamiliar to them. We tested monolingual speakers from two different linguistic backgrounds, English and Italian, who had had no contact (other than, perhaps, occasional) with either Spanish or Catalan. If they can discriminate Spanish from Catalan, then we will not be able to rule out the alternative account that there is some kind of nonlinguistic information in the visual stimuli used in Experiment 1 that aids in discrimination. However, if the unfamiliar observers are unable to discriminate between the two sets of sentences, the potential account based on extralinguistic information will be seriously compromised.

Furthermore, because Spanish and Catalan are Romance languages and also very similar from a rhythmic point of

view, knowledge of a related language (Italian) might help a subset of the participants to discriminate between them. Previous research in adaptation to time-compressed speech has shown that listeners can generalize perceptual adaptation mechanisms to rhythmically similar languages (Sebastián-Gallés, Dupoux, Costa, & Mehler, 2000). In Sebastián-Gallés et al.'s (2000) study, Spanish natives were able to better understand highly time-compressed Spanish sentences if they were previously exposed to time-compressed Spanish sentences than to time-compressed English ones. Results also showed that previous exposure to time-compressed Italian or Greek (two languages that are rhythmically similar to Spanish) produced the same benefits as listening to time-compressed Spanish. If this generalization were applicable to visual input, Italian monolinguals might be able to use their knowledge about their native language to better parse Spanish and Catalan visual sentences and, therefore, perceive the differences between them.

Method

Participants. Twenty English native speakers (undergraduate students at the University of British Columbia) and 15 native speakers of Italian (5 undergraduate volunteers from the Fondazione Santa Lucia in Rome and 10 graduate students from the Scoula Internazionale Superiore di Studi Avanzati in Trieste) were recruited for this experiment. They had only occasional or no previous experience with Catalan or Spanish. All had normal hearing and normal or corrected-to-normal vision. The students from the University of British Columbia received course credit for their participation.

Apparatus, Materials, and Procedure. The experimental methods and testing conditions were closely matched to those used in Experiment 1 except that the participants were now tested in different locations. The Italian participants in Rome were tested using a laptop Pentium PC with a 15-in. LCD monitor in a quiet room, whereas the Italian participants in Trieste were tested in a laboratory room using a Pentium III PC with a 17-in. CRT monitor. The Canadian participants were tested in a sound-attenuated room using a Pentium 4 PC with an 18-in. CRT monitor. All the participants sat at eye level with the monitor, approximately 75 cm from the screen.

Results

We submitted the percentage correct scores of all the participants to two ANOVAs (one in which participants was the random factor and the other in which items was the random factor), with number of syllables (16 vs. 22 vs. 32) and linguistic group (English vs. Italian) as independent variables. Both the main effect of number of syllables ($F_1 < 1$, $F_2 < 1$) and the main effect of language background ($F_1 < 1$, $F_2 < 1$) failed to reach statistical significance. The interaction between number of syllables and group was significant by participants but not by items [$F_1(2,66) = 3.32$, $p = .042$; $F_2(2,45) = 2.32$, $p = .110$]. This interaction, however, was not accompanied by a significant effect of number of syllables in either of the two linguistic groups when tested individually [English speakers, $F(2,38) = 1.47$, $p = .242$; Italian speakers, $F(2,28) = 2.17$, $p = .133$]. Moreover, further analyses revealed that, unlike the Spanish–Catalan bilinguals tested in Experiment 1, the participants in this experiment were not able to perform at a better-than-chance level [$M = 52\%$ overall; $t(34) = 1.5$, $p = .179$]. This was true for each linguistic group [English group, $M = 53\%$, $t(19) = 1.3$, $p = .213$; Italian group, $M = 52\%$, $|t| < 1$] and for each syllable length tested separately for each of the two linguistic groups [English group, $t(19) = 1.5$, $p = .140$, $|t| < 1$, and $t(19) = 1.2$, $p = .235$ for the 16-, 22-, and 32-syllable sentences, respectively; Italian group, $|t| < 1$, $t(14) = 1.8$, $p = .095$, and $|t| < 1$ for the 16-, 22-, and 32-syllable sentences, respectively].

We compared the accuracy of participants in this experiment with that of the Spanish–Catalan bilinguals of Experiment 1. We found a significant overall difference between the Spanish–Catalan bilinguals in Experiment 1 and the linguistically unfamiliar observers in Experiment 2 [$t(81) = 3.54$, $p < .001$]. In a further effort to confirm the reliability of the result, we used one-tailed t tests to evaluate the hypothesis that every linguistic group included in Experiment 1 (Spanish dominant, Catalan dominant, and simultaneous bilinguals) had performed better than both linguistic groups included in Experiment 2 (English speakers and Italian speakers). The analyses showed that, indeed, all comparisons between the bilingual participants of Experiment 1 and the unfamiliar participants of Experiment 2 resulted in significant differences ($p < .05$).[2]

In the analysis of the signal detection parameters in Experiment 2, sensitivity was significantly lower than that in Experiment 1 ($d'_{Exp1} = 0.52$, $d'_{Exp2} = 0.15$, $p < .001$), and overall d'_{Exp2} was not different from zero ($p = .102$). This was also true for each language group separately ($d'_{English} = 0.14$, $p = .286$; $d'_{Italian} = 0.13$, $p = .327$). The criterion was also marginally lower than what was seen in Experiment 1 ($c_{Exp1} = 0.24$, $c_{Exp2} = 0.07$; difference, $p < .03$), suggesting that the participants in this experiment had less of a tendency to prefer the "different" response.

Discussion

Several conclusions emerge from the results of Experiment 2. First, there are no extralinguistic cues in the visual message that the observers were able to use to classify the stimuli. This demonstration is important because it restricts the set of potential explanations for the results of Experiment 1 to the realm of linguistic information. Second, in contrast with the results of Experiment 1, here the length of the utterances did not have any relevant effect on accuracy,[3] although the participants in Experiment 2 took an average of 1,034 msec longer to inspect the sentences than did the participants in Experiment 1 ($p < .001$). The fact that, unlike performance in Experiment 1, performance in Experiment 2 did not vary with sentence length provides yet more support for the linguistic nature of discrimination in Experiment 1, since speech-reading performance is usually positively correlated with the length of the sentence or word.[4]

As we discussed in the introduction, one of the potentially important cues for distinguishing two unfamiliar languages, at least when sounds are available, is their rhythmic patterns. This is the most prominent cue that prelinguistic infants and adults unfamiliar with the test languages resort to when confronted with an auditory language discrimination task. Because Spanish and Catalan belong to the same rhythmic class (both are syllable based—see, e.g., Ramus et al., 1999), the overall prosodic characteristics of the two languages may sound (and look) very similar to the unfamiliar ear (and eye). Unlike English, which is stress based, Italian belongs to the same rhythmic class as Spanish and

Catalan (i.e., it is syllable based). One might therefore expect that Italian speakers, more familiar with the rhythmic aspect of the two languages, would be better able to extract some language-specific information from the visual speech in this experiment. Yet, the Italian speakers were no better than the English speakers at distinguishing whether they were being exposed to Spanish or to Catalan. This suggests that familiarity with the rhythmic class alone does not allow successful language discrimination.

Taken together, the results of Experiments 1 and 2 lead to the conclusion that a high degree of familiarity with the languages is a necessary condition for successful visual discrimination. In Experiment 3, we investigated whether the observer must be familiar with both languages, or whether familiarity with one language is sufficient for discrimination.

EXPERIMENT 3
Spanish–Catalan Discrimination by Spanish Monolinguals

The next experiment of this series addressed whether or not familiarity with only one of the two languages is sufficient to enable visual discrimination between Spanish and Catalan. To this end, we tested a group of Spanish monolingual speakers born and raised outside Catalonia who had only very occasional, if any, previous contact with spoken Catalan.

Method

Participants. Seventeen undergraduate students at the University of the Basque Country (Vitoria, Spain) participated in this study. All spoke Spanish as their native language and had very occasional, if any, previous experience with Catalan. They reported normal hearing and normal or corrected-to-normal vision.

Apparatus, Materials, and Procedure. These were exactly as in Experiment 1 except that the experiment was conducted in a different geographical location (the Basque Country) using a different computer (Acer Travelmate 654LC laptop PC with a 15-in. LCD monitor). The participants were tested inside a sound-attenuated laboratory room.[5]

Results

We submitted the percentage of correct responses to two one-way ANOVAs (one by averaged participants and the other by items) with number of syllables (16 vs. 22 vs. 32) as the factor. The analyses returned null results ($F_1 < 1$, $F_2 < 1$). We then tested the group average against chance level (in this case, 50%) and found a significant difference [$M = 56\%$, $t(16) = 3.8$, $p = .001$], indicating that the Spanish monolinguals were able to discriminate between Spanish and Catalan sentences. Finally, the comparison between the Spanish monolinguals and the entire sample of Spanish–Catalan bilinguals tested in Experiment 1 revealed a numerical, albeit only marginally significant, difference in favor of the latter [$t(63) = 1.7$, $p = .098$]. We then performed one-tailed t tests to test the hypothesis that each group of bilinguals included in Experiment 1 had performed better than the monolinguals in Experiment 3. The analyses revealed that the Catalan-dominant group achieved significantly better discrimination scores than did the Spanish monolinguals ($p < .05$), whereas the scores of the simultaneous bilinguals were marginally better than those of the Spanish monolinguals ($p = .07$). In contrast, no differences were observed between the Spanish-dominant bilinguals and the Spanish monolinguals ($|t| < 1$).

When testing the signal detection parameters for the Spanish monolingual group, we found that sensitivity was marginally lower than that in Experiment 1 overall ($d'_{Exp1} = 0.52$, $d'_{Exp3} = 0.32$, $p = .092$). Groupwise, this difference was significant when the Spanish monolinguals were compared to the Catalan-dominant bilinguals ($d'_{\text{Catalan dominant}} = 0.63$, $p < .05$) but not in the other comparisons (both ps $> .1$). With respect to response criterion, there were no differences overall with respect to Experiment 1 ($c_{Exp1} = 0.24$, $c_{Exp3} = 0.20$, $p = .662$).

Discussion

The results of Experiment 3 are clear in demonstrating that knowledge of only one of the two test languages (i.e., Spanish) enables observers to successfully discriminate natural connected speech in Spanish and Catalan on the basis of visual cues alone. The comparison between the results of Experiment 3 and those of Experiment 1 suggests that knowledge of both languages (rather than of only one of them) further facilitates performance in the visual discrimination task. Although the difference between Experiments 1 and 3 was only marginally significant, this particular conclusion is supported by two secondary findings. First, when the groups were tested separately, the only group in Experiment 1 not to show a reliable advantage with respect to the monolinguals in Experiment 3 was the Spanish-dominant group. Second, the monolingual participants in Experiment 3 required significantly more time to make their decisions than did the bilinguals tested in Experiment 1 (660-msec mean difference in RTs between groups, $p < .05$).[6] One could raise the potential concern that, since we did not measure our participants' lipreading ability in advance, there might be individual differences across linguistic groups that could explain the differences found. However, we think this is unlikely, since we can assume that the chance of finding good lip-readers is equivalent for both linguistic groups.[7] Thus, overall, the present pattern of data provides support for the idea that experience with both languages might provide an advantage in decoding (and using) the available visual correlates necessary for successful language discrimination.

Once it was established that the average bilingual speaker can discriminate between his or her two languages on the basis of visual speech correlates and that the relevant information is likely to be linguistic in nature, we revisited the data collected in Experiment 1 in order to find out more about the potential cues that our speakers might have used.

FURTHER ANALYSES OF THE SENTENCE DISCRIMINATION DATA OF EXPERIMENT 1

In order to further investigate visual speech as the basis for successful discrimination between Spanish and Catalan, we conducted additional analyses as a function of several phonological and lexical properties of the stimuli that could potentially help to characterize each language. We extracted indexes about features regarding segmental

information (characteristic phonemes, phonotactic combinations), correlates of rhythm (vowel reduction, ratio of vowels to consonants), and lexical information (number of cognates). We calculated a numerical index (see below) for each of these cues and for each of the sentences used (independent of language). Then, we performed regression analyses for each of the cues, separately and grouped, with the average discrimination scores in Experiment 1 as the dependent variable (given that this is where the participants achieved the best discrimination results). We based the sentence statistics on the (acoustic) phonological properties, not directly on visemic information. As we discussed earlier, and in consistency with the results of the present experiments, traditionally defined visual-only information (visemes) is generally a poor tool for differentiating speech samples from these two languages. Therefore, we decided to use (richer) phonological descriptions instead, given that, no matter what visual cues (or combination of cues) the participants might be using to discriminate, these should manifest themselves in terms of the (auditory) phonological makeup.

Data extraction and transformation. First, from the data of Experiment 1 (in which successful discrimination was observed), we calculated a discrimination index for each sentence in the experiment. This index was the average of correct responses across all trials in which the sentence was present ($M = .60$, $SD = .09$; range, .30–.80). Second, for each sentence we assessed the following values: characteristic phonemes (Phon), defined as the number of phonemes in the sentence that exist in one language but not in the other (see the introduction)[8]; ratio of vowels to consonants (RVC); and number of cognates (Cog), defined as the number of content words in a sentence that are similar phonologically, etymologically, and semantically between the two languages. Two more variables were extracted for the Catalan sentences only, since these properties are rarely if ever present in Spanish: characteristic phonotactic combinations (Clust), which are those clusters or word endings that are present in Catalan but not allowed in Spanish; and vowel reduction (Reduc), which was the number of weak (reduced) vowels in a sentence. All these values (except for RVC, which is naturally expressed as a proportion) were divided by the number of syllables in the sentence, so sentence length (which has already been singled out as a significant factor for discrimination) was canceled out. Finally, we transformed each variable into an index ranging from 0 to 1, so that values between variables were expressed along the same scale (see Table 2).

Data analyses. All the variables that could be computed for both Catalan and Spanish sentences were compared using a Student's *t* test in order to find potential overall differences between the two languages. None of the comparisons yielded significant results (for Phon, $p = .595$; for RVC, $p = .08$; for Cog, $p = .132$). We then introduced these three variables as independent factors in a linear regression analysis with discrimination performance as the dependent variable. Both in combination and individually, all failed to account significantly for the variation observed in the discrimination index [$F(3,76) = 1.0$, $p = .379$; for Phon, $p = .403$; for RVC, $p = .141$; for Cog, $p = .681$]. We used the results of the Catalan sentences only to incorporate all the variables (Phon, RVC, Cog, Clust, and Reduc) in a new regression analysis. Again, neither the full regression equation nor any of the individual regressors included in the analysis explained a significant amount of variation [$F(5,26) < 1$, $p = .777$; for Phon, $p = .462$; for RVC, $p = .285$; for Cog, $p = .450$; for Clust, $p = .399$; for Reduc, $p = .896$]. See Figure 2 for the partial correlations of each variable against the discrimination index.

None of the variables extracted from the sentences was able to clearly explain the ability of bilingual speakers of Spanish and Catalan to discriminate languages visually. The best predictor of performance, although far from significant, was RVC. This ratio has usually been considered a good correlate of the rhythmic characteristics of languages, including Spanish and Catalan (Ramus et al., 1999). In our sample of sentences, the difference in RVC between languages was only marginally significant ($p = .08$), since it corresponds to languages belonging to the same rhythmic class (see Ramus et al., 1999). To increase the sensitivity of our comparison to this potential cue, we decided to rank the Catalan and Spanish sentences in terms of their RVCs and assess the discrimination index for the 12 sentences with the highest RVCs ($RVC_{Catalan} = .48$; $RVC_{Spanish} = .49$) and the 12 sentences with the lowest RVCs ($RVC_{Catalan} = .40$; $RVC_{Spanish} = .41$) in each language. We found that, for Catalan sentences, people were better at discriminating those with lower RVCs than those with higher RVCs (66% vs. 59%, $p < .05$). For Spanish sentences, however, this was not true, and discrimination was similar for high- and low-RVC sentences (56% vs. 58%, $p = .647$). Similar analyses based on low versus high ratings did not provide any significant result for any of the remaining variables.

Discussion. The outcome of the statistical analyses based on the sentence phonological profile provided only a limited amount of information regarding the source of discrimination ability. Neither the individual linguistic variables included nor their combination seemed to constitute a clear basis for the prediction of the accuracy scores observed. Of course, this does not mean that these variables are necessarily idle with respect to the mechanisms used

Table 2
Average-per-Syllable and Transformed Indexes (Range, 0–1) Corresponding to the Properties Analyzed for the Sentences Used in Experiments 1–3

	Avg./Syll.		Transformed	
Property	Spanish	Catalan	Spanish	Catalan
Phon	0.062	0.068	.329	.365
RVC	0.452	0.438	.546	.464
Cog	0.192	0.164	.513	.438
Clust	–	0.092	–	.294
Reduc	–	0.456	–	.530

Note—Spanish and Catalan materials are presented separately. Phon, phonological segments present in one language but not in the other; RVC, ratio of vowels to consonants; Cog, cognate words (content words only); Clust, characteristic phonological combinations existing in Catalan but not in Spanish; Reduc, vowel reduction.

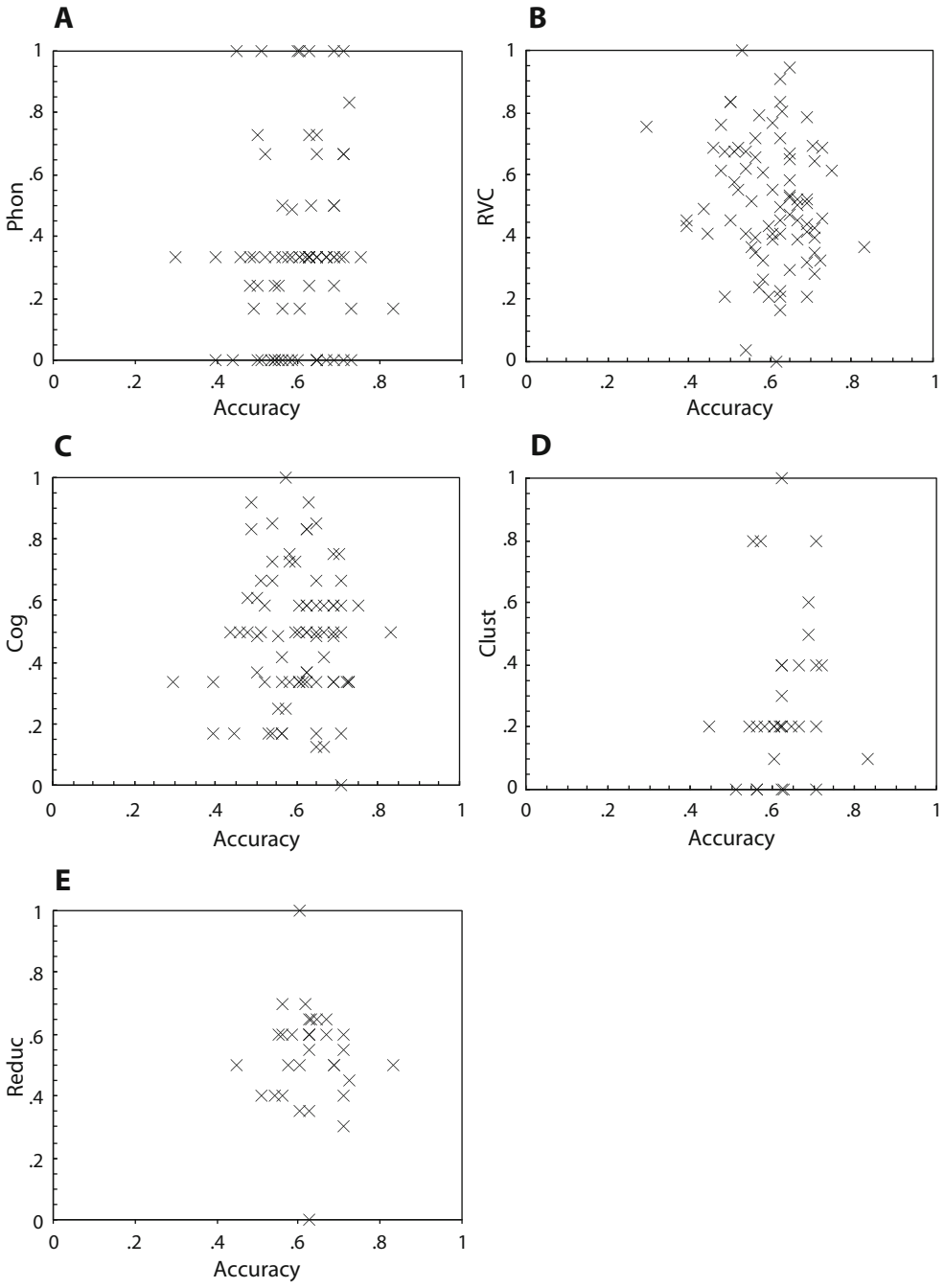

Figure 2. The panels plot each sentence in Experiment 1 as a function of discrimination performance (*x*-axis) and five indexes regarding the phonological properties of the sentences (*y*-axis). These indexes are (A) characteristic phonemes (Phon), (B) ratio of vowels to consonants (RVC), (C) number of cognates (Cog), (D) characteristic phonotactic combinations (Clust), and (E) vowel reduction (Reduc).

during discrimination, but that their role might be more subtle than can be picked up from our post hoc regression analyses. The particular case of the RVC, which is a correlate of the rhythmic characteristics of the languages, is perhaps illustrative, since more powerful analyses might reveal a potential role of this parameter.

Another important aspect for language discrimination that has possibly been underestimated in the preceding analyses is the potential ability of the participants to identify which language was being spoken by picking up just one or two occasional words by speech-reading. Given the inconclusive results of the regression analyses based on segmental and suprasegmental properties, it is perhaps important to test how much lexically based information can contribute to the successful discrimination between languages. This possibility will be addressed in the next and final experiment.

EXPERIMENT 4

The aim of this experiment was to address the potential ability to identify the actual language being looked at and to determine whether or not this ability may rely on speech-reading isolated words. To this end, participants were asked to classify visually presented sentences according to language (Spanish or Catalan) and to report any words they were able to pick up by speech-reading. This manipulation provides an alternative measure of language discrimination based on an identification task rather than on the two-interval (same–different) task used in Experiments 1–3. It allows us to measure sentence identification performance as a function of how well the words in the sentence can be speech-read. This can help uncover the role that lexically based strategies might have played in the discrimination ability seen in Experiment 1.

Method

Participants. Sixteen students at the University of Barcelona participated in this study. We indiscriminately included participants from all three linguistic backgrounds tested in Experiment 1 (Spanish dominant, $n = 6$; Catalan dominant, $n = 6$; simultaneous bilinguals, $n = 4$), since they did not display any significant differences on the language discrimination task. They reported normal hearing and normal or corrected-to-normal vision.

Apparatus, Materials, and Procedure. These were as in Experiment 1, except for the following. The sentences were presented on the same monitor used in Experiment 1 but with a different computer (Acer Travelmate 292 Pentium processor). The participants were presented with a single sentence in each trial. After each sentence, they were asked to respond verbally whether the sentence was in Catalan or in Spanish and then to try to retrieve as many words as they could from the sentence. The experimenter kept a written record of the participants' responses and started a new trial after the response had been recorded. The experiment included a single block in which all 96 sentences used in Experiments 1–3 were presented in random order for each participant. This time, no color frame surrounded the video clip display.

Results

We scored the correct responses in the language identification task and calculated performance accuracy. We submitted the accuracy data to two ANOVAs (one by participants and one by items) with sentence language (Spanish vs. Catalan) and sentence length (16 vs. 22 vs. 32 syllables) as factors. Overall, the participants identified the language on 71% of the sentences, which is significantly above chance ($p < .001$). There was a positive effect of sentence length (significant only by participants) whereby language identification was easier for longer sentences [$F_1(2,30) = 5.1$, $p = .012$; $F_2(2,90) = 1.7$, $p = .190$], with averages of 67%, 71%, and 74% for the 16-, 22-, and 32-syllable sentences, respectively. Sentence language produced null effects ($F_1 < 1$, $F_2 < 1$), confirming that the two languages were not particularly distinct (70% and 72% for Spanish and Catalan, respectively).

Speech-reading ability was measured as the percentage of content words[9] that the participants could correctly retrieve from the sentences. We scored as correct morphological derivations of the word actually presented (i.e., plural vs. singular, past vs. present tense) and words that were phonologically very close to a word present in the sentence. Overall, the speech-reading task was very challenging, with participants reporting an average of 0.87 words per sentence (about 10% of the actual number of words presented per sentence). The average number of correct words picked up from viewing the sentences was 0.24 out of 9.1—that is, 2.5% (2.0% for Catalan sentences and 3.0% for Spanish sentences). In general, speech-reading scores correlated with language identification (by participants, $r = .79$, $p < .001$; by items, $r = .24$, $p = .018$).

In order to evaluate further the potential role of speech-reading, we also assessed the number of reported words that, correct or not, were distinctive of language—that is, words that are not acoustically similar to words in the other language and could therefore, by themselves, signal which language was being spoken. Of the total 1,536 sentence presentations (96 sentences × 16 participants), responses contained a potentially distinctive word on 553 of them (36%). The probability of a sentence's being correctly classified as Spanish or as Catalan given that a distinctive word had been reported was .79, whereas the correct language identification rate when no distinctive word had been reported was .64 (which is significantly smaller [$p < .001$] but still above chance [$p < .001$]).

Discussion

The language identification accuracy seen in Experiment 4 (71%) was very much in line with the discrimination scores observed in Experiment 1 (59%). In fact, we can argue that the participants performed the same–different discrimination task of Experiment 1 by correctly identifying languages at $p = .71$. Since discrimination is correct when participants identify the language of each sentence correctly, as well as when they identify both sentences incorrectly, discrimination performance should approximate a simple model of the form $p^2 + (1 - p)^2$. Using the result of Experiment 4 to establish p at .71, this rather simple model predicts a discrimination accuracy of about .59 in the two-alternative forced-choice task, which compares well with the result observed in Experiment 1 (.59).

The second finding arising from this experiment relates to the contribution that word recognition by speech-reading might have had on language identification (and consequently, according to the argument mounted above, on discrimination in Experiments 1 and 3). The correlation between speech-reading ability and language identification was high both when the participants' scores were used and when the slightly less biased correlation based on items accuracy scores was performed. This suggests a possible relation between the two factors and thereby supports a possible role of lexical strategies. This conclusion is further supported by the fact that sentences whose responses contained language-distinctive words were more likely to be classified correctly than sentences whose responses did not contain such words. This is, again, circumstantial evidence given that it stems from correlational analyses. Yet, so far, and given the limited success of tests addressing other levels of linguistic analysis, it provides some indication that lexical strategies might have played a role. It is important to stress, however, that the participants must have used other kinds of information or strategies as well, given

that when no distinctive words were reported, performance was still above what one would expect by chance.

GENERAL DISCUSSION

The results of this study have revealed that it is possible to discriminate between phonologically and visually very similar languages on the basis of speech-reading alone (Experiments 1 and 3) and to identify them when each of them is seen in isolation (Experiment 4). This indicates not only that the information necessary for discrimination is present in the visual speech signal, but that observers can decode and use it. However, this result does not reflect a general speech-reading capacity, because we have found that this ability is constrained by specific linguistic experience (Experiments 2 and 3). In the remainder of this article, we discuss the potential implications of these findings.

The Possible Bases of Visual Language Discrimination

The present results provide evidence supporting the hypothesis that visual discrimination between Spanish and Catalan by individuals fluent in one or both languages was achieved on the basis of the linguistic information contained in the signals. Although it is difficult to make strong claims as to which kind of linguistic information, or which combination of linguistic cues, the observers may have used, we have begun to explore several possibilities. For instance, we suspect that the discrimination ability shown by our participants is supported by the use of a combination of different features, given that none of the several potential cues alone is very informative as to the differences between the two languages. In terms of segmental cues, as discussed in the introduction, most of the distinctive sounds in one language or the other actually map onto visemes (visible linguistic gestures) that are very similar to (if not indistinguishable from) the gestures produced during the pronunciation of sounds existing in both languages. This was supported by the results of our regression analyses based on the proportion of distinctive sounds present in the sentences (Phon index). Therefore, for discrimination, one can appeal to the use of only a few subtle visual phonological correlates within the categories of the visemic repertoire.

The rhythmic pattern of the language is another possible candidate. Although Spanish and Catalan belong to the same general rhythmic class (i.e., both are syllable-based languages—see Ramus et al., 1999), there are some other possible cues that could have an effect on prosody. For instance, the distribution and frequency of vowels in the language have been associated with discrimination abilities in young prelinguistic children using acoustic stimuli (see Bosch & Sebastián-Gallés, 2001, for the specific case of Spanish–Catalan discrimination). Supporting the potentially important role of these types of suprasegmental cues in the present results, recent studies reveal that there are reliable visual cues to rhythm, especially in head movements, when observers watch someone speak (Munhall et al., 2004; Vatikiotis-Bateson et al., 1996; Yehia et al., 2002). Therefore, the subtle rhythmic distinctiveness between Spanish and Catalan remains a potentially important basis for discrimination in this study. This has been partially confirmed in our regression analyses, in which RVC stood out as a potentially important cue for discrimination.

Finally, we have also investigated the potential role played by lexical cues in allowing language discrimination. Since Catalan and Spanish have largely overlapping lexicons—a limitation that is aggravated here because lexical distinctiveness is highly reduced when it comes to visually based speech-reading—one might think that lexical distinctions are unlikely to play a large role. This is somewhat supported by the poor performance of the cognate index as a regressor of participant performance in our sentence analyses. In Experiment 4, we tested directly the ability to speech-read words from the sentences and, as we expected, we obtained a quite limited result (the participants were able to retrieve a mere 2.5% of the words). Nevertheless, this rather low amount of information seemed to be enough to contribute to successful language discrimination, since it appears that the participants might have picked up isolated words that could be language distinctive.

Whatever the cue or, possibly, the combination of cues underlying this ability, what seems clear from our results is that participants must be familiar with the visual correlates of at least one of the languages in order to decode them. The results from the Italian participants indicate that familiarity with the rhythmic characteristics of the languages alone is not sufficient, because the Italian speakers performed no better than the English speakers. However, speakers who are familiar with the prosodic patterns and who know at least one of the to-be-discriminated languages are able to extract sufficient information for discrimination. This combination of results seems to point to the idea that either the subtle rhythmic differences between Catalan and Spanish can be appreciated only by a native speaker, or that other linguistic cues (perhaps based on visual lexical access) are required to facilitate discrimination of highly similar languages.

Extrapolation of Other Language Combinations and Linguistic Groups

As is the case with most psycholinguistic research, one must be cautious in generalizing the results obtained in one language or cross-linguistic combination to other languages. In our case, this is perhaps even more important because there are several reasons to predict different outcomes as a function of the language combination being tested (as well as of the native languages of the observers). For example, different languages may differ in the number of sounds represented by a given viseme (and, therefore, in the ambiguity of the viseme). They also may differ in terms of lexical distinctiveness (see MacEachern, 2000) and possibly in other critical aspects, such as the visual correlates of prosody (e.g., tonal languages make extensive use of lexical prosody). As was mentioned above, in the present study we tested one of the most difficult cases because the two languages used are very similar in most aspects, and we have found that speakers of both (or

even only one) of the two languages are able to decode the relevant visual cues present in the signal. Follow-up studies using two test languages that are more distinct in a number of important features, such as French and English, are necessary. (Such a study is currently under way in our laboratories.) French has a distinctive (syllable-based) rhythmic pattern that is different from that of English (which is stress based). It will be interesting to determine whether, in this easier discrimination test, specific knowledge about the languages is still a requirement or whether any observer (regardless of his or her linguistic background) would be able to discriminate them. Given the potential role of rhythmic cues in these results and previous research showing a significant role of rhythmic class in the discrimination of low-pass filtered (acoustic) speech, one might speculate that visual and acoustic discrimination might bear on some common mechanisms. Looking at visual and low-pass filtered acoustic speech within a common set of languages could provide some support for this hypothesis (Navarra et al., 2007).

Independent of the similarities or differences between the two test languages, another possibly important factor to investigate is the linguistic background of the observer. As psycholinguistic research in auditory speech processing has shown, the sensitivity to specific speech features varies as a function of linguistic background (see Soto-Faraco, Sebastián-Gallés, & Cutler, 2001). For example, French listeners are not sensitive to lexical stress (Dupoux, Pallier, Sebastián, & Mehler, 1997; Dupoux, Peperkamp, & Sebastián-Gallés, 2001), arguably because lexical stress is not a useful cue for word discrimination in their language (French has a fixed stress pattern). Finnish native speakers are sensitive to vowel harmony within words, unlike speakers of many other languages, including English (Suomi, McQueen, & Cutler, 1997).

Cross-linguistic manipulations will offer a wide range of possible tests to refine the characterization of visual discrimination ability and, therefore, to gain specific knowledge about the basis of speech-reading and what it may contribute to language processing in general.

Conclusions and Future Directions

The present findings support the following conclusions: (1) The visual speech signal carries sufficient cues for language discrimination even between very similar languages; (2) the average observer (not specifically trained or selected on the basis of lipreading ability) is able to decode these cues; and (3) this ability is limited to observers who are experienced speakers of at least one of the two languages. These results support the idea that the observer is able to extract more information from the signal than is predicted by the classic measures of visual speech distinctiveness. We have proposed that the mechanisms underlying this ability are based on segmental, suprasegmental, and lexical processes or, possibly, a combination thereof.

The present findings provide a foundation for future investigation of several relevant issues. First, testing of additional linguistic groups and language discriminations will help refine our knowledge of the feature or combination of features that are present (and usable) in the signal. A second important aspect for further research relates to the role of visual speech information in development. It is known that prelinguistic infants are sensitive to several visual correlates of speech, such as audiovisual synchrony (see, e.g., Dodd, 1979), gender correspondence (Patterson & Werker, 2002), and even segmental correspondence between sounds (vocalic and consonantal) and gestures (see, e.g., Burnham & Dodd, 2004; Desjardins, Rogers, & Werker, 1997; Desjardins & Werker, 2004; Kuhl & Meltzoff, 1982, 1984; Patterson & Werker, 1999, 2002). It will be interesting to address whether or not infants are also sensitive to language information and, if so, what use they can make of it to facilitate language acquisition in bilingual environments (for preliminary evidence, see Weikum et al., 2004).

AUTHOR NOTE

This work was supported by a grant from the Human Early Learning Partnership of British Columbia, Grant TIN2004-04363-C03-02 from the Ministerio de Educación y Ciencia of Spain and the "Ramón y Cajal" Program, Grant 410-2004-0744 from the Social Sciences and Humanities Research Council of Canada, Human Frontier Science Program Grant RPG 68/2002, and Grant JSMF-20002079 from the James S. McDonnell Foundation. We thank Ruth de Diego, who served as the speaker in the video recordings, Xavier Mayoral for technical assistance, Agnes Caño and Juan Manuel Toro for their help testing the Italian speakers, Mikel Santisteban and Itziar Laka for their help testing the Spanish monolingual speakers, and Aida Mallorquí for help scoring the results of Experiment 4. Correspondence concerning this article should be addressed to S. Soto-Faraco, Parc Científic de Barcelona, Hospital de Sant Joan de Deu, Edifici Docent, c/Santa Rosa 39-57 planta 4a, 08950 Esplugues, Barcelona, Spain (e-mail: salvador.soto@icrea.es).

REFERENCES

AUER, E. T., JR. (2002). The influence of the lexicon on speech read word recognition: Contrasting segmental and lexical distinctiveness. *Psychonomic Bulletin & Review*, **9**, 341-347.

AUER, E. T., JR., & BERNSTEIN, L. E. (1997). Speechreading and the structure of the lexicon: Computationally modeling the effects of reduced phonetic distinctiveness on lexical uniqueness. *Journal of the Acoustical Society of America*, **102**, 3704-3710.

BERGER, K. W. (1972). Visemes and homophenous words. *Teacher of the Deaf*, **70**, 396-399.

BERNSTEIN, L. E., AUER, E. T., JR., & TUCKER, P. E. (2001). Enhanced speechreading in deaf adults: Can short-term training/practice close the gap for hearing adults? *Journal of Speech, Language, & Hearing Research*, **44**, 5-18.

BERNSTEIN, L. [E.], & BENOÎT, C. (1996). For speech perception three senses are better than one. In *Proceedings of the 4th International Conference on Spoken Language Processing (ICSLP 96)* (Vol. 3, pp. 1477-1480). New York: IEEE Press.

BERNSTEIN, L. E., DEMOREST, M. E., & TUCKER, P. E. (1998). What makes a good speechreader? First you have to find one. In R. Campbell, B. Dodd, & D. Burnham (Eds.), *Hearing by eye II: Advances in the psychology of speechreading and auditory–visual speech* (pp. 211-227). Hove, U.K.: Psychology Press.

BERNSTEIN, L. E., DEMOREST, M. E., & TUCKER, P. E. (2000). Speech perception without hearing. *Perception & Psychophysics*, **62**, 233-252.

BERNSTEIN, L. E., IVERSON, P., & AUER, E. T., JR. (1997). Elucidating the complex relationships between phonetic perception and word recognition in audiovisual speech perception. In C. Benoît & R. Campbell (Eds.), *Proceedings of the ESCA/ESCOP Workshop on Audio–Visual Speech Processing* (pp. 89-92).

BOSCH, L., & SEBASTIÁN-GALLÉS, N. (1997). Native-language recognition abilities in 4-month-old infants from monolingual and bilingual environments. *Cognition*, **65**, 33-69.

BOSCH, L., & SEBASTIÁN-GALLÉS, N. (2001). Evidence of early lan-

guage discrimination abilities in infants from bilingual environments. *Infancy*, **2**, 29-49.

BURNHAM, D., & DODD, B. (2004). Auditory–visual speech integration by prelinguistic infants: Perception of an emergent consonant in the McGurk effect. *Developmental Psychobiology*, **45**, 204-220.

CALLAN, D. E., JONES, J. A., MUNHALL, K., CALLAN, A. M., KROOS, C., & VATIKIOTIS-BATESON, E. (2003). Neural processes underlying perceptual enhancement by visual speech gestures. *NeuroReport*, **14**, 2213-2218.

CALVERT, G. A., BRAMMER, M. J., BULLMORE, E. T., CAMPBELL, R., IVERSEN, S. D., & DAVID, A. S. (1999). Response amplification in sensory-specific cortices during crossmodal binding. *NeuroReport*, **10**, 2619-2623.

CALVERT, G. A., BULLMORE, E. T., BRAMMER, M. J., CAMPBELL, R., WILLIAMS, S. C., MCGUIRE, P. K., ET AL. (1997). Activation of auditory cortex during silent lipreading. *Science*, **276**, 593-596.

CALVERT, G. A., SPENCE, C., & STEIN, B. E. (EDS.) (2004). *The handbook of multisensory processes*. Cambridge, MA: MIT Press.

CAMPBELL, R., DODD, B., & BURNHAM, D. (EDS.) (1998). *Hearing by eye II: Advances in the psychology of speechreading and auditory–visual speech*. Hove, U.K.: Psychology Press.

CONRAD, R. (1977). Lipreading by deaf and hearing children. *British Journal of Educational Psychology*, **47**, 60-65.

DESJARDINS, R. N., ROGERS, J., & WERKER, J. F. (1997). An exploration of why preschoolers perform differently than do adults in audiovisual speech perception tasks. *Journal of Experimental Child Psychology*, **66**, 85-110.

DESJARDINS, R. N., & WERKER, J. F. (2004). Is the integration of heard and seen speech mandatory for infants? *Developmental Psychobiology*, **45**, 187-203.

DODD, B. (1979). Lip reading in infants: Attention to speech presented in- and out-of-synchrony. *Cognitive Psychology*, **11**, 478-484.

DODD, B., & CAMPBELL, R. (EDS.) (1987). *Hearing by eye: The psychology of lip-reading*. London: Erlbaum.

DODD, B., & MURPHY, J. (1992). Visual thoughts. In R. Campbell (Ed.), *Mental lives: Case studies in cognition* (pp. 47-60). Oxford: Blackwell.

DUPOUX, E., PALLIER, C., SEBASTIÁN, N., & MEHLER, J. (1997). A destressing "deafness" in French? *Journal of Memory & Language*, **36**, 406-421.

DUPOUX, E., PEPERKAMP, S., & SEBASTIÁN-GALLÉS, N. (2001). A robust method to study stress "deafness." *Journal of the Acoustical Society of America*, **110**, 1606-1618.

FISHER, C. G. (1968). Confusions among visually perceived consonants. *Journal of Speech & Hearing Research*, **11**, 769-804.

FOWLER, C. A. (1996). Listeners do hear sounds, not tongues. *Journal of the Acoustical Society of America*, **99**, 1730-1741.

HADAR, U., STEINER, T. J., GRANT, E. C., & ROSE, F. C. (1983). Head movement correlates of juncture and stress at sentence level. *Language & Speech*, **26**, 117-129.

HADAR, U., STEINER, T. J., GRANT, E. C., & ROSE, F. C. (1984). The timing of shifts in head posture during conversation. *Human Movement Science*, **3**, 237-245.

HEIDER, F., & HEIDER, G. M. (1940). An experimental investigation of lipreading. *Psychological Monographs*, **52**, 124-153.

KAMACHI, M., HILL, H., LANDER, K., & VATIKIOTIS-BATESON, E. (2003). "Putting the face to the voice": Matching identity across modality. *Current Biology*, **13**, 1709-1714.

KUHL, P. K., & MELTZOFF, A. N. (1982). The bimodal perception of speech in infancy. *Science*, **218**, 1138-1141.

KUHL, P. K., & MELTZOFF, A. N. (1984). The intermodal representation of speech in infants. *Infant Behavior & Development*, **7**, 361-381.

LIBERMAN, A. M., & MATTINGLY, I. G. (1985). The motor theory of speech perception revised. *Cognition*, **21**, 1-36.

LYXELL, B., & RÖNNBERG, J. (1991). Word discrimination and chronological age related to sentence-based speech-reading skill. *British Journal of Audiology*, **25**, 3-10.

MACEACHERN, M. R. (2000). On the visual distinctiveness of words in the English lexicon. *Journal of Phonetics*, **28**, 367-376.

MACMILLAN, N. A., & CREELMAN, C. D. (1991). *Detection theory: A user's guide*. Cambridge: Cambridge University Press.

MASSARO, D. W. (1998). *Perceiving talking faces: From speech perception to a behavioral principle*. Cambridge, MA: MIT Press.

MASSARO, D. W., COHEN, M. M., & GESI, A. T. (1993). Long-term training, transfer, and retention in learning to lipread. *Perception & Psychophysics*, **53**, 549-562.

MATTYS, S. L., BERNSTEIN, L. E., & AUER, E. T., JR. (2002). Stimulus-based lexical distinctiveness as a general word-recognition mechanism. *Perception & Psychophysics*, **64**, 667-679.

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

MEHLER, J., JUSCZYK, P., LAMBERTZ, G., HALSTED, N., BERTONCINI, J., & AMIEL-TISON, C. (1988). A precursor of language acquisition in young infants. *Cognition*, **29**, 143-178.

MOGFORD, K. (1987). Lip-reading in the prelingually deaf. In B. Dodd & R. Campbell (Eds.), *Hearing by eye: The psychology of lip-reading* (pp. 191-211). London: Erlbaum.

MUNHALL, K., JONES, J. A., CALLAN, D. E., KURATATE, T., & VATIKIOTIS-BATESON, E. (2004). Head movement improves auditory speech perception. *Psychological Science*, **15**, 133-137.

NAVARRA, J., SEBASTIÁN-GALLÉS, N., & SOTO-FARACO, S. (2005). The perception of second language sounds in early bilinguals: New evidence from an implicit measure. *Journal of Experimental Psychology: Human Perception & Performance*, **31**, 912.

NAVARRA, J., & SOTO-FARACO, S. (2007). Hearing lips in a second language: Visual articulatory information enables the perception of second language sounds. *Psychological Research*, **71**, 4-12.

NAVARRA, J., SPENCE, C., & SOTO-FARACO, S. (2007). *Visual discrimination of rhythm in speech*. Manuscript submitted for publication.

NAZZI, T., BERTONCINI, J., & MEHLER, J. (1998). Language discrimination by newborns: Toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception & Performance*, **24**, 756-766.

NAZZI, T., JUSCZYK, P. W., & JOHNSON, E. K. (2000). Language discrimination by English-learning 5-month-olds: Effects of rhythm and familiarity. *Journal of Memory & Language*, **43**, 1-19.

NITCHIE, E. B. (1916). The use of homophenous words. *Volta Review*, **18**, 85-93.

OWENS, E., & BLAZEK, B. (1985). Visemes observed by hearing-impaired and normal-hearing adult viewers. *Journal of Speech & Hearing Research*, **28**, 381-393.

PATTERSON, M. L., & WERKER, J. F. (1999). Matching phonetic information in lips and voice is robust in 4.5-month-old infants. *Infant Behavior & Development*, **22**, 237-247.

PATTERSON, M. L., & WERKER, J. F. (2002). Infants' ability to match dynamic phonetic and gender information in the face and voice. *Journal of Experimental Child Psychology*, **81**, 93-115.

RAMUS, F. (2002). Language discrimination by newborns. *Annual Review of Language Acquisition*, **2**, 85-115.

RAMUS, F., HAUSER, M. D., MILLER, C., MORRIS, D., & MEHLER, J. (2000). Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science*, **288**, 349-351.

RAMUS, F., & MEHLER, J. (1999). Language identification with suprasegmental cues: A study based on speech resynthesis. *Journal of the Acoustical Society of America*, **105**, 512-521.

RAMUS, F., NESPOR, M., & MEHLER, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, **73**, 265-292.

REISBERG, D., MCLEAN, J., & GOLDFIELD, A. (1987). Easy to hear but hard to understand: A lip-reading advantage with intact auditory stimuli. In B. Dodd & R. Campbell (Eds.), *Hearing by eye: The psychology of lip-reading* (pp. 97-113). London: Erlbaum.

RÖNNBERG, J., SAMUELSSON, S., & LYXELL, B. (1998). Conceptual constraints in sentence-based lipreading in the hearing-impaired. In R. Campbell, B. Dodd, & D. Burnham (Eds.), *Hearing by eye II: Advances in the psychology of speechreading and auditory–visual speech* (pp. 143-153). Hove, U.K.: Psychology Press.

SACKS, O. (1990). *Seeing voices: A journey into the world of the deaf*. New York: HarperPerennial.

SAMS, M., AULANKO, R., HAMALAINEN, M., HARI, R., LOUNASMAA, O. V., LU, S. T., & SIMOLA, J. (1991). Seeing speech: Visual information from lip movements modifies activity in the human auditory cortex. *Neuroscience Letters*, **127**, 141-145.

SAMUELSSON, S., & RÖNNBERG, J. (1993). Implicit and explicit use of scripted constraints in lip-reading. *European Journal of Cognitive Psychology*, **5**, 201-233.

SEBASTIÁN-GALLÉS, N., DUPOUX, E., COSTA, A., & MEHLER, J. (2000).

Adaptation to time-compressed speech: Phonological determinants. *Perception & Psychophysics*, **62**, 834-842.

SEBASTIÁN-GALLÉS, N., ECHEVERRÍA, S., & BOSCH, L. (2005). The influence of initial exposure on lexical representation: Comparing early and simultaneous bilinguals. *Journal of Memory & Language*, **52**, 240-255.

SEBASTIÁN-GALLÉS, N., & SOTO-FARACO, S. (1999). Online processing of native and non-native phonemic contrasts in early bilinguals. *Cognition*, **72**, 111-123.

SOLÀ, J., LLORET, M. R., MASCARÓ, J., & PÉREZ SALDANYA, M. (2000). *Gramàtica del català contemporani: Volum 1. Introducció: Fonètica i fonologia. Morfologia* [Grammar of contemporary Catalan: Vol. 1. Introduction: Phonetics and phonology. Morphology]. Barcelona: Editorial Empúries.

SOTO-FARACO, S., NAVARRA, J., & ALSIUS, A. (2004). Assessing automaticity in audiovisual speech integration: Evidence from the speeded classification task. *Cognition*, **92**, B13-B23.

SOTO-FARACO, S., SEBASTIÁN-GALLÉS, N., & CUTLER, A. (2001). Segmental and suprasegmental mismatch in lexical access. *Journal of Memory & Language*, **45**, 412-432.

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). London: Erlbaum.

SUMMERFIELD, Q., MACLEOD, A., MCGRATH, M., & BROOKE, M. (1989). Lips, teeth, and the benefits of lipreading. In A. W. Young & H. D. Ellis (Eds.), *Handbook of research on face processing* (pp. 223-233). Amsterdam: North-Holland.

SUOMI, K., MCQUEEN, J. M., & CUTLER, A. (1997). Vowel harmony and speech segmentation in Finnish. *Journal of Memory & Language*, **36**, 422-444.

VATIKIOTIS-BATESON, E., MUNHALL, K. G., KASAHARA, Y., GARCIA, F., & YEHIA, H. (1996). Characterizing audiovisual information during speech. In *Proceedings of the 4th International Conference on Spoken Language Processing (ICSLP 96)* (Vol. 3, pp. 1485-1488). New York: IEEE Press. Available at www.asel.udel.edu/icslp/cdrom/vol3/1004/a1004.pdf.

WALDEN, B. E., ERDMAN, S. A., MONTGOMERY, A. A., SCHWARTZ, D. M., & PROSEK, R. A. (1981). Some effects of training on speech recognition by hearing-impaired adults. *Journal of Speech & Hearing Research*, **24**, 207-216.

WALDEN, B. E., PROSEK, R. A., MONTGOMERY, A. A., SCHERR, C. K., & JONES, C. J. (1977). Effects of training on the visual recognition of consonants. *Journal of Speech & Hearing Research*, **20**, 130-145.

WEIKUM, W. M., WERKER, J. F., VOULOUMANOS, A., NAVARRA-ORDOÑO, J., SOTO-FARACO, S., & SEBASTIÁN-GALLÉS, N. (2004, November). *When can infants discriminate languages using only visual speech information?* Poster presented at the 29th Boston University Conference on Child Development, Boston.

YEHIA, H. C., KURATATE, T., & VATIKIOTIS-BATESON, E. (2002). Linking facial animation, head motion and speech acoustics. *Journal of Phonetics*, **30**, 555-568.

ZATORRE, R. J. (2001). Do you see what I'm saying? Interactions between auditory and visual cortices in cochlear implant users. *Neuron*, **31**, 13-14.

NOTES

1. The participants' language dominance was classified according to the language spoken by their parents and/or caretakers at home when they were growing up. This type of classification is strongly correlated with early linguistic experience, a critical determinant of later phonological abilities such as, for example, the ability to discriminate between certain Catalan phonemes (e.g., /e/ and /ɛ/) that do not exist as separate sounds in Spanish (see Navarra et al., 2005; Sebastián-Gallés, Echeverría, & Bosch, 2005; Sebastián-Gallés & Soto-Faraco, 1999).

2. We detected an error in the distribution of the materials of the protocol used to test the English-speaking group whereby there were more trials of different languages than of the same language (always coinciding with the 22-syllable sentences). This did not help the participants to perform better by estimating the duration of the trial, as can be seen from the lack of differences in correct responses. In any case, to ensure that this could not have been the cause of their failure to discriminate, we decided to run a group of 42 Catalan–Spanish bilinguals (14 Spanish dominant, 15 Catalan dominant, and 13 simultaneous bilinguals) with the imbalanced materials. The results of this experiment closely replicated those found in Experiment 1—to be specific, no differences between the groups tested (both Fs < 1), a positive linear effect of number of syllables [$F_1(2,78) = 3.0$, $p = .056$; $F_2(2,45) = 3.4$, $p < .05$], and above-chance discrimination overall (58.5%, $p < .001$). The comparison with the English and Italian speakers of Experiment 2 confirmed, again, the superiority of the Spanish–Catalan bilinguals overall [$F(1,75) = 6.5$, $p < .05$].

3. In Experiment 1, there was a significant linear trend between number of syllables and performance [$F(1,45) = 5.4$, $p < .05$]; in Experiment 2, the fit to a linear equation was far from significant [$F < 1$].

4. There is still a chance that the speaker whose recordings were used gave away certain culturally determined cues (e.g., in the form of facial expressions), which only viewers familiar with the Catalan and Spanish languages/cultures could pick up in order to classify the sentences. Though far-fetched, this culturally based account must be considered a logical possibility at present.

5. Four of the participants were tested outside the laboratory, in a quiet dormitory room. Since the results of these participants did not differ from those of the others ($|t| < 1$), the data were pooled.

6. Although it was not the purpose of this study to measure response latencies, we were able to examine the total amount of time from trial onset until a response was made for each trial. Of course, we cannot compare across trials of different lengths, but we can compare RT differences across participant groups, since they were exposed to exactly the same sentences. The three bilingual linguistic groups tested in Experiment 1 did not show any reliable difference among themselves ($|t| < 1$). However, both the English and the Italian speakers in Experiment 2 (both ps $< .001$) and the Spanish monolingual group in Experiment 3 ($p < .05$) required significantly more time to respond than did the bilinguals tested in Experiment 1.

7. Further examination of the data distribution supports this conclusion: We found that the difference (3.5%) that we observed between the mean scores of the Spanish monolingual group and the Catalan–Spanish bilingual sample of Experiment 1 also appeared in the first (51% vs. 54%) and third (61% vs. 64%) quartiles of their distributions, as well as in their median values (55% vs. 59%). Moreover, neither of the two distributions contained outliers (± 3 *SD*s). The most extreme value of the Spanish monolingual sample was 69% (2.5 *SD*s above the mean), whereas that of the Catalan–Spanish bilingual sample was 81% (2.8 *SD*s above the mean).

8. We did not include the Catalan schwa in the Phon index because it is very frequent and would have biased the final scores significantly, obscuring the potential impact of the other vowels. Thus, we analyzed the schwa as part of a separate variable (vowel reduction).

9. In particular, we included nouns, verbs, adjectives, adverbs, personal pronouns, possessive pronouns, and demonstrative pronouns. We chose to include the pronoun categories for their potential relevance in lexically based strategies for discriminating one language from the other, since some of them are quite distinctive of language.

(Manuscript received March 24, 2005;
revision accepted for publication May 2, 2006.)