Visual speech interference in an auditory shadowing task: The dubbed movie effect

Jordi Navarra
SPPB Group (Universitat de Barcelona)
E-mail: jnavarra@psi.ub.es

ABSTRACT

Experiment 1 was designed to investigate if incongruent linguistic lip movements interfere with the perception of auditory sentences. We tested Catalan/Spanish bilinguals with Spanish as L1 using a "shadowing task". The relevant auditory sentences (degraded with white noise) were synchronized with a silent video recording of a speaker pronouncing other sentences (in Catalan or Spanish) or energetically chewing gum. Only Spanish lip movements produced significant interference (as compared with the non-linguistic condition), although, there were no differences between Spanish and Catalan. A second experiment was designed to see if Catalan/Spanish bilinguals can discriminate between these languages using only visual information. Catalan/Spanish bilinguals can do this type of discrimination above chance, whereas English and Italian monolinguals cannot. This result highlights the importance of linguistic experience and rules out the role of extra-linguistic factors.

INTRODUCTION

There is an anecdotic phenomenon that suggests the automaticity of speechreading during the perception of audiovisual linguistic material. When a bilingual spectator watches a movie originally spoken in one of his languages (L1 or L2), dubbed to his other language (L2 or L1, respectively); normally, the spectator has a subjective impression of mismatch between vision and audition. The present study investigates this phenomenon experimentally. We expected to find an interference effect due to incongruent visual information (when it is linguistic) in the comprehension of an audio message.

Behavioral [1,2] and fMRI studies [3] emphasize the relevance of visual information (extracted from articulatory movements) in speech perception. Sumby et al. [1] showed the benefit that is obtained from speechreading during the perception of words in a noisy context. Reisberg et al. [2] concluded, in different experiments, that visual information can help to understand unfamiliar languages, or messages pronounced with a foreign accent. In an fMRI study, presenting bimodal congruent speech stimuli, Calvert et al. [3] observed supra-additive activity enhancements in the ventral bank of the superior temporal sulcus (BA 22/21). This study also revealed activity depression in this area for incongruent bimodal stimuli [3]. Surprisingly, Reisberg [2] did not find any interference effect during the perception of spoken messages with audio-visual incongruent speech (in fact, asynchronies of 500 ms), see also [4].

It seems, as Calvert et al. [3] showed, that there are some inhibitory cortical activation effects when audio-visual incongruent speech is presented, although the existing behavioral data regarding these effects cannot be considered conclusive. The goal of experiment 1 was to study if incongruent linguistic lip-movements have more influence in speech perception than non linguistic ones; in this way, experiment 1 tries to reproduce the cited anecdotic phenomenon (the dubbed movie effect).

In the experiments presented here, we used sentences, because this kind of linguistic material includes several types of information as, for example, rhythm. The characteristics and relevance of this kind of information has been highlighted by Ramus et al. [5].

We also addressed the language-specificity of this potential interference effect. Some experimental studies have shown that the audiovisual perception of certain phonetic (and visemic) categories within a language are influenced by the degree of experience within that language [6-9]. Accordingly, we would expect that a Spanish/Catalan bilingual with a clear dominance for Spanish will process more accurately the visual information in his/her mother tongue than in the non dominant language, in this case Catalan.

In spite of the similarity between Spanish and Catalan, these languages present some phonological differences, such as a smaller vowel repertoire in Spanish and some relevant consonant particularities. Other aspects related to rhythm (the so called *vocalic reduction* and the existence of consonantal clusters at the end of the word, features present only in Catalan) provide additional examples of the differences between these languages.

We can hypothesize that (1) some of the characteristics that differentiate Catalan and Spanish can be manifested visually, and (2) linguistic experience can influence the perception of visual speech in the first language (L1) and in the second language (L2) differentially.

ISBN 1-876346-48-5

EXPERIMENT 1

The main goal of this experiment was to investigate if, just as the literature seems to indicate [10-12], our perceptive system is not able to obviate visual linguistic information from lip-movements, even when it is incongruent with the auditory information. In addition, the inclusion of two different linguistic conditions (Catalan and Spanish) will permit us to see if the degree of experience with the language presented visually influences its level of interference on auditory speech perception.

Method

Participants

Twelve Catalan/Spanish bilinguals with dominance for Spanish (students at the University of Barcelona) were tested.

Stimuli

We used 108 video-clips of a balanced bilingual female speaker pronouncing sentences (36 Catalan, 36 Spanish) or chewing gum energetically (36 clips). A further 108 un-related audio sentences (all of them in Spanish) were dubbed with the visual clips. The visual and the auditory sentences were matched in number of syllables (or duration in the case of chewing gum). Sentences of 16, 22 and 32 syllables were used.

Procedure

The 108 dubbed clips were presented on a monitor screen and the audio stimuli appeared in two speakers located at both sides of the screen. Each of the three different blocks, ordered according to a Latin square, contained a different visual condition (visual-Catalan, visual-Spanish or visual non-linguistic).

The participants performed a *shadowing task* on the auditory sentences (repeating back the sentences, all of them in Spanish; the subject's L1). Finally, each block included an additional 33% catch trials (randomly interspersed) in which the speaker seen on the screen stopped moving the lips at some unexpected instant, while the acoustic sentence continued. Participants had to detect these trials by pressing a button. This secondary task served to ensure that the participants were looking at the speaker's face. These trials were not analyzed. All the auditory stimuli were masked with white noise. The white noise intensity level was regulated for every participant individually, in order to adjust it at a level in which participants shadowed about 80% of the words correctly. Participants received prior training without visual stimuli to be accustomed to the task. After the noise level adjustment, a second training with audiovisual stimuli was carried out to familiarize the participants with the materials.

Results

The number of syllables incorrectly shadowed or omitted by the participants was counted for every participant and condition (Figure 1). Participants made an average of 184.2 errors when the visual sentences were in Spanish, 168.1 errors when the sentences were in Catalan and 160.3 errors when the speaker was chewing gum. An ANOVA including the factor *visual condition* (VSp, VCat and VNonL) reached significance ($F(1, 11) = 5.99$, $p < .05$).

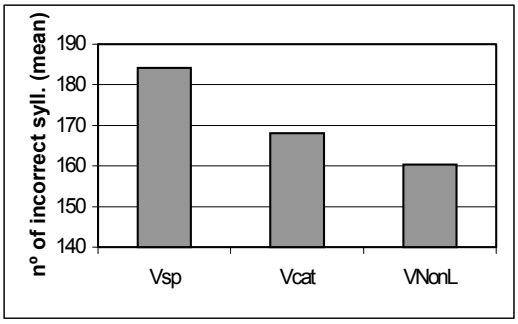

Figure 1: Mean % of syllables incorrectly shadowed or omitted by the participants in each condition. VSp refers to visual-Spanish, VCat to visual-Catalan and VNonL to visual not linguistic.

In a pair-wise comparison, a significant difference between the conditions VSp and VNonL ($t(11) = 2.45$, $p < .05$) was observed. This analysis also revealed that there were no significant differences between VNonL and VCat ($|t| < 1$, $p = .506$), and a marginally significant difference between VCat and VSp ($t(11) = 1.69$, $p = .079$). Independently of the condition, the effect of block order was marginally significant ($F(1, 11) = 3.34$, $p = .095$), indicating a potential practice effect.

Discussion

It is possible to conclude that the linguistic information of incongruent lip-movements in L1 produces a greater interference than the non linguistic information in auditory speech perception.

These results seem to indicate that it is difficult to obviate the linguistic lip-movements in complex material (sentences), even when an evident audiovisual mismatch exists. These results point to the same direction as the McGurk effect [10] supporting the hypothesis that the visual processing of speech is automatic.

The effect presented here, using sentences in distinct languages, shows that it is possible to study how some characteristics of languages (the rhythm, at the *supra-segmental* level, or the visemes, at the *segmental* level, for example) influence the audiovisual integration of speech. The results found in this experiment are certainly related to the dubbed movie effect previously cited. From our data it seems that L2 interferes to a lesser degree than L1, however, the language specificity of this interference is not clear cut.

EXPERIMENT 2

Experiment 1 did not show significant differences between the two visual linguistic conditions (VSp versus VCat). It is

not clear whether the linguistic dominance in Spanish that the subjects manifest affects the visual interference effect. There are two possible explanations to consider. On the one hand, Spanish dominant bilinguals generally show a high proficiency in Catalan. On the other hand, Catalan and Spanish are very similar in visual aspects of the speech. The main objective of experiment 2 was to test this second possibility by checking if Catalan/Spanish bilinguals can discriminate between Catalan and Spanish visually. This experiment also included a small sample of English and Italian monolingual participants to check the relevance of experience with these languages in the discrimination task.

Method

Participants

Fifty seven subjects participated in this experiment, 42 undergraduates from the University of Barcelona (Spain), 10 students from the University British Columbia (Canada) and 5 students from distinct universities of Rome (Italy). All the Spanish participants were Catalan/Spanish bilinguals, and they were divided in three groups. One group was formed by 13 balanced bilinguals, that is to say, bilinguals who one parent spoke to them in Catalan (in 7 of them was the mother and, in the other 6, the father) while the other parent spoke Spanish. A second group was formed by 16 bilinguals with a clear dominance for Catalan. A third group, formed by 13 bilinguals with a clear dominance for Spanish. Finally, we included 10 English monolinguals and 5 Italian monolinguals.

Stimuli

Additional sentences were added to the materials of experiment 1 to make a total of 40 sentences pronounced in each language. They consisted of 80 silent video-clips of the bilingual speaker; 40 pronouncing Catalan sentences and 40 pronouncing Spanish sentences.

Procedure

Each participant received 40 trials. Each trial started with the speaker's image on the screen, inside a red frame, pronouncing a sentence in Catalan or Spanish. Then, the screen faded into black for 1 second. The speaker then reappeared, now inside a green frame, saying another sentence in either language. The two sentences of each trial had an equivalent length (16, 22 or 32 syllables). Each potential combination (Cat-Sp, Sp-Sp, Sp-Cat or Cat-Cat) was equiprobable and the two sentences paired within a trial were never the same.

The task consisted of pressing one of two buttons depending on whether s/he judged that the two sentences were in the same or distinct languages. The instructions said explicitly that the participants should try to press the button while the speaker was pronouncing the second sentence. Thus the green frame signaled the period in which the answer should be emitted. The percentage of correct responses was counted for each participant.

Results

Catalan/Spanish bilinguals obtained a 58.4% of correct responses (see figure 2), the English monolinguals 50.9% and, finally, the Italian monolinguals 51%.

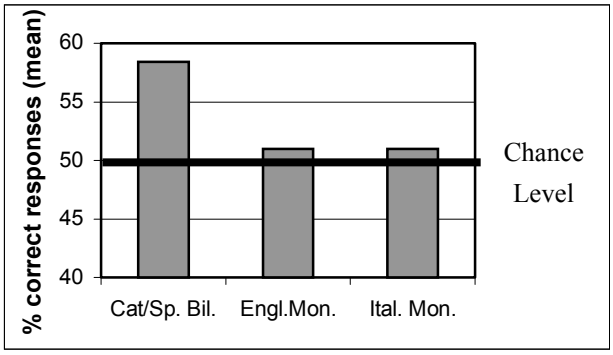

Figure 2: Proportion of correct discriminations in Experiment 2 for the different language groups tested: Spanish/Catalan bilinguals (Cat/Sp. Bil.), English mono- linguals (Engl. Mon.) and Italian monolinguals (Ital. Mon).

A Student-T revealed that the 58.4% of correct responses observed in the Catalan/Spanish bilinguals was significantly above chance (by participants, $t1(41) = 6.97$, $p <. 001$; by items, $t2(79) = 5.833$, $p <. 001$). Subsequently the distinct Catalan/Spanish bilingual groups were compared. This test did not show significant differences between them ($F <1$). Another two comparisons showed that the English monolingual's correct responses were not significantly distinct from chance ($t(9) = 1.70$, $p =. 115$), nor in the case of the Italian monolinguals ($t(4) =. 18$, $p =. 863$). Finally, the last analyses revealed that the Catalan/Spanish bilinguals and the English monolinguals differed significantly ($t(50) = 2.56$, $p <. 05$), and that the Catalan/Spanish bilinguals and the Italian monolinguals differed, but only marginally ($t(45) = 1.87$, $p =. 069$), maybe because the sample composed by Italian monolinguals was too small. There were no significant differences between the English and Italian groups ($|t| <1$).

Discussion

The goal of Experiment 2 was to verify if the Catalan/Spanish bilinguals are capable to differentiate between these languages using visual cues alone. The results suggest that these participants can establish this distinction, although the task seems to be difficult. The analysis by items suggests that the correct responses are not accumulated in specific stimuli. These data also showed that neither English nor Italian monolinguals are capable of distinguishing between the two languages (despite the similarities between Italian, Catalan and Spanish).

It can be concluded, therefore, that the Catalan/Spanish bilinguals are sensitive to the visual differences between their two languages. The fact that the English and Italian monolinguals cannot do the distinction ensures that the Catalan/Spanish bilinguals used linguistic information of

ISBN 1-876346-48-5

the articulatory movements and not another kind of extra-linguistic information.

It is important to highlight that this ability to perceive the visual particularities of Catalan and Spanish when attention is focused on speech-reading (i.e. Experiment 2) does not necessarily imply that this ability will be at play when attention is focused on another task (or another type of signal, like the auditory-linguistic one, as in Experiment 1).

CONCLUSIONS

Complex linguistic materials (sentences) were used in the experiments presented here. Previous studies seem to indicate that congruent visual information can facilitate the auditory comprehension of words in a noisy context [1] and the comprehension of sentences in a non-familiar second language [2].

The goal of experiment 1 was to explore if it is possible to find an interference of incongruent visual linguistic information in auditory speech perception. The results of this experiment revealed that the linguistic visual stimuli produced a greater interference than the non-linguistic stimuli. These data aim in the same direction as other previous empirical results, showing the relevance of the visual information in the perception of speech and its automatic processing [10]. Future research can study a possible relation between our behavioral data and the fMRI evidence that showed sub-additive response in STS during the audio-visually incongruent speech perception.

Experiment 1 did not reveal significant differences between the condition in which the visual information was in Catalan (L2) and in Spanish (L1), although, we observed a tendency to make more mistakes when the "visual" language was L1. These results raised some questions: are Spanish and Catalan visually different? What is the role of the experience with a language in its visual perception?

Experiment 2 tried to answer these questions. The results showed that a very heterogeneous sample of Catalan/ Spanish bilinguals can do this distinction, while a relatively small sample of English and Italian monolinguals cannot, especially in the case of the English monolinguals. First, these results suggest that there are some visual cues that these bilinguals can use to differentiate Catalan and Spanish. And, second, the experience and knowledge of these languages seems to be crucial for carrying out the distinction.

One interesting question for future research is what is the specific role of finer phonological elements such as rhythm and visemic cues in the dubbed movie interference effect.

ACKNOWLEDGMENTS

The research reported here was supported by a grant from the James S.McDonnell Foundation JSMF-20002079, and the Catalan Government Research Grant SGR00034.

REFERENCES

[1] Sumby, W. & Pollack, I. "Visual contribution to speech intelligibility in noise", *Journal of the Acoustical Society of America*, **26**, pp. 212-215, 1954.

[2] Reisberg, D., McLean, J. & Goldfield, A. "Easy to hear but hard to understand: A lip-reading advantage with intact auditory stimuli". In B. Dodd and R. Campbell (eds.). *Hearing by eye: The psychology of lip-reading*. Hillsdale: LEA, pp. 97-111, 1987.

[3] Calvert, G. A., Campbell, R. & Brammer, M. J. "Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex", *Current Biology,* **10(11)**, pp. 649-657, 2000.

[4] Grant, K. W., Walden, B. E. & Seitz, P. F. "Auditory-visual speech recognition by hearing impaired subjects: Consonant recognition, sentence recognition, and auditory-visual integration", *Journal of Acoustical Society of America*, **103(5)**, PP. 2677-2690, 1998.

[5] Ramus, F., Nespor, M. & Mehler, J. "Correlates of linguistic rhythm in the speech signal", *Cognition*, **73**, pp. 265-292, 1999.

[6] Werker, J., Frost, P. & McGurk, H. "Cross-language influences on bimodal speech perception", *Canadian Journal of Psychology,* **46 (4)**, pp. 551-568, 1992.

[7] Sekiyama, K. & Tohkura, Y. "Inter-language differences in the influence of visual cues in speech perception", *Journal-of-Phonetics,* **21 (4)**, pp. 427-444, 1993.

[8] Massaro, D. W., Cohen, M. M., Gesi, A. & Heredia, R. "Bimodal speech perception: an examination across languages", *Journal of Phonetics,* **21**, pp. 445-478, 1993.

[9] Hardison, D. M., "Bimodal speech perception by native and nonnative speakers of English: Factors influencing the McGurk effect", *Language Learning*, **46 (1)**,pp. 3-73, 1996.

[10] McGurk, H. & MacDonald, J., "Hearing lips and seeing voices", *Nature,* **265**, pp. 746-748, 1976.

[11] Dodd, B. & Campbell (Eds.). Hearing by Eye: The Psychology of Lip-Reading. London: LEA, 1987.

[12] Campbell, R., Dodd, B., & Burnham, D. (Eds.). Hearing by Eye II: Advances in the Psychology of Speechreading and Auditory-visual Speech. East Sussex: Psychology Press Ltd, 1998.

[13] Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C. R., McGuire, P. K., Woodruff, P. R., Iversen, S. D. & David, A. S., **"**Activation of auditory cortex during silent lip-reading", *Science,* **276 (5312)**, pp. 593-596, 1997.