The Role of Visual Stimuli in the Perception of Prosody in Brazilian Portuguese

Daniel Oliveira Peres, Beatriz Raposo de Medeiros, Waldemar Ferreira Netto, and Maria de Fátima de Almeida Baia
Universidade de São Paulo

1. Introduction

This experimental pilot study aims to analyze the role of visual stimuli in the recognition of prosodic characteristics in the discrimination task between statements and yes-no questions in Brazilian Portuguese (BP). Studies conducted by Massaro (1998), Abelin (2007), Fagel (2006) and Ronquest *et al.* (2010) show that there is strong evidence for bimodality in speech perception. The same pattern is supported by Barbosa *et al.* (2002) in a study on the relationship between the acoustic signal of speech and facial movements, and by Ronquest *et al.* (2010), who investigate in a visual-only experiment whether participants are able to perceive different languages through visual stimuli

We analyze intonation (in which any manifestation of prosodic variations as F0, time and amplitude are included) and its modal function (Fonagy, 2003) encompassing statements and questions. This study follows a strict approach which states that intonation can be understood as the melodic variation that occurs during speech (Hirst and Di Cristo, 1998). The choice is based on the fact that modal prosodic variations in BP permit identification without any help from other indices such as morphemes or syntactic inversion (Moraes, 1998). In other words, in BP yes-no questions and statements differ only in prosody, while in other languages grammatical changes are needed to transform statements into yes-no questions.

To verify the role of each modality, visual and auditory, in the perception of statements and yes-no questions in BP, we carried out two experiments. The first was based on the study by McGurk and MacDonald (1976) that consists of switching segments of audio-visual language. In a conventional McGurk test, people usually say they hear the syllable /na/, which was the result of two overlapped syllables: /da/ visual stimulus, and /ma/ acoustic stimulus. In our experiment, only pitch was changed and segments were left intact. It is important to mention that we do not intend to verify if McGurk effect can be observed when the issue is related to prosody. The aim is to know what is more relevant when people need to judge a sentence with visual and acoustic stimuli switched. The second experiment consisted of showing the participants only visual stimuli in order to verify if they were able to recognise statements and yes-no questions from purely visual modality.

We intended to answer the following questions with this brief study:

(i) What is the contribution of each of the senses involved (sight and hearing) in recognition of prosodic characteristics?
(ii) Can the informants, from only visual stimulus, perceive modal prosodic variations?

* We are grateful to Erik Willis, Rebecca Ronquest, Ian Maddieson, Jeffrey Steely, the audience at LARP, and to two anonymous reviewers for their comments and suggestions. The remaining mistakes are our own. Contact: first author danielperes@usp.br .

2. Methodology

2.1. Preparation of the stimuli

The first step was to capture 16 sentences (8 statements and 8 yes-no questions) with a digital camera (Kodak C743 - 7.1 mega pixels). One of the authors of this study recorded himself speaking the sentences in an environment properly illuminated to avoid problems regarding the display of facial expressions. The sentences used in the experiment are in the following table:

1	O João estava contente ontem. (?)	João was happy yesterday. *or* Was João happy yesterday?
2	A Maria chegou hoje. (?)	Maria arrived today. *or* Did Maria arrive today?
3	O Antonio estava triste ontem. (?)	Antonio was sad yesterday. *or* Was Antonio sad yesterday?
4	O Pedro estava cansado ontem. (?)	Pedro was tired yesterday. *or* Was Pedro tired yesterday?
5	O João estava bravo de manhã. (?)	João was angry in the morning. *or* Was João angry in the morning?
6	A Ana foi a escola cedo. (?)	Ana went to school early. *or* Did Ana go to school early?
7	O Rodrigo chegou hoje. (?)	Rodrigo arrived today. *or* Did Rodrigo arrive today?
8	O Paulo estava alegre ontem. (?)	Paulo was happy yesterday *or* Was Paulo happy yesterday?

Table 1: Sentences used in the experiment.

The second step was to extract one track in *mp3* format from each sentence for F0 manipulation. It is important to mention that the choice of audio file format does not affect the features of the original audio recorded. The bit rate was 320 kbps and the sampling rate was 44kHz. The script MoMEL (*Melodic Modelisation*) (Hirst, 2007) was used for the extraction. This script is an automatic stylisation of fundamental frequency. The extraction was performed in Praat (Boersma and Weenink, 1996) and.

We manipulated 8 statements in order to sound like yes-no questions in BP. Similarly, 8 yes-no questions were manipulated to sound like statements. It was necessary to manipulate both sentence types in order to make the lip synchronisation of audio and video switched and to preserve the time of the original sentences. The software *VirtualDubMod*[1] was used for editing images and audio in order to avoid audio and video mismatch.

Two steps were taken in the methodology to manipulate sentences. The first step was the adequacy of melodic contour based on studies by Moraes (1993, 1998). The main characteristic of yes-no questions in BP is higher F0 value when compared to F0 value of statements. In general, in yes-no questions there is an elevation of F0 on the last stressed syllable, the initial tone is slightly higher than in statements, and the unstressed final syllable is lower than the same syllable in statements. Below is an example of statement into yes-no question manipulation:

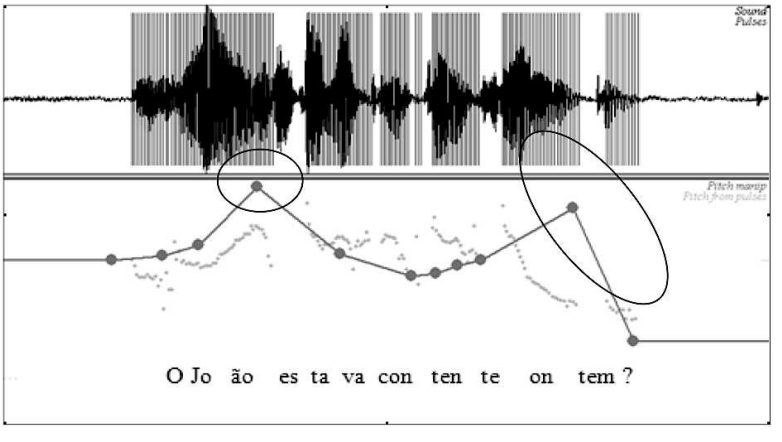
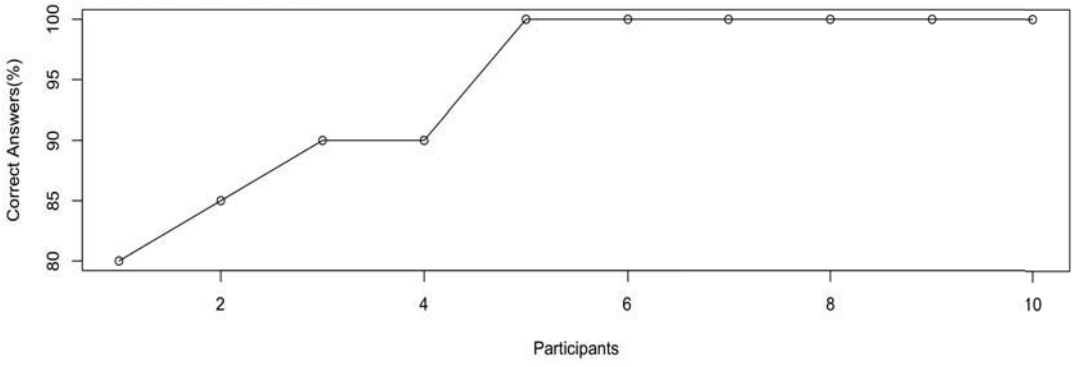


Figure 1 - Illustration of a statement sentence manipulated into a yes-no question – *Was João happy yesterday?*

[1] *VirtualDubMod* is free software: http://virtualdubmod.sourceforge.net.

The second step taken in order to check if sentences would be judged as statements or yes-no questions was the application of an F0 manipulation test to 10 participants, in which each listener had to judge 20 manipulated sentences presented in a random order. The participants were 5 women and 5 men, aged between 20 and 30, and non-linguists.

In the total of 200 judgments given by the listeners, there were only 10 errors, which consisted of 9 inversions and one instance where the listener was in doubt about the modality heard. This result indicates that the F0 manipulations sounded natural. The following graph illustrates the satisfactory performance at the discrimination task:



Graph 1: Percentage of correct answers per participant at the discrimination task.

2.2. Experiment with audio and video switched

After the satisfactory results obtained in the F0 manipulation test, another experiment was conducted with commutations between acoustic and visual stimuli. The test was applied to the same 10 informants who watched 20 videos with audio and video switched. The 20 sentences were composed of 8 statements and 8 yes-no questions with the audio and video switched. The other 4 sentences were used as distraction stimuli without switching audio and video (2 of each modality). The combinations of sentences can be seen in Table 2:

	Audio	**Vídeo**
Sentences	**Types**	**Types**
8	S	Q
8	Q	S
2	S	S
2	Q	Q

Table 2: Representation of acoustic and visual stimuli combinations (S - statements and Q - questions).

The videos were presented within an interval of 7 seconds, so that the informants would have time to write down the answer. The participants had to complete a form that had the sentence number and space to write 'S' for statements and 'Q' for yes-no questions.

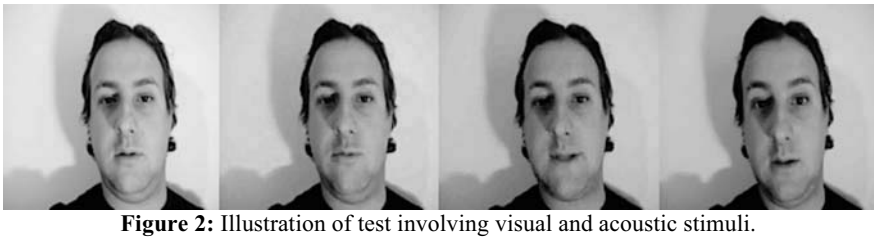


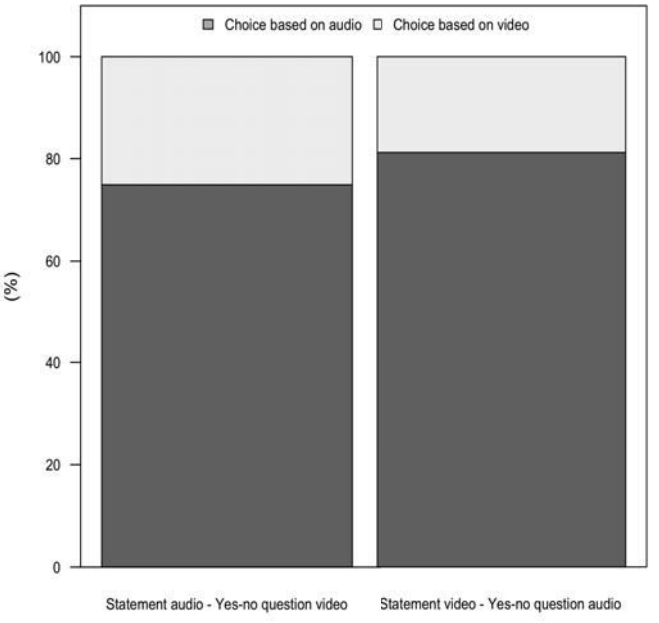
Figure 2: Illustration of test involving visual and acoustic stimuli.

2.3. Visual-only experiment

Twenty sentences were presented in a random order without acoustic stimulus. Participants were asked to judge if a statement or yes-no question was being produced. As for the other experiment, the participants had to complete a form that had the sentence number and space to write 'S' for statements and 'Q' for yes-no questions.

3. Results and Discussion

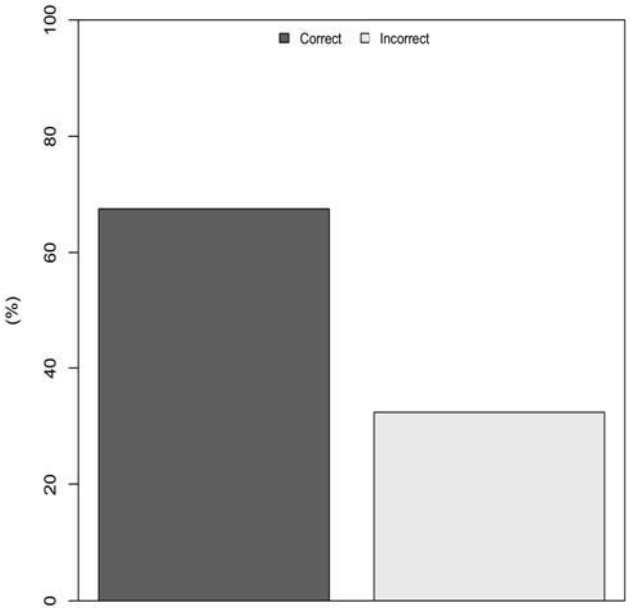
The results of the experiment with video and audio switched showed that the acoustic signal prevailed in the participants' choice. In 80 sentences that had statement acoustic stimulus and yes-no question visual stimulus, 20 (25%) answers were guided by the latter and 60 (75%) by the former. In 80 sentences that had statement visual stimulus and yes-no question acoustic stimulus, 15 (18.8%) answers were guided by the former and 65 (81.2%) by the latter. The 40 judgments of sentences without audio and video switched were not taken into account. Graph 2 represents the distribution of choices based on sentence type and visual and audio stimuli:



Graph 2: Perceptual test with audio and video switched.

A two-way ANOVA test was performed involving participants` choice based on audio or video stimuli. The results point to a non-random choice in favour of acoustic stimulus: $F(2, 17) = 6.94$, $p < .05$. The type of sentence (statement or yes-no question) did not have a significant influence on participants' choice: $F(4, 37) = 1.04$, $p > .05$

In the visual-only experiment, there were 65 (32.5%) errors in 200 answers. The number of correct answers (135, 67.5%) points to a satisfactory recognition of facial configurations. Graph 3 shows the distribution of answers in this experiment:



Graph 3*:* Distribution of answers in the visual-only experiment.

A two-way ANOVA test was performed involving the correct and incorrect answers of participants $F(2, 17) = 7.95$, $p < .05$ and its results demonstrate that the choice in favour of visual stimuli was not random. As we found in the experiment with video and audio switched, the type of sentence (statement or yes-no question) did not have significant influence on participants' choice $F(4, 38) = 0.05$, $p > .05$.

After consideration of the results, the initial questions can be answered:

(i) What is the contribution of each of the senses involved (sight and hearing) in recognition of prosodic characteristics?

The results of the first experiment (McGurk) showed that acoustic stimuli are the more important for the recognition of yes-no questions or statements. The statistical results also demonstrate that type of sentence had no influence on participants' choice.

(ii) Can the informants, from only visual stimulus, perceive modal prosodic variations?

In the experiment without audio, the results pointed to a satisfactory recognition of visual stimulus without sound correspondence.

In sum, the results of the experiments present evidence for bimodality in speech perception. Although the experiment with audio and video switched showed that acoustic stimuli played a more important role, we cannot affirm that this is the crucial stimulus type for speech perception. The visual-only experiment showed an important role is played by visual stimuli as well, in the absence of sound, but again we could not affirm that they are the essential stimulus type for speech perception. We believe that speech perception is bimodal as other authors have proposed (cf. Massaro 1998, Abelin 2007, Fagel 2006 and Ronquest *et al.* 2010) and when one modality is absent, the other is able to restore information somehow. However, it is important to remember that when video and audio are switched, the use of acoustic stimuli prevails.

4. Further Work

Although satisfactory results were obtained by conducting the experiments, this study only discussed and analyzed data regarding the modal function proposed by Fonagy (2003); also, the number of participants who judged the sentences was small. Therefore, there is a need to extend this experimental study to other modalities, as well as to increase the number of participants.

References

Abelin, Åsa (2007). Emotional McGurk effect – an experiment. *Proceedings of the Seventh International Conference on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems*. Lund University Cognitive Studies, 135.

Barbosa, Adriano V., Vatikiotis-Bateson, Eric and Yehia, Hani C. (2002). *Modeling the relation between speech acoustics and 2D facial motion*. Eurasip. Submitted.

Boersma, Paul and Weenink, David (2011). Praat: doing phonetics by computer [Computer program]. Version 5.1.10, retrieved 22 November 2009 from http://www.praat.org/.

Fagel, Sascha. 2006. "Emotional McGurk effect". *Speech Prosody 2006*, paper 006.

Fònagy, Ivan. 2003. "Des fonctions de l'intonation: Essai de synthèse". *Flambeau*. 29:1-20.

Hirst, Daniel and Di Cristo, Albert., eds. 1998. *Intonation Systems*. Cambridge University Press.

Hirst, Daniel. 2007. A Praat plugin for the Momel and INTSINT with improved algorithms for modeling and coding intonation. *Proceedings of ICPhS.*

McGurk, Harry and MacDonald, John. 1976. "Hearing lips and seeing voices." *Nature*. 264(5588):746-748.

Moraes, João A. 1993. "A Entoação Modal Brasileira: Fonética e Fonologia." *Caderno de Estudos Linguísticos*. 25:101-11.

Moraes, João A. 1998. "Intonation in Brazilian Portuguese." Pp. 179-194 in *Intonation Systems,* edited by Daniel Hirst and Albert Di Cristo. Cambridge: Cambridge University Press.

Ohala, John J. 1996. "Speech Perception is Hearing Sounds, not Tongues". *Journal of the Acoustical Society of America*. 99:1718-1725.

Ronquest, Rebecca. E., Levi, Susannah. V. and Pisoni, David. B. 2010. "Language Identification from Visual-only Speech Signals". *Attention, Perception & Psychophysics*. 72:1601-1613.

Vaissière, Jacqueline. 2004. "Perception of Intonation". Pp. 236-263 in *Handbook of Speech Perception,* edited by David. B. Pisoni and Robert. E. Remez . Oxford: Blackwell.

Selected Proceedings of the 5th Conference on Laboratory Approaches to Romance Phonology

edited by Scott M. Alvord

Cascadilla Proceedings Project Somerville, MA 2011

Copyright information

ISBN 978-1-57473-449-2 library binding

Ordering information

Orders for the library binding edition are handled by Cascadilla Press.
To place an order, go to www.lingref.com or contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, fax: 1-617-776-2271, sales@cascadilla.com

Web access and citation information

This entire proceedings can also be viewed on the web at www.lingref.com. Each paper has a unique document # which can be added to citations to facilitate access. The document # should not replace the full citation.

This paper can be cited as:

Peres, Daniel Oliveira, Beatriz Raposo de Medeiros, Waldemar Ferreira Netto, and Maria de Fátima de Almeida Baia. 2011. The Role of Visual Stimuli in the Perception of Prosody in Brazilian Portuguese. In *Selected Proceedings of the 5th Conference on Laboratory Approaches to Romance Phonology*, ed. Scott M. Alvord, 136-141. Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #2642.