Are eyebrow movements linked to voice variations and turn-taking? An experimental investigation

Isabelle Guaïtella, Serge Santi, Christian Cavé

Links between Eyebrow Movements and Voice Variations : an Experimental Investigation

Isabelle Guaïtella, Serge Santi, and Christian Cavé

Laboratoire Parole et Langage, UMR CNRS 6057, Université de Provence, Aix-en-Provence, France
&
Association GeVoix

Introduction

Taking a multimodal approach to communication, we experimentally test some of our hypotheses regarding the links between rapid eyebrow movements and vocalization during speech production in dialogue situations. In particular, we examine the relationship between eyebrow movements and variations in the fundamental frequency of the voice. We have already found event simultaneity and resemblance of form in the eyebrow-movement and fundamental-frequency curves (Cavé, Guaïtella, & Santi, 1993; Cavé, Guaïtella, Bertrand, Santi, Harlay, & Espesser,1996). We use an experimental device that supplies objective gestural data via an automatic movement acquisition and analysis system (Elite). Once the eyebrow data is plotted on curves, it can be related to and compared with the fundamental frequency curves obtained using sound signal processing software.

1. Theoretical background

Darwin (1874) described the gestures that contribute to the expression of emotions as resulting from the interaction between physiological constraints and functional needs. For example, an activity like raising the eyebrows to enlarge the visual field and see as far away as possible will be preserved -- through the influence of the so-called association principle -- as a means of expressing surprise and astonishment, and will thus become a mark of special attention given to what is happening. Scientists like Ekman and Friesen (1975) have studied facial cues to gain insight into the emotional aspects of communication, and Ekman (1989) in particular attempted to characterize rapid movements, of which eyebrow movements are a part. Such rapid movements are thought to be mainly linked to discourse production, where they serve as a rhythmic demarcation device (see also Birdwhistell, 1970). For many authors, the difference between rapid and slow movements is a categorical one, for rapid movements are said to be rhythmic, while slow ones are viewed as semiotic. We have criticized this point of view (Guaïtella, 1991, 1995), and contend that both rapid and slow movements contribute not only to rhythm, but to the semiotic production as well. Rapid eyebrow movements thus have a semiotic component that is rooted -- probably by association -- in the emotional expressions of the face. However, the function of these rapid movements is essentially discourse-based, for their use in the expression of emotions has come to take on a linguistic function (an eyebrow movement occurring on an unpredictable word, for example, can be interpreted as "Pay special attention to the word I'm pronouncing" ...). For Bolinger (1998),the eyebrows even more than other gestural parameters are related in a privileged way to speech, and benefit from a status equivalent to vocality.

2. Experimental procedure

2.1 Corpus

Ten unpaid subjects volunteered to participate in the experiment. They were questioned about their work by means of interviews with a highly conversational style. Thus, the corpus consisted of ten dialogues between an interviewer and an interviewee, each lasting approximately five minutes. For technical reasons, we were only able to process seven of the ten dialogues.

2.2 Data acquisition

For each subject, three types of recordings were obtained: (1)the speech signal, (2) the video image, and (3) head and face movement data. Subjects were filmed on videotape for the full duration of the experimental session. The subject's speech was recorded using a DAT recorder and a Sennheiser microphone (type ME 40). The techniques used to record the head and face movements are described below.

2.3 Kinematic data acquisition and processing

Head and eyebrow movements were recorded using a motion analysis system with a 100-Hz automatic optical TV-image processor (Elite system) enabling computer reconstruction of three-dimensional trajectories of small infrared retro-reflective markers (Ferrigno and Pedotti, 1985).The kinematic properties of five key points were captured by means of hemispherical markers (5 mm in diameter) attached to the subject's skin. The marker for recording head movements was placed on the forehead at the upper extremity of the frontal suture. For scowl detection, a marker was placed at each of the inner ends of the left and right supra-orbital arches. Markers for analyzing eyebrow raising and lowering were placed at the outer ends of the left and right supra-orbital arches. For the last two markers, a lateral position was chosen to avoid confusion between the inner and outer markers during head rotation. An external reference marker was placed on a vertical rod behind the subject.

Two video cameras equipped with wide-angle lenses were positioned in front of the subject parallel to his/her frontal plane at a distance of 1.5 m. The sight angles of the two cameras in the horizontal plane were -35°and +35°, respectively .

The Elite system detected the markers in real-time by means of a hardware shape-recognition procedure. It then calculated their centroid x-y coordinates and generated a 3-D reconstruction of the marker trajectories. The powerful algorithms used for the centroid calculation and the stereometric procedure allowed us to attained a measurement accuracy of about 1/3000, i.e., 0.5 mm for the linear displacements in our experimental conditions. After digital filtering based on the preliminary automatic selection of the appropriate bandwidth for each signal (D'Amico and Ferrigno, 1990), the first and second derivatives of the markers' linear displacements were computed.

2.4 Data analysis

The kinematic data output by the Elite system (x, y, and z coordinates of the eyebrow movements) and the speech signal were transferred to a Sparc SUN station in a Unix environment. An automatic procedure was used to display the spectrogram of the speech signal, the corresponding Fo curve, and the eyebrow movement curve. In order to separate the real movements of the eyebrows from of those resulting from the movements of the head (which also sent data to the eyebrow sensors), the head movement component was factored out by an automatic procedure. The entire output data set was verified a posteriori by viewing of the video documents.

3. Hypotheses

Our first hypothesis regarding the role of the eyebrows was related to the part they might play in managing the conversational interaction, i.e., in turn-taking. It was assumed that speakers would use this device to signal their desire to say something and thus, to be given a speaking turn. It appeared valid to test this hypothesis by measuring the interval between eyebrow movements and the beginning and ending of the speaking turns. Eyebrow movements that fell closer to the onset of a speaking turn than to its offset would lend support this hypothesis.

Our second hypothesis concerned the role of the eyebrows in rhythmic demarcation. During speech production, a speaker may accentuate a syllable(of an important word, for example) by means of an intonation variation (by using a fundamental frequency contour with an "accentuating" value, i.e., one likely to lead to the perception of a stressed unit) in conjunction with an eyebrow movement. This hypothesis was tested in a previous analysis of video documents (Cavé et al., 1993; Cavé et al. 1996) which revealed a systematic link between rapid upward-downward eyebrow movements and the fundamental frequency. The Fo curve had the same rising-falling contour (and thus, an accentuating value) as the eyebrow movement curve, and the two curves were synchronized. The question raised by this study was whether this link is truly systematic, or even automatic, and whether it is rooted in the effects of muscular synergy and/or is the outcome of a communicational strategy on the part of the speaker.

4. Data reduction

4.1 Selection of eyebrow movements

We limited the present study to rapid upward-downward eyebrow movements (eliminating slow movements, considered to be "postural"). Only those movements where both eyebrows moved simultaneously were retained for analysis. The cutoff point for inclusion of a movement was set at a displacement of at least 3 mm by at least one eyebrow.

4.2 Relationship to vocalization

To relate the occurrence of eyebrow movements to the vocal production, various intervals were measured in the following manner:

a) If the eyebrows moved when the speaker was talking, two intervals were measured: the time between the onset of the vocal signal and the eyebrow movement, and the time between the eyebrow movement and the end of the vocal signal.

beg spkr voc   eyebrow mvmt   end spkr voc

interval (ms)   interval (ms)

These cases -- where the eyebrow movement occurred during vocalization -- were classified in category 1 (C1).

b) If the movement took place while the speaker was silent (even if the interviewer was talking), the intervals measured were between the eyebrow movement and the end of the speaker's last speaking turn, and between the eyebrow movement and the onset of the speaker's next speaking turn.

end spkr voc	eyebrow mvmt	beg spkr voc
interval (ms)	interval (ms)	

Whenever the eyebrow movement and the end of the vocal production were synchronized (difference of 30 ms or less), the interval between the movement and the beginning of the next vocalization was measured. Similarly, whenever the movement was synchronized with the beginning of the vocalization (difference of 30 ms or less), the time since the end of the preceding vocalization was measured.

end spkr voc + eyebrow mvmt	beg spkr voc
interval (ms)	

end spkr voc	beg spkr voc + eyebrow mvmt
interval (ms)	

These cases -- where the eyebrow movement took place during speaker silence or at a vocalization boundary -- were grouped into category2 (C2).

4.3 Coding fundamental frequency

For the speaker's vocal production, the fundamental frequency contours were classified into seven categories: rising-falling (RF),falling-rising (FR), plateau

(P), rising (R), falling (F),rising-falling-rising (RFR), and falling-rising-falling (FRF). In the passages where the speaker's and the interviewer's speech overlapped, the fundamental frequency could not be analyzed, so they were discarded. An intonation contour was considered to have an accentuating value when it contained points where there was a change in direction ("target points" in D.J. Hirst's terminology) (Guaïtella, 1991). Simple contours R and F were checked at the beginning and end to determine the presence of target points within the observed sequence. Only flat and falling contours had no target points, and thus no accentuating value.

5. Results

5.1 Eyebrows and vocalization

For category C1 (movement during vocalization), the eyebrow-movement/end-of-vocalization interval averaged 2100 ms and the beginning-of-vocalization/eyebrow-movement interval averaged 1500 ms. For category C2, the mean end-of-vocalization/movement interval was 1200 ms and the mean movement/beginning-of-vocalization was 490 ms. These means take into account the data for a speaker whose interactive strategy was the opposite of the others. (C1: beginning = 543 ms, end = 780 ms; C2:beginning = 1300 ms, end = 1300 ms). This speaker's data contributed to weakening the overall tendency, nevertheless exhibited by the majority of the subjects.

For the two categories pooled, the interval data as a whole showed that eyebrow movements occurred closer to the beginning of a vocal production than to the end. This effect was statistically significant ($p < .05$).

5.2 Eyebrows, accentuation, and fundamental frequency

Only C1 data was analyzed here. This category contained 43.5%of the observed eyebrow movements (excluding the 16 cases of overlapping out of 96). RF intonation contours -- which are similar in shape to eyebrow movement curves -- were found for 67.5% of the movements. Contours RF, FR,RFR, FRF, and R, all of which are accentuating, accompanied 93.75% of the movements. This left a mere 6.25% for contours with a non-accentuating value (F: 5%, P: 1.25%).

6. Discussion

The observed proximity of vocalization onset and eyebrow movements supports the idea that eyebrow movements act as marks of a new speaking turn.

Concerning the relationship between eyebrow movements and fundamental frequency, the data obtained here refine the results of previous studies (Cavé et al., 1993; Cavé et al., 1996). Although 93.75% of the eyebrow movements were associated with accentuating intonation contours(mostly RFs), the presence of eyebrow movements during non-accentuating productions (in particular, during periods of hesitation) seems to argue for the idea that the link between eyebrow movements and fundamental frequency variations is not an automatic one. In other words, eyebrow movements are not solely dictated by Fo variations, but may be speaker-controlled, being the consequence of linguistic and communicational choices. However, hand-voice coordination during the production of accentuating markers is known to be highly prevalent. The muscular synergy hypothesis cannot be rejected, then, even if it does not apply in all cases. In the contour distribution obtained here, we can see a distinction between what might be called "rhythmic" movements, which are more or less expressive and associated with accentuating variations in the fundamental frequency, and movements we could call "dynamogenic", which contribute to discursive planning and are associated with flatter contours (Guaïtella,1995).

Thus, rapid raising/lowering of the eyebrows does indeed appear to act as an attention-getting device, whether during the speaker's silent periods ("Look, I would like to speak"), or during speech production("Listen, what I'm saying is important"). Accordingly, we can contend that this gesture in fact derives from the eyebrow-raising movement whose primary function was to manifest enhanced visual attention.

In the future, we plan to look at what goes on for the interviewer, from the vocal and visual standpoints, when the speaker's eyebrows move, in hopes of finding out whether these attention signals may also function in a more interactive way, and, at least in certain cases, reflect either focusing on the interviewer's speech or between-speaker rhythmic harmonization, as proposed in Condon's (1976) intersynchrony hypothesis.

7. References

Birdwhistell, R.L.(1970). Kinesics and context, University of Pensylvania Press.

Bolinger, D.L. (1998, in press). "Intonation in American English", in D.J. Hirst & A. Di Cristo (eds), Intonation systems: a survey of twenty languages, Cambridge University Press.

Cavé, C., Guaïtella, I., Bertrand, R., Santi, S., Harlay, F., & Espesser, R. (1996). "About the relationship between eyebrow movements and Fo variations", in: Proceedings of the ICSLP, Philadelphia, 4, 2175-2178.

Cavé, C., Guaïtella, I., & Santi, S. (1993). "Relations entre geste et voix: Le cas des sourcils et de la fréquence fondamentale", in: Images et Langages: Multimodalité et modélisation cognitive. Actes du colloque interdisciplinaire du Comité National de la Recherche Scientifique, Paris, 261-268.

Condon, W.S. (1976). An analysis of behavioral organisation, Sign Language Studies, 13, 285-318.
D'Amico, M., & Ferrigno, G. (1990). Technique for the evaluation of derivatives from noisy biomechanical displacement data using a model-based bandwidth-selection procedure. Med. and Biol. Eng. and Comp,28.
Darwin, C. (1874). L'expression des émotions chez l'homme et les animaux, French translation by S. Pozzi & R. Benoit, Reinwald & Cie Libraires-Editeurs.
Ekman, P. (1989). "La mesure de l'expression faciale, Sciences et vie, 168, 24-31.
Ekman, P., & Friesen, W. (1975). Unmasking the face. A guide to recognizing emotions from facial clues, Prentice-Hall.
Ferrigno, G., & Pedotti, A. (1985). ELITE: A digital dedicated hardware system for movement analysis via real-time TV signal processing. IEEE Transactions on Biomedical Engineering, BME-32, II,943-949.
Guaïtella, I. (1991). Rythme et parole: comparaison critique du rythme de la lecture oralisée et de la parole spontanée, Doctoral Thesis, Aix-en-Provence.
Guaïtella, I. (1995). "Mélodie du geste, mimique vocale?", Semiotica, 103, 3/4, 253-276.