An investigation into non-articulatory facial gestures and intonation

María Luisa Flecha-García

marisa@ling.ed.ac.uk

ABSTRACT

Facial expressions have been mainly studied as expressions of underlying emotional states, while the linguistic information that can be conveyed in the face, with the exception of articulatory movements, has received little attention. There is some evidence to suggest that brow movements are used as conversational signals during speech, and research on Sign Language has shown how signers fulfil prosodic functions via facial movements, or facial gestures as they are referred to here. However a line of investigation involving non-articulatory facial gestures and prosodic events seems to have been neglected in spoken language. This paper presents an ongoing project to investigate a possible relation between non-articulatory facial gestures and intonation, focusing in particular on the temporal alignment of such gestures with pitch movements. The final goal is not only to gain a better understanding of audio-visual speech production but also to contribute to the design of multimodal communication systems.

1 INTRODUCTION

The science of face perception has benefited from a substantial amount of research in such areas as the relation between emotion and facial expression (e.g. Ekman & Rosenberg 1997). This relationship has become the main focus of attention and has been treated as a route into psychological processes. In the area of speech, research has concentrated on the movements of the articulators in the lower part of the face, that is, lips, jaw, teeth and tongue. The influence that articulatory movements have on speech perception was clearly demonstrated with the discovery of the McGurk Effect (McGurk & MacDonald 1976) and has been investigated in detail since then (e.g. Massaro & Cohen 1983). Articulatory movements have also been studied extensively in the fields of audio-visual speech synthesis and recognition. On the contrary, the relationship between non-articulatory facial gestures and speech awaits characterisation.

“Non-articulatory” refers to facial movements that are not necessarily tied to the articulation of speech and involve mainly the upper face (e.g. raising eyebrows or blinking). The term “gesture” has been used in the literature to refer to hand movements, while facial movements have been mainly referred to as “expressions”. However, I will use “facial gesture”, or Chovil’s term “facial display” (1989), because “facial expressions” have been linked to the expression of underlying emotional states, whereas the present research concentrates on the linguistic information conveyed in the face. Although hand gestures are not included in the present study, it is important to mention how researchers have found interesting correlates between hand gestures and features of speech, including intonation (e.g. Kendon 1980; McNeill 1992; McClave 1998). Compared to this, studies on non-articulatory facial gestures in connection with speech are scant. However, as I explain below, there is some evidence for the linguistic function of such gestures.

One area in which the linguistic function of upper facial movements has been well investigated is Sign Language (SL). It is widely accepted among SL researchers that some of these nonmanual markers are used to fulfil prosodic functions (Wilbur 1994; Corina 1989). Facial displays have been found to mark the introduction of a topic, and of certain clauses or questions. For example, yes/no questions in ASL are accompanied by an eyebrow raise, whereas wh-questions have an eyebrow furrow. Obviously, the use of these displays with linguistic functions in SL is much more systematized than in the speech of hearing people, simply because signers do not have the auditory modality. But it is nevertheless very important to note how it is upper face movements that substitute for the auditory suprasegmental cues that signers do not have.

Important evidence for linguistic functions of upper face gestures has also been found in spoken language. Ekman and Friesen (1978) developed a detailed system to measure movements of the facial muscles. Although this system has been mainly used to study expressions of emotion, Ekman (1979) observed how two of the seven brow actions described in it seem to have important linguistic roles too, such as: giving emphasis to words and phrases, marking punctuation, and signalling questions.

More recently, Chovil (1989; 1991/92) analysed videos of spontaneous dialogues between two people in order to examine how facial displays are used in conversation and whether their frequency may be affected by social factors. She found both syntactic and semantic linguistic functions in facial displays. Within the syntactic displays the most common were eyebrow movements that served as emphasisers and question markers. Chovil (1997) described two different approaches to the study of facial expression: the Emotional Expression Approach, that has been followed by most researchers, and the Social Communicative Approach, an emerging view that considers expressions of emotion as a subset within a much broader domain of facial displays. Following the Social Communicative Approach, Bavelas and Chovil (1997) have emphasised the communicative functions of facial displays and the need to study faces in dialogue. They proposed that "the facial displays of conversants are *active*, *symbolic* components of *integrated* messages (including words, intonation, and gestures)" (p. 334).

So far there have been some studies suggesting the linguistic function of certain facial displays. However, there is no information about the relation between these gestures and suprasegmental aspects of speech. This project aims at characterising the possible relation between upper face gestures and pitch movements. One aspect of particular interest is the synchronisation between eyebrow movements and pitch phenomena. I will describe below the methodology involved in a preliminary study and the final goals of this research.

2 METHODOLOGY

Two male speakers of English were filmed in a television studio for a pilot experiment using the Map Task design, with some modifications. The Map Task (Anderson *et al*. 1991) is a cooperative task involving two participants with a map each that the other cannot see. One of the maps has a route marked on it and the speakers are told that their goal is to reproduce this route on the other map. The data from the original design consist of dialogues between pairs of participants carrying out the task. For the purpose of the present study new conditions were introduced in the design, as described below.

2.1 Procedure
The recording was divided into four different parts:
2.1.1 Recorded rehearsal by one participant
One of the participants, Subject A, was given a map with a route and was asked to rehearse aloud the instructions he would give to a potential viewer trying to draw that route on another map of the same area. He was told that this session would not be recorded but was intended to help adjust the cameras and sound. However both sound and image were recorded using a clip microphone attached to the speaker's chest and a camera placed in front of him. The cameraman was instructed to act as if he was only doing a test recording to adjust the camera.
2.1.2 Recording aimed at potential viewers
Subject A was asked to give the route instructions to the camera as if he was addressing a potential viewer who would later on watch the video and had to draw that route following his instructions. The session was recorded as before, this time with the awareness of the participant.
2.1.3 Dialogue with a second participant
Subject A and Subject B performed the task as in the original design of the Map Task. They sat facing each other across a desk, with their maps placed on sloping boards. Subject A gave instructions to Subject B so that B could draw the missing route on his map. They could say anything at any time but were asked to avoid hand gestures. Both maps were not identical and they had to negotiate to reproduce the route without looking at each other's map. Two cameramen recorded simultaneously each subject from a frontal view, and clip microphones recorded their speech as above. A third camera placed behind Subject B recorded his map as he drew the route on it.
2.1.4 Different task
Subject A was asked to describe the general features of their bedroom to Subject B, while B drew it on an A3 sheet of paper. Then they switched roles so that B described his room while A drew it. The session was recorded as above.

2.2 Materials
The resulting materials were an SVHS videotape and a DAT audiotape of approximately 30 minutes. In the videotape, the first two sessions showed a frontal view of the Subject A's head and shoulders. In the following sessions there was a similar view of both participants on a split screen with a small inset at the bottom showing the B's map, where the route was being drawn.

2.3 Preliminary measurements
There were some preliminary observations on the pilot experiment. Part of the speech was analysed using the ToBI system (Silverman *et al.* 1992) to transcribe pitch boundaries and accents. As for the videotape, the Subject B's eyebrows were measured during some of his utterances on the third session described above. An acetate grid was placed on a TV screen playing the video and measures were taken of the vertical distance that his eyebrows moved from an axis aligned with corners of the eyes. This was done in order to assess the difficulties involved in such measurements and the possibility of finding a baseline position of his eyebrows that would allow further measurements. There has been no further analysis of the data but the main objective of this pilot study was achieved, namely to confirm that some facial gestures seemed to correlate with prosodic features of the speech, and that the available tools would allow measurements of the auditory and visual data. New measuring techniques for the visual data are being searched for, such as computer programs that will allow faster and more precise measurements.

3 FUTURE DEVELOPMENTS AND GOALS

There have been new recordings made with four female speakers of English. The design of this experiment was similar to the one above except for the fourth part that was not included here. Another difference was the introduction of a final recording where participants read out a list of all the landmark names that had appeared on the maps, as is done in the original Map Task design.

It is expected that from the different conditions in the above design we will be able to extract information about facial gestures that are planned versus spontaneous ones, and also gestures which are responses to the interlocutor. At the same time the speech accompanying those facial displays will be analysed paying special attention to pitch events. Based on preliminary observations from the pilot study, the main interest is focused on the time-course of those gestures and their correlation with pitch phenomena.

One obvious goal of the present research is to gain a better understanding of audio-visual speech production. A further goal is to contribute to the design of multimodal communication systems. Computer animated agents are becoming more and more popular in multimedia, in such areas as speech technology. However, most of the "talking heads" in these systems seem to lack accurate coordination of facial movements and speech, which may impair the comprehension of messages and may affect the credibility of the animated face. This lack of accuracy results not only from technical difficulties but also from the scarce knowledge available about this coordination in human speech. By investigating the time-course of non-articulatory facial gestures in human speech, and their relationship with prosodic events, such as pitch movements, the present research aims at providing some of the missing information for the improvement of animated faces in communication systems.

ACKNOWLEDGEMENTS

Many thanks to my supervisors, Dr. Ellen G. Bard and Prof. Robert Ladd, for their help in this project.

References

Anderson, A. H., M. Bader, E. G. Bard, E. Boyle, G. Doherty, S. Garrod, S. Isard, J. Kowtko, J. McAllister, J. Miller, C. Sotillo, H. S. Thompson & R. Weinert. 1991. The HCRC Map Task Corpus. *Language and Speech,* 34: 351-366.

Bavelas, J. B. & N. Chovil. 1997. Faces in dialogue. In J. A. Russell & J. M. Fernández-Dols (eds.) *The Psychology of Facial Expression.* Cambridge: Cambridge University Press, pp. 334-346.

Chovil, N. 1989. *Communicative functions of facial displays in conversation*. Unpublished PhD thesis, University of Victoria, BC, Canada.

Chovil, N. 1991/92. Discourse-oriented facial displays in conversation. *Research on Language and Social Interaction,* 25: 163-194.

Chovil, N. 1997. Facing others: A social communicative perspective on facial displays. In J. A. Russell & J. Fernández-Dols (eds.) *The Psychology of Facial Expression.* Cambridge: Cambridge University Press, pp. 321-333

Corina, D. P. 1989. Recognition of affective and noncanonical linguistic facial expressions in hearing and deaf subjects. *Brain and Cognition*, 9: 227-237.

Ekman, P. 1979. About brows: Emotional and conversational signals. In M. von Cranach, K. Foppa, W. Lepenies, & D. Ploog, (eds.), *Human Ethology.* Cambridge: Cambridge University Press, pp. 169-248

Ekman, P. & W. V. Friesen, 1978. *The Facial Action Coding System (FACS): A technique for the measurement of facial action.* Palo Alto, CA: Consulting Psychologists Press.

Ekman, P. & E. Rosenberg (eds.). 1997. *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS).* Oxford: Oxford University Press.

Kendon, A. 1980. Gesticulation and speech: Two aspects of the process of utterance. In M. R. Key (ed.) *The Relationship of Verbal and Nonverbal Communication.* The Hague: Mouton, pp. 207-227.

Massaro, D. W. & M. M. Cohen. 1983. Evaluation and integration of visual and auditory information in speech perception. *Journal of Experimental Psychology: Human Perception and Performance,* 9: 753-711.

McClave, E. 1998. Pitch and Manual Gestures. *Journal and Psycholinguistic Research,* 27(1): 69-89.

McGurk, H. & J. MacDonald. 1976. Hearing lips and seeing voices. *Nature*, 264: 746-748.

McNeill, D. 1992. *Hand and Mind: What Gestures Reveal About Thought.* University of Chicago Press.

Silverman, K., M. E. Beckman, J. Pitrelli, M. Ostendorf, C. Wightman, P. Price, J. Pierrehumbert & J. Hirschberg. 1992. ToBI: a standard for labeling English prosody. *Proceedings, Second International Conference on Spoken Language Processing* 2: 867-70. Banff, Canada.

Wilbur, R. B. 1994. Eyeblinks and ASL phrase structure. *Sign Language Studies*, 84: 221-240.