Berthouze, L., Prince, C. G., Littman, M., Kozima, H., and Balkenius, C. (2007).
Proceedings of the Seventh International Conference on Epigenetic Robotics: Modeling
Cognitive Development in Robotic Systems. *Lund University Cognitive Studies, 135*.

Emotional McGurk effect – an experiment

Åsa Abelin
Department of Linguistics, Göteborg University
abelin@ling.gu.se

Abstract

Video and audio recordings of emotional expressions were used to construct conflicting stimuli in order to perform a McGurk experiment. Perception tests were made on bimodal, auditory, visual and McGurk stimuli. In unimodal condition the auditive channel showed more expected interpretations than the visual channel. In the McGurk condition however, the visual channel was more reliable than audio at conveying emotions. However, most frequently the listener hears something which is present neither in video nor audio. Bimodal non-conflicting information is most reliable. The results are relevant for construction of robots and avatars, since facial emotional expression is always present.

1. Introduction

Speech is seldom perceived only auditorily, but bimodally. In fact, even when perceiving only the speech of a person we might be affected by earlier experiences of seeing that person articulate (Rosenblum 2005). This paper presents the results of a perception experiment of emotional Mc Gurk effect in Swedish. The phenomenon of McGurk has been widely studied since McGurk and McDonald (1976) and concerns the perception of conflicting visual and auditive input, see e.g. Massaro (1998a) or Traunmüller (2006).

The experiment is part of a set of experiments which aim at elucidating the question of innate versus learned expression of emotions. The aspects to be investigated are cultural differences versus universals, and interpretation of emotional expressions by the child.

The questions under study in this experiment are the following: 1. Is the auditory or the visual modality dominant in perception of emotions? 2. Are the different emotions connected to different modalities? 3. What happens when the auditive and visual stimuli are conflicting ?

2. Method

Video and audio recordings of emotional expressions were used to construct conflicting stimuli in order to perform a McGurk experiment. Perception tests were also made on bimodal, only auditory and only visual stimuli.

One Swedish female speaker was video and audio recorded using a MacBookPro inbuilt camera and microphone. She expressed the six basic emotions *happy*, *angry*, *surprised*, *afraid*, *sad*, *disgusted* plus *neutral,* saying "hallo, hallo". These bimodal expressions of emotions were subjected to a perception test with five perceivers (Test 1). This showed that the most successful productions were *happy* (4 out of 5), *surprised* (5 out of 5) and *afraid* (4 out of 5). These three expressions were chosen, along with *angry*

The audio and the video for these emotions were separated and then combined to form the McGurk stimuli shown in Table 1. There were no problems with the synchronization of audio and video. The stimuli were presented to ten perceivers (seven female, three male, aged 19–33 years) in test 2. In test 3 and test 4 audio and video were separately presented to a group of nine perceivers (all female, aged around 20 years). All tests employed free choice.

The stimuli of the McGurk test, were all the possible combinations of *happiness*, *fear*, *surprise* and *anger* are shown in table 1.

Table 1. The stimuli of the McGurk condition.

Stim	Visual	Auditive
1	happy	angry
2	surprised	afraid
3	angry	happy
4	afraid	surprised
5	angry	afraid
6	surprised	angry
7	happy	afraid
8	afraid	angry
9	surprised	happy
10	angry	surprised
11	afraid	happy
12	happy	surprised

Method of analysis

As the subjects responded with free choice the answers had to be classified. As an example *afraid* included *insecure* and *worried* but not *sad*; *surprised* included *questioning* and *curious* but not *hesitating* and *angry*

included *irritated* but not *serious*. In the McGurk test the answers were labeled "video" if the response was in accordance with the video stimulus, "audio" if the response was in accordance with the audio stimulus, and "other" if the response was not in accordance with either video or audio.

3. Results

The general result from the McGurk test is that perceivers interpreted either 1) in accordance with the face, 2) in accordance with the voice 3) as other emotion which means a) a fusion: an interpretation of the stimulus as something else than what face or voice intended, or b) separation of voice and face which gave two different answers, however mostly not in accordance with either voice nor face. The results for the different emotions are summarized in Figure 1.

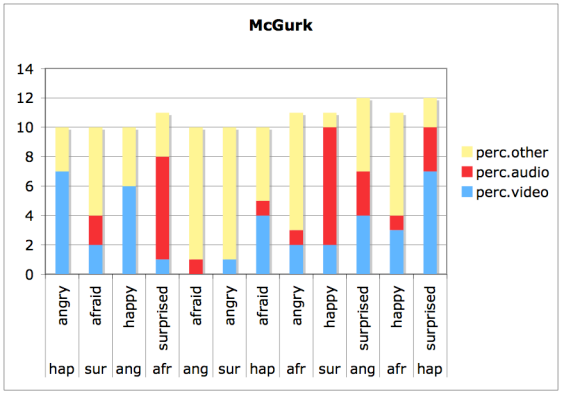

Figure 1. Interpretation of the twelve conflicting audio and video stimuli. Bottom row text: video, upper row text: audio.

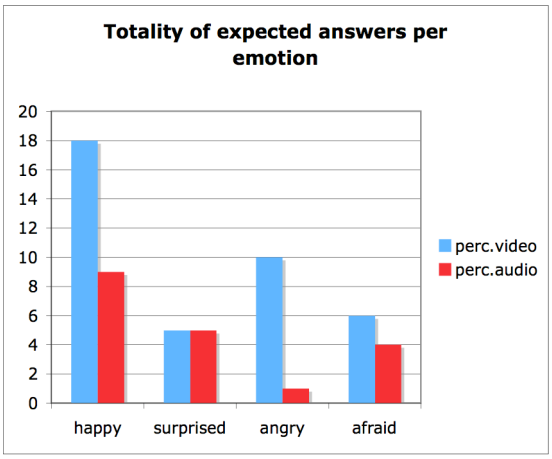

Figure 2. Interpretation of different emotions as expected, in video and audio, in the McGurk test. In a situation with conflicting stimuli the visual channel is generally preferred, but the dependency is different for different emotions.
Question number 2 was: Are the different emotions connected to a certain modality? Figure 2 shows that the emotions which were not perceived as "other" were generally perceived best visually. *Happy* and *angry* were the most visual emotions.

4. Summary and discussion

In unimodal condition the auditive channel showed more expected interpretations than the visual channel. In the McGurk condition however, the visual channel was more reliable than audio at conveying emotions ; in case of conflicting information – invent something, otherwise trust the face. However, most frequently the listener hears something which is present neither in video nor audio. Bimodal non-conflicting information is most reliable. Specific emotions are not connected to either the visual or the auditive modality.

An earlier cross linguistic experiment on multimodal interpretation of emotional expressions (Abelin, 2004) showed that visual stimuli improved interpretation of emotional expressions in an other language. The importance of the visual modality in linguistic development can be expected to follow the same line.

The results are relevant for construction of e.g. robots and avatars, since facial emotional expression is always present; even if there is no intentional emotional expression, an emotions will always be interpreted by the perceiver.

References

Abelin Å (2004). Spanish and Swedish Interpretations of Spanish and Swedish Emotions – the influence of facial expressions, in *Proceedings of Fonetik 2004*, Stockholm

Massaro D W (1998). *Perceiving talking faces: From speech perception to a behavioural principle.* Cambridge, Massachusetts: MIT Press.

McGurk H and McDonald I (1976). Hearing lips and seeing voices. *Nature,* 264, 746-748

Rosenblum L D (2005). Primacy of multimodal speech perception in: D B Pisoni and R E Remez, eds *The handbook of speech perception*, Blackwell publishing

Traunmüller H (2006). Cross-modal interactions in visual as opposed to auditory perception of vowels. In: G. Ambrazaitis, S. Schötz, eds, *Proceedings from Fonetik 2006*. Lund, 137-140.