# Cross-modal Bias of Voice Tone on Facial Expression: Upper versus Lower Halves of a Face

*Beatrice de Gelder, Jean Vroomen and Paul Bertelson*

Tilburg University, Dept of Psychology
Tilburg, The Netherlands

## ABSTRACT

Emotional states are communicated by seeing the face as well as by listening to affective prosody. The two inputs can also be present at the same time and are then processed concurrently. The present experiment examines the role of the voice on the upper versus lower halves of a face. Previous research using an angry-fear facial expression continuum showed that recognition of the lower half of a face was nearly at chance level. Our experiment asked whether in these circumstances the impact of the voice would be the same for both face halves. The results showed that the cross-modal effect of the voice was the same for the two face conditions

## 1. INTRODUCTION

A central theme in research on face perception concerns the question whether a face stimulus is perceived as a whole (or holistically) [1], as a configuration [2], or simply as a bundle of independent features. This issue is extensively debated in research on facial identity recognition but has not attracted much attention so far from the side of students of facial expression recognition.

One way to approach this issue is to study whether the classic inversion effect obtains for expression recognition as it does for identity recognition. The effect of up-down face inversion has attracted much attention in the literature on face identification (see [3] for a review). Starting with the findings of [4], the fact that inversion results in a stronger reduction of identification performance for faces than for other objects has been taken as evidence for a special status of faces. The general notion has been that the recognition of upright faces takes advantage of configural attributes which are not available in inverted faces.

Is the configuration also important in telling the emotion from the face? This question has received less attention, in part because very little is known about how the different regions of the face convey expression: do they operate separately, or is it the facial configuration that is the bearer of the expression? The well-known phenomenon called the Thatcher illusion [5] suggests the latter. Inverting the eyes and the mouth in an upright face causes a perception of gruesome expression. But when such a face is presented upside-down, the unpleasant impression tends to disappear. The inverted stimulus does not prompt recognition of a familiar person, but there is nothing particularly striking about the expression. More directly relevant evidence was obtained by [6] who had subjects judge the emotion expressed by faces presented upright or upside-down. Whereas upright faces were identified with a high degree of accuracy, matching classical data by [7], the performance with inverted faces varied considerably among emotions, from good performance for happiness and surprise to very poor performance for emotions like sadness, fear, or anger. Related evidence was provided by [8] in a study in which stimuli from several facial expression continua were presented either upright or upside-down. Inversion made discrimination along an angry-sad continuum impossible.

For two other continua, happy-sad and angry-fear, non-random discrimination was observed in spite of the inversion. However, all the evidence of categorical processing obtained for upright faces disappeared with inversion.

Two studies have addressed the issue of the role of separate facial features in expression recognition. In [9], three facial expression continua were used (happy/sad, angry/fear and angry/sad) and expression recognition was compared for upper half, lower half, and full faces. Of particular relevance for the present study are the results for the angry/fear continuum. When presented with full faces, subjects perceived these two emotions clearly categorically. The situation was different, though, when only part of the face caried an emotional expression and the other half is neutral. For the continuum consisting of an upper half with expressions going from angry to sad, the results were the same as for the full morphs and these emotions were thus perceived categorically. But when the emotional expression was entirely a matter of the information in the lower half of the face, subjects no longer made consistent category assignments. In other words, at least for these two emotions the lower part of the face, most prominently the shape of the mouth, did not by itself allow an interpretation of the expression as either angry or fear. This result suggests that there is a very unbalanced contribution of the lower versus the upper face half. However, it does not allow us to conclude that facial features rather than the whole face or its configuration are the bearers of information. This inference would require data from a comparison between the upper (expressive) face part when presented with a neutral lower face half vs. in isolation. Only if no difference is found between the context and the isolation condition can one conclude that expression information is feature-based. In [10] a synthetic face was used with either a happy or an angry expression which was achieved by changing independently the contribution of the curvature of the eyebrow or that of the mouth. They report that the results indicate that participants evaluated and integrated information from both features. In this study the role of features cannot be properly appreciated since features always appeared in a full face context.

The question of the critical role of the whole face versus its separate features in expression recognition is interesting in its own right but it is also particularly relevant for understanding how information from the face is combined when cues other than facial ones are present. Among non-face cues, the voice occupies a privileged position because in natural circumstances a face expression and a tone of voice are often present concurrently. Previous studies have shown that information from the voice is combined with that from the face in an automatic fashion so that a judgement of the facial expression is affected by a concurrently presented voice tone [11, 12, 13, 14]. Here we relate questions about such cross-modal effects to that of the role of face parts. The experiment we report used the angry/fear face continuum in which the lower face half did not show evidence for categorical emotion perception. In this bi-modal study we asked whether the non-expressive lower face half would benefit from being presented in combination with a voice expressing either fear or anger. We consider two possibilities. One is that the nonexpressive or 'silent' lower face stimuli now benefit maximally from an addition of the voice while the impact of the voice on the already highly recognizable upper face expressions is minimal. The other alternative is that the impact of the voice is the same in each condition so that the overall pattern of the results is not different from that obtained in our previous face-only study.

## 2. METHOD

*Visual materials.* Preparation of the materials started with a set of 11 B/W pictures taken from a previous study of categorical perception of facial expressions [8]. These

consisted of an angry and a fearful facial expression of the same male actor taken from a standard picture series [7]. Between these two extremes, nine intermediate synthesized expressions were made by a morphing technique. As was done in a previous mono-modal study [9], each picture was cut in two halves at the horizontal midline. The halves were then reassembled in such a way that two different picture sets were generated that corresponded to two experimental conditions. One condition consisted of the full set of morphed upper parts that were each combined with the same lower face part that had a neutral expression. The other condition was the reverse and presented each of the lower face halves combined with the same neutral upper face part.

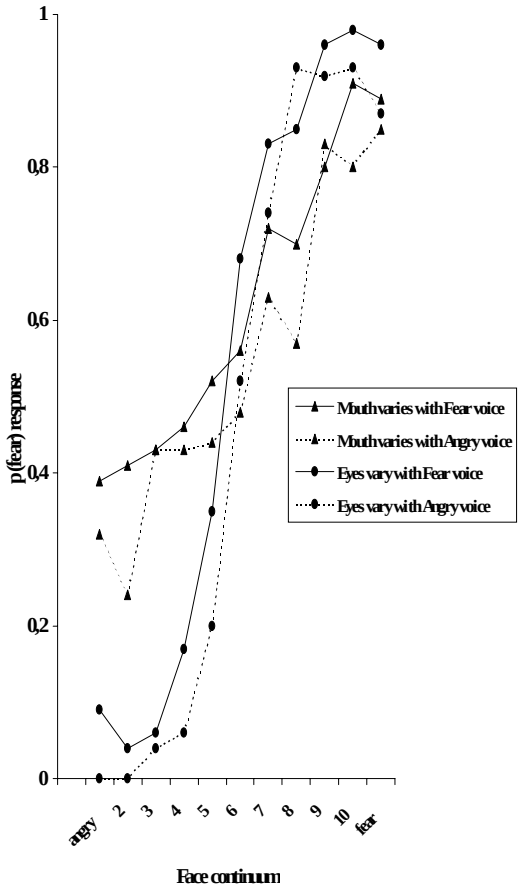
*Auditory materials.* A sentence with a neutral content ('his girlfriend arrives by plane') was spoken by a semi-professional actor who had been instructed by way of examples of the relevant situation to pronounce the sentence in either a angry or a fearful tone. The length of each utterance was approximately 1.80 s.
The face and the voice were presented at the same time and the face remained on the screen for the duration of the utterance. The face was presented on a 15 inch computer screen, and the utterance was played over two loudspeakers.

*Design and procedure.* The experiment consisted of one block of 144 randomly presented trials (6 repetitions x 22 different pictures x 2 voices). Subjects were instructed to judge the face thereby ignoring the voice. The subjects pressed one of two keys in front of them. One key was labeled 'angry', the other 'fear'. Each new trial was started by the subject pushing a middle key.

## 3. RESULTS

The proportion of 'fear' responses was computed for each subject, and the averages are displayed in Figure 1. As can be seen, the eyes were more informative about the happy-fear contrast than the mouth, because the range of the responses in the eye-varied continuum is bigger than that in the mouth-varied continuum. However, the audio-part contributed in both continua equally to the identification response. In an overall ANOVA there was, as expected, a main effect of the visual, $F(10,80) = 46.91$, $p < .001$, and the auditory stimulus, $F(1,8) = 6.43$, $p < .05$. There was also a significant interaction between the visual stimulus and the mouth/eye-varied continuum, $F(1,8) = 7.76$, $p < .001$, indicating that the eyes were more informative than the mouth. All other effects were non-significant.

## 4. DISCUSSION



Our results show very clearly that voice and face expressions combine and that as a consequence, the judgement of the face reflects the impact of the voice in the upper as well as the lower face condition. When we compare these two conditions, the actual impact of the voice on the face expression turns out to be the same in the two conditions. In other words, the effect of the voice is not modulated by the degree of expressiveness of the actual face stimulus. Further research must examine how information from two modalities is combined when the input in one of these is minimally informative. The FLMP model developed by Massaro and collaborators predicts that the more ambiguous a feature is, the more it will benefit from other cues that are concurrently presented. This suggests that the lower face half condition should show a greater cross modal impact that the upper half. Our data suggest that cross modal combinations may take place following a more complex recipe where there may be boundaries to the degree to which input channels can compensate for lack of informativeness in one of the other channels.

## 5. REFERENCES


1. Farah, M.J., Tanaka, J.W., and Drain, H. M. "What causes the face inversion effect?," *JEP:HPP, Vol 21*: 628-634, 1995.

2. Carey, S., and Diamond, R. "Are faces perceived as configurations more by adults than by children?," *Vis. Cogn., Vol 1* : 253-274, 1994.

3. Valentine, T. "Upside-down faces: A review of the effect of inversion upon face recognition," *Brit. J. of Psy., Vol. 79*: 161-204, 1988.

4. Yin, R.K. "Looking at upside-down faces," *JEP:HPP, Vol. 81*: 141-145, 1969.

5. Thompson, P. "Margaret Thatcher: A new illusion," *Perception, Vol. 9*: 483-484, 1980.

6. McKelvie, S.J. "Emotional expression in upside-down faces: Evidence for configurational and componential processing," *Brit. J. of Soc. Psy. Vol. 34*: 325-334, 1995.

7. Ekman, P., and Friesen, W.V. *Unmasking the face.* Englewood Cliffs: Prentice Hall, 1975.

8. de Gelder, B., Teunisse, J-P., and Benson, P.J. "Categorical perception of facial expressions: Categories and their internal structure," *Cog. and Emo, Vol. 11*: 1-23, 1997.

9. de Gelder, B., Vroomen, J., and Popelier, T. "Facial expressions: Do part play a full role?. *J. of Int. Neuropsy. Soc., Vol. 2:* 206, 1996.

10. Ellison, J.W. and Massaro, D.W. "Featural evaluation, integration, and judgement of facial affect," *JEP:HPP, Vol 23*: 213-226, 1997

11. de Gelder, B., and Vroomen, J. "Hearing smiles and seeing cries: The bimodal perception of emotion," *Bull of the Psych. Soc.*, 1995

12. de Gelder, B., and Vroomen, J. "The perception of emotions by ear and by eye", submitted.

13. Massaro, D.W., and Egan, P.B. "Perceiving affect from the voice and the face," *Psych. Bull. and Rev. Vol. 3*: 215-221, 1996.

14. de Gelder, B., Vroomen, J., and Bertelson, P. "Upright but not inverted faces modify the perception of the emotion in the voice," *Cahiers de Psychologie Cognitive, 17*: 1021-1032, 1998.