FACIAL EXPRESSIONS IN SINGING
(A Pilot Study)

Nicole Scotto Di Carlo and Isabelle Guaïtella
Laboratoire Parole et Langage, U.R.A. 261 CNRS, Aix-en-Provence, France

ABSTRACT

An experiment dealing with the recognition of emotions in speech and in singing for two subject populations (opera amateurs and non-amateurs) based on visual, auditory, and audiovisual perception tests was used to assess the respective roles of the singer's voice and facial expressions in the perception of emotions produced by a professional soprano, and to determine how spectators decode these emotions.

INTRODUCTION

How do opera singers manage to produce the physiological or functional facial movements (required for emitting a given sound) at the same time as they produce the facial expressions (aimed at displaying the wide range of human feelings) in order to reflect the emotions they must transmit to the public? Do spectators use a specific strategy to decode the emotions expressed by lyrical artists? Do they ignore the functional part of a facial expression and focus their attention solely on the expressive part? We designed a pilot experiment that allowed us not only to assess the respective roles of the voice and the face in the perception of emotions produced by a professional soprano, but also to determine how spectators decode these emotions.

EXPERIMENTAL PROCEDURE

Corpus

The soprano recruited had to be both an excellent professional singer and a good actress. She was filmed in an anechoic chamber using two synchronized cameras so as to obtain simultaneous profile and front view videotapes of her face as she carried out a certain number of tasks of increasing complexity. For the purposes of the present study, the videofilms were used to extract a corpus of the vowel [a], spoken and sung on C_4 (262 Hz) for the lower register, C_5 (523 Hz) for the middle register, and C_6 (1047 Hz) for the upper register, with four basic emotions : joy, sorrow, fear, and anger (which taken two at a time, can be opposed along the activation/inhibition dimension). The best sequences were selected and the most characteristic part of the facial expression in each was photographed from the videotape. This gave us a corpus composed of both sound sequences and photographs.

Testing

Tests were administered to two populations of subjects : 20 opera lovers who regularly watch videotapes of lyric works and 20 subjects with no particular interest in the lyrical arts. Using a computer-driven projection system, the slides used in the visual tests were presented to the subjects for four seconds. Testing was done in three phases : (1) a solely visual phase where the subjects had to judge static images only (photographs), (2) a purely auditory phase where they had to judge sound sequences only, and (3) an audiovisual phase which combined the sound sequences and the static images.

- *Visual test.* The visual test consisted of two steps. In the first, the subjects had to identify the emotions expressed. In the second, they were told what the emotion would be and had to assess its intensity.

The first step (identification of emotions) included two series, one containing real images where each face corresponded to a given emotion and a given register (vowel [a] spoken and sung in the lower, middle, and upper registers) and one containing mixed images where the top and bottom of the face did not correspond to the same emotion. Joy was permuted with sorrow, and fear with anger, always within the same register.

The second step (assessment of emotion intensity) consisted of three phases. In the first, the subjects' task was to choose the most expressive face out of two simultaneously-presented real photographs, each representing a given emotion (of which the subjects were informed) expressed in the lower and upper registers. In the second phase, the two photographs represented the same emotion (again, of which the subjects were informed) expressed in the lower and upper registers, but this time the two halves of the faces were interchanged across the lower and upper registers. For example, a top half (lower register)/bottom half (upper register) face had to be compared with a top half (upper register)/bottom half (lower register) face. The third phase dealt solely with emotions expressed in the spoken voice. For each emotion, the subjects had to select the face they felt was the most expressive among three simultaneously presented photographs, one of a real face and two of reconstructed faces made from two right halves and two left halves.

- *Auditory test.* The subjects' task was to identify the emotions by listening to two series of randomly mounted sound sequences containing the vowel [a] spoken and sung in the three registers.

- *Audiovisual test.* The subjects had to identify the various emotions by looking at photographs of real faces on which each facial expression corresponded to a given emotion and a given register (vowel [a] spoken and sung in the lower, middle, and upper registers) while listening to the corresponding synchronized sound sequences.

RESULTS

Soprano's Morphological Features

- *Expressivity index.* The soprano who volunteered for this experiment has a relatively symmetrical face. Because of this, her right side was judged to be as expressive as her left side (36% and 37%, respectively) for all emotions.

It is interesting to note that, for all emotions pooled, the real face was judged to be expressive by only 24% of the subjects. It appears as though joining two right halves and two left halves reinforces the facial symmetry, making the mixed faces more expressive.

For sorrow, fear, and anger, the faces made up of two left halves were considered by the greatest number of subjects (49%) to be the most representative of the emotion in question (compared to the real face and the face made up of two right halves). These results are consistent with those obtained in an experiment by Sackheim and Gur [1], where the emotion in faces composed of two left halves was judged to be more intense than in faces composed of two right halves. For joy, the right side was judged to be more expressive by 50% of our subjects, a finding which is in line with Bruyer's [2] theory that the left hemisphere is involved in the expression of pleasant emotions.

- *Within-register salience index.* For all subjects and emotions pooled, when joy and sorrow or fear and anger were mixed within a given register, the bottom of the face clearly dominated in speech (B = 54% and T = 32%), whereas this effect was not as pronounced in singing (B = 37% and T = 32%). Moreover, the differences between the top and bottom of the face gradually decreased from the lower register to the upper register (lower : T/B = 37/43; middle : T/B = 33/37; upper : T/B = 27/31), probably because emotion information becomes increasingly difficult to perceive, making it necessary to use all available cues regardless of whether they are on the top or bottom of the face.

For all subjects and all registers pooled, the bottom of the face dominated for joy (T = 13%, B = 58%) and sorrow T = 16%, B = 51%), whereas the top dominated for fear (T = 36%, B = 18%) and anger (T = 61%, B = 46%). These results partially corroborate Bassili's [3] findings, which showed that the top part of the face is used to detect anger; the bottom, joy, sorrow, and disgust; and both parts, surprise and fear.

Between-medium Comparisons

Regardless of the medium used (sound, image, sound+image), the emotions expressed in speech were identified the best (83% vs. 56% for singing).

For all subjects, emotions, and registers pooled, sound was found to be the poorest conveyor of information about emotions (39%), whereas images attained 74% and sound accompanied by images, 76%. The combination of image and sound improved the score by 7% over images alone for the spoken voice (I = 88%, S+I = 95%) and by 4% over singing in the lower register (I = 72% and S+I = 76%). In contrast, the scores were virtually the same in the middle (I = 81% and S+I = 82%) and upper (I = 53% and S+I = 54%) regis-

ters. As the acoustic cues of emotion are partially destroyed in these registers, almost no additional information is provided by the sound. In the other registers, the acoustic cues are added to the visual cues and thereby contribute to improving recognition.

Among the four basic emotions, for all subjects, registers, and media combined, sorrow (73%) and joy (72%) were recognized the best. Then came fear (54%), with anger in last place (53%).

Comparison of Lower and Upper Registers

- *Between-register comparison of expressivity.* When subjects had to choose the most expressive of two real photographs representing the same emotion in the lower and upper registers, the lower register predominated at 50%, vs. 48% for the upper register. For the emotions taken separately, the upper register was judged to be more expressive for joy, sorrow, and fear, while the lower register was judged so for anger. This is no doubt due to the wide buccal opening in the upper register, which gives the impression of a more intense emotion, whether it be joy, sorrow, or fear. The small buccal opening in the lower register is more representative of anger, whose main visual feature is the clenching of the teeth.

- *Between-register comparison of salience.* When for a given emotion, the lower and upper registers were mixed to obtain L/U and U/L pairs, the lower register dominated in 55% of the cases. The faces judged to be the most representative of joy and sorrow were the U/L mixtures (lower register on bottom of face). The faces judged to be the most expressive of fear and anger were the L/U mixtures (lower register on top of face). This confirms our previous results. The dominance of the lower register can be explained by the fact that, in the upper register where all facial muscles are being used to produce the sound, achieving good vocal technique takes precedence over expressing emotions.

Auditory Identification of Emotions

At the auditory level, for all registers and subjects pooled, the recognition rate for sorrow was 56%, fear 38%, joy 33%, and anger 28%.

In the upper register, for all subjects pooled, joy was recognized the best (48%). In contrast, none of the subjects recognized sorrow (0%) : 45% thought it was joy, 30% fear, and 23% anger. This is most likely related to existing emotional stereotypes. Joy, fear, and anger are the three emotions which, when speaking, the soprano produced in the upper register (sometimes even in the upper-upper register). It seems that joy is associated with the upper register. Inversely, one can hypothesize that the lower register is associated with sorrow. If this hypothesis is valid, not only should sorrow be recognized better in the lower register (which was indeed the case : 73%) than in the upper register (0%), but joy should be poorly recognized in the lower register (which was also the case : 18%). Fear and anger seem to follow the sorrow pattern and are treated as "internalized" versions of the traditionally accepted emotional stereotypes, with recognition rates of 28% in the lower register and 8% in the upper register for fear, and 38% in the lower register and 13% in the upper register for anger.

CONCLUSION

The face of the soprano who volunteered for this experiment has a mixed salience index. This means that whether she is speaking or singing, the part of her face which determines the recognition of emotions depends on the emotion being expressed : the top of the face dominates for fear and anger, and the bottom dominates for joy and sorrow. Moreover, this soprano has a neutral expressivity index. In other words, the right side of her face is judged to be as expressive as the left, probably due to the regularity of her features.

Regardless of the medium, the emotions expressed in speech were identified better than those expressed in singing. The visual medium was the best, whereas sound was a poor vehicle of emotion information. The combination of sound and image led to a slight improvement in emotion recognition in speech, but did not turn out to be very effective in singing, where the partly destroyed acoustic emotion cues are unable to supply any additional information.

For the visual aspect of this study, the emotions expressed in the upper register appear as a whole to be overridden by those expressed in the lower register, judged to be more expressive. This finding is not surprising in that the entire facial musculature is involved in upper register emissions, which jeopardizes the expression of emotions. However, the upper register appears to be considered more expressive for joy, sorrow, and fear, most likely because of the wide buccal opening which reinforces the intensity of the emotion being expressed. The lower register in turn is judged to be more expressive for anger, no doubt due to the fact that the narrower buccal opening in the lower register corresponds more closely to this emotion, generally expressed by a set jaw.

For the auditory aspect of this study, the emotion identified the best was sorrow, although none of the subjects recognized it in the upper register where it was mistaken for joy. Due to the existence of strong emotional stereotypes, joy seems to be associated with the upper register, and sorrow, with the lower register. This accounts for the fact that sorrow was not recognized in the upper register and that joy was poorly identified in the lower register.

In summary, for singing, the identification of emotions and the assessment of their intensity appears to be influenced by parasitic phenomena related to their extreme production conditions. For instance, buccal opening plays a role in the assessment of emotional intensity, which is judged to be greater when the mouth is wide open (upper register) than when it is only slightly open (lower register). Likewise, the fact that unconsciously, most people associate joy with the upper register and sorrow with the lower register, has an impact on the recognition of these emotions in the outer registers.

Whereas in speech, the recognition of emotions is not subject to any particular production constraints, in singing the constraints are great and are manifested in the face by specific movements which interfere with the expression of emotions. Due to this fact, emotions are only correctly identified when the functional and expressive movements are compatible. In all other cases, the functional movement takes precedence over the emotional expression, because even when an opera singer is an excellent actor or actress, he or she cannot run the risk of jeopardizing the quality of the sound emitted in order to express an emotion or a feeling.

REFERENCES

[1] Sackheim, H.A. & Gur, R.C. (1982), Facial asymmetry and the Communication of Emotion, in *Social Psychophysiology*, Cacioppo J.T. & Petty R.E. (eds) Guilford Press, New York.

[2] Bruyer, R. (1980), Implication différentielle des hémisphères cérébraux dans les conduites émotionnelles, *Acta Psychiatrica Belgica*, 80, 266-284.

[3] Bassili, J.N. (1979), Emotion recognition : the role of facial movement and the relative importance of upper and lower areas of the face, *Journal of Personality and Social Psychology*, 37, 2049-2058.

ACKNOWLEDGMENTS

We would like express our sincere gratitude to all those who contributed to making this study come to be : Nathalie Labry lyrical artist, who so kindly and patiently complied with our scientific demands, Bernard Teston, C.N.R.S. research engineer, who designed part of the experimental equipment, Georges Bourrat and Louis Seimandi, University of Provence technicians, who skillfully ran the various phases of this extremely complex experiment, Robert Espesser, C.N.R.S. research engineer, who developed the software for the random auditory testing procedure, Claudette Bonnaud, head of the A.E.L, who lent us the projection equipment, Annie Poujol and Marie-Françoise Borel, in charge of the International Lyrical Arts Video Library in Aix-en-Provence, who with their usual friendliness and conscientiousness, allowed us to use the video libary auditorium and helped in recruiting part of the subjects, Vivian Waltz, who agreed to do the English translation in an amazingly short time, and finally, all the subjects, known and unknown, who were willing to devote some of their time to taking the various tests.