# The "Montreal Affective Voices": a validated set of nonlinguistic emotional vocal expressions for research on auditory affective processing

Pascal Belin[1,2], Sarah Fillion-Bilodeau[1], Frédéric Gosselin[1]

[1]Départment de Psychologie, Université de Montréal, Montréal, Québec, Canada

[2]Department of Psychology, University of Glasgow, Glasgow, Scotland, United Kingdom

## ABSTRACT

## INTRODUCTION

Past years have seen an increase in research on the neural correlates of paralinguistic aspects of voice perception, in particular affective processing. Thus far, the vast majority of the research on auditory affective processing has been conducted in the context of speech prosody, the "third element of language" (Monrad-Krohn, 1963). Typically, test material consists of speech material (words, sentences) spoken with various emotional tones (Adolphs, Tranel, & Damasio, 2001; Buchanan et al., 2000; Kotz et al., 2003; Mitchell, Elliott, Barry, Cruttenden, & Woodruff, 2003; Pell, 2005; Ross, 2000; Schirmer, Kotz, & Friederici, 2005).

Several limitations, however, are associated with the use of speech material to study auditory affective processing. One important limitation is the potential interaction between the affective tone of speech (emotional prosody) and the affective value carried by its semantic content. Different strategies have been used to attempt to minimize this interaction: controlling the affective value of the semantic content (Schirmer et al., 2005), for example by selecting "neutral" sentences (Imaizumi et al., 1997; Kotz et al., 2003; Laukka, 2005); eliminating the semantic content by using meaningless speech (Grandjean et al., 2005) or by employing acoustic manipulations such as low-pass filtering (Friend, 2000; McNally, Otto, & Hornig, 2001). Another important limitation of the use of speech stimuli is they are language-specific and thus cannot be used to compare results across different countries, for example to test for cross-cultural effects in auditory affective processing.

In contrast, research in affective processing in the visual modality is not subject to the same limitations. Studies indeed typically use nonverbal visual stimuli, such as the International Affective Picture System (P. Lang, Öhman, & Vaitl, 1988), or the set of affective faces by Ekman & Friesen (1978). The so-called "Ekman faces" consist of a standardised, validated set of photographs of the face of several actors portraying six discrete emotions (anger, disgust, fear, happiness, sadness, surprise) plus a neutral expression (P Ekman & Friesen, 1978) and do not convey any linguistic information. The "Ekman faces" have been introduced nearly three decades ago; they consist of greyscale, static stimuli that are relatively un-ecological compared to the colour, dynamic visual stimuli of the real world. Moreover, they sample a somewhat restricted set of emotions, based on a categorical account of facial expression of emotions that is still largely debated (P. J. Lang, 1995; Russell, Bachorowski, & Fernandez-Dols, 2003; Schlosberg, 1954). Nevertheless, the "Ekman faces" are still widely used in cognitive neuroscience research (Calder, Burton, Miller, Young, & Akamatsu, 2001; Krolak-Salmon, Fischer, Vighetto, & Mauguiere, 2001; Morris et al., 1996; Smith, Cottrell, Gosselin, & Schyns, 2005; Vuilleumier, Armony, Driver, & Dolan, 2001; Young et al., 1997). Two important advantages may contribute to this popularity: they are non-linguistic and thus can be used in several different countries (potentially allowing cross-cultural comparisons); several different actors portray the same emotions, allowing to use several different stimuli for each discrete emotion and avoiding potential confounds with actor's identity.

We argue that the same advantages can be obtained in the auditory modality by using nonlinguistic emotional interjections - as shown by recent studies (Morris, Scott, & Dolan, 1999; Sander & Scheich, 2001, 2005). . Nonlinguistic emotional interjections such as screams of fear are vocal expressions that usually accompany very intense emotional feelings - along with the corresponding facial expressions. They are closely parallel to animal affect vocalizations (Scherer, 1995), and generally sound similar in different cultures - although this issue has been less explored than in the domain of facial expressions probably because of the lack of appropriate material.Here we present a standardized set of emotional vocal expressions designed to constitute the auditory equivalent of the "Ekman faces" and avoid the potential confound of linguistic content. The "Montreal affective voices" consist of seventy short, nonlinguistic interjections expressing anger, disgust, fear, happiness, sadness, and surprise (plus a neutral expression) recorded in ten different actors. This set of vocalizations was validated based on ratings of valence, arousal, and perceived intensity along the six discrete emotions provided by a group of thirty judges.

## METHODS

### Recording

*Participants*

Twenty-two amateur or professional actors participated in the recording sessions after giving written informed consent. They received a compensation of CA$ 20 per hour of recording.

*Procedure*

Actors were instructed to produce short emotional interjections using the French vowel «ah», and were played an auditory demonstration of the expressions they would be asked to generate before the recording session. They had to produce vocal expressions corresponding to: happiness, sadness, fear, anger, pleasure, pain, surprise, disgust as well as a neutral expression. Each category of vocalizations was performed several times until our qualitative criterion was reached, i.e. until the affective vocalization produced was clearly recognizable by the experimenter as the one they were asked to produce. A short practice session was performed at the beginning of each recording bout for each emotion during which the sound level was adjusted. Constant feedback was given to the participants during the entire session so they could improve their performance.

Vocalizations were recorded in the sound-proof room of the Vocal Neurocognition laboratory (University of Montréal) using a UMT800 condenser microphone (Microtech Geffell) at a distance of approximately 30 cm. Recordings were pre-amplified using a Millenia Media HV3B pre-amplifier, and digitized at a 96 kHz sampling rate and 16-bits resolution. using an Audiophile 2496 PCI sound card (M-audio). They were then edited in short meaningful segments and normalized peak value (90%) and downsampled at 44.1 kHz using Adobe Audition (Adobe Systems Inc.). For each actor and vocalization category, only the best occurrence was kept for the validation stage.

**Validation**

*Participants*

Thirty francophone participants (15 males, 15 females) were recruited (average age 23.3 +/- 3 years old) through notices posted at the University of Montréal. Each participant gave written informed consent and filled a socio-demographic information sheet prior to the judgment phase. They were compensated CA$ 10 per hour for their participation.

*Procedure*

Each participant was instructed to evaluate each of the 198 vocalizations (22 actors x 9 categories) on ten different rating scales: the perceived emotional valence (from extremely negative to extremely positive), the actor's perceived arousal (from not at all aroused to extremely aroused), and perceived intensity in each of 8 rating scales corresponding to the eight targeted affective states: happiness, sadness, fear, anger, surprise, disgust, pleasure, pain (from not at all angry to extremely angry, for example). For each sound they had to judge, participants were played the sound and displayed a judgment board on a computer screen, consisting of a small speaker icon at the top of the screen and ten horizontal visual analogue scales. Each scale consisted of an identical unmarked horizontal line with verbal labels at the left and right extremities (e.g., for the Arousal scale: not at all aroused on the left; extremely aroused on the right). Participants could hear at will the sound they were judging by clicking on the speaker icon. Each of the ten ratings was performed by clicking with the computer mouse on the point of the scale corresponding to the intended judgment. All ten judgments had to be done before the next sound was played and the screen was displaying a blank judgment board. Ratings along the visual analogue scales were linearly converted to an integer number ranging from 0 to 100.

Stimuli were presented in a pseudo-randomized order in 4 blocks (2 blocks of 50 stimuli and 2 blocks of 49 stimuli) at a self-adjusted comfortable level over DT770 headphones (Beyerdynamics). During the session, the participants could take breaks at will between blocks.

**Selection**

Ten actors (5 males and 5 females) have been selected out of the 22 actors based on the results of the listener's judgments as those who produced the most unambiguous set of affective vocalizations. The vocalizations of pain and pleasure were removed from the set as they led to too many ambiguous judgments; the following descriptive and inferential statistics are thus only based on the remaining six emotions (plus the neutral one) and the ratings on the corresponding scales. Physical characteristics of the ten selected actors are summarized in Table 1.

**Acoustical analyses**

Acoustic characteristics of the vocalizations were measured using Straight (Kawahara, Katayose, de Cheveigne, & Patterson, 1999) and Praat (www.praat.org). They included: minimum, maximum, median and standard deviation of the fundamental frequency (f0) measured over the voiced portions (in Hertz); sound duration (in milliseconds); median and standard deviation of intensity (in decibels). These characteristics, averaged across the ten actors, are given for each vocalization category in Table 2. The complete set of acoustic measures can be found in Appendix 1.

**Statistical analyses**

The effects of judge's and actor's gender on ratings of perceived valence, arousal and intensity in the scale corresponding to each of the six basic emotions (Anger, Disgust, Fear, Happiness, Sadness and Surprise) were investigated for each vocalization category. Eighteen 2X2 mixed-design ANOVAS were performed with ratings averaged across the five actors of a same gender as the dependant variable, actors' gender as the repeated within-subject factor, and judge's gender as the between-subject factor.

Correction for multiple tests were performed using Bonferroni's correction: effects were considered to be significant if they reached a threshold of $p<(0.05/18)= 0.0027$.

**RESULTS**

The “Montreal affective vocalizations” consist of seventy nonlinguistic vocal expressions of anger, disgust, fear, happiness, sadness, and surprise (plus a neutral expression) recorded in ten different actors. They re available at the URL: ftp://132.204.126.245/. Figure 1 shows the waveforms and spectrograms of the 70 vocalizations.

*Acoustic characteristics*

Table 2 summarizes the acoustic characteristics of the Montreal affective voices. They are characterized by a substantial degree of variation between actors, but the values averaged across actors show important but consistent differences between vocalizations categories, in line with the existing literature (Juslin & Laukka, 2003; Scherer, 1986, 1995).

The neutral vocalizations were characterized by the lowest averaged values of median and maximum f0, and the standard deviation of their f0 and intensity was much smaller than for other vocalization categories, consistent with the stable pitch and intensity of a sustained vowel. The angry vocalizations were characterized by a high f0 variation, and the highest intensity variation. The disgusted vocalizations were those with the lowest averaged minimum f0 and a very low maximum f0, with a small f0 variation compared to the other affective vocalizations. The fearful vocalizations were relatively brief, and were those with the highest median f0, maximum f0, and f0 variation. They were also the category of affective with the highest averaged median intensity. The vocalizations of happiness (laughs) were relatively long, and with a low averaged f0 variation comparable to that of the disgusted vocalizations. They were characterized by the lowest median intensity values and large variations in intensity due to the silent intervals between the laugh bursts. The sad vocalizations were the longest ones, with relatively high values of maximum f0, and a low averaged median intensity comparable to that of the laughs. Finally, the surprised vocalizations were on average much shorter than the other vocalizations, and were characterized by large values of minimal and median f0.

*Emotional ratings*

Ratings along the eight visual analogue affective scales, averaged across the 30 participants and the 10 actors, are reported in Table 3 for each category of vocalization. Consistent with the choice of an unambiguous set of vocalizations, each vocalization category obtained significantly higher average ratings for the rating scale evaluating the corresponding discrete emotion (see shaded cells in Table 3) than to all other rating scales (Fisher's protected LSD, column-wise). A fair degree of variation can be noted, though: most vocalizations categories (e.g., happiness) obtained very significantly higher ($p<.0001$) average ratings in the corresponding compared to other scales. This was less true for the fearful vocalizations that obtained relatively high ratings on the Surprise scale. This could indicate a potentially high degree of confusion by the listeners for some of the vocalizations. Alternatively, this could indicate that some of the vocalizations express a mixture of several discrete emotions, which caused the listeners to give high ratings to more than one scale.

*Decoding accuracy*

These two alternative explanations can be partly disambiguated by taking into account only the dominant perceived emotion, i.e., the rating scale with maximal value for each vocalization and listener. Table 4 reports the proportion of judgments (across the 70 vocalizations X 30 participants) for which each scale corresponding to the six discrete emotions obtained maximal value. For each stimulus, the highest rating emotion scale was coded as 1, and the other emotion scales were rated as 0. If two or more scales obtained the maximum value, they were coded as 1 divided by the number of scales obtaining the maximum score. As expected, these proportion values are highest for the scales corresponding to the portrayed emotion (shaded areas along the diagonal in Table 4). Some vocalization categories yielded very high proportion of maximal ratings in the corresponding scale (e.g., happiness), indicating an unambiguous recognition of the dominant portrayed emotion by most listeners. But other categories (e.g., fear and surprise) yielded rather low proportions of maximal rating, and high proportion of maximal ratings in scales other than the corresponding one, indicating confusion by the listeners.

Ratings obtained on each scale and effects of actor's and judge's gender on these ratings are discussed below for each category of affective vocalization. These results are summarized in Figure 2.

*Anger*

Angry vocalizations were considered as the most negative, yielding the lowest average ratings on the Valence scale (Table 3). Angry vocalizations were also judged on average as conveying the highest level of arousal. The average rating on the Angriness scale was 75, which was three times as high as the second highest score obtained on the Surprised scale. Angry vocalizations obtained the maximal rating in the Angriness scale in 81% of the judgments, and were perceived as conveying more surprise than anger in only 11% of the judgments (Table 4). There was no significant effect of actor's gender on ratings on the Valence ($F(1,28)=3.453$, $p>.07$), Arousal ($F(1,28)=8.232$, $p>.005$), or Anger scales ($F(1,28)<2.502$, $p>.125$), nor any effect of judge's gender (all $F(1,28)<1.54$, $p>.22$).

*Disgust*

Disgusted vocalizations were rated as negative (below 50), although less than fear or sadness. A significant main effect of actor's gender on the Valence scale was found ($F(1,28)=12.905$, $p=.001$), but no main effect of judge's gender ($F(1,28)=.276$, $p>.5$) and no interaction between the two factors ($F(1,28)=1.53$, $p>.22$), indicating that both male and female judges gave lower valence ratings to the vocalizations from the female (21±16) than the male (29±13) actors (Fig. 2). The disgusted vocalizations were judged as conveying a moderate level of arousal (Table 3), comparable to that of the neutral vocalizations, with no significant effect of actor's ($F(1,28)=7.577$, $p=0.01$) or judge's ($F(1,28)=.33$, $p>.5$) gender. Average ratings were much higher on the Disgust scale than other scales. The second highest rating was again obtained on the Surprise scale (Table 3). The proportion of maximal ratings for the disgusted vocalizations was quite comparable to that of the angry vocalizations: above 80% of the maximal ratings on the Disgust scale, followed by 9% for surprise (Table 4). A significant main effect of actor's gender was observed for the ratings on the Disgust scale ($F(1,28)=51.739$, $p<.001$), but no effect of judge's gender ($F(1,28)=.481$, $p>.4$) and no interaction between the two factors ($F(1,28)=.332$, $p>.5$). As shown in Figure 2, both male and female judges rated the male vocalizations lower (60±22) than the female vocalizations (79±16).

*Fear*

The fearful vocalizations were rated as negative, with an average valence below 50 and identical to that of the disgusted vocalizations. The effect of actor's gender on the valence ratings was just above significance ($F(1,28)= 10.429$, $p=0.003$), with no effect of judge's gender ($F(1,28)=2.011$, $p>.15$) and no interaction between the two factors ($F(1,28)=.493$, $p>.4$), suggesting a trend for lower perceived valence for the female (22±8) than the male (27±8) screams. The fearful vocalizations were judged on average as conveying the highest level of arousal, comparable to that of angry vocalizations (Table 3). A significant effect of actor's gender on ratings on the Arousal scale was found ($F(1,28)=60.841$, $p<0.001$), but no effect of judge's gender ($F(1,28)=.025$, $p>.5$) and no interaction ($F(1,28)=.294$, $p>.5$), indicating that both male and female listeners judged female fearful vocalizations as conveying more arousal (78±11) than male vocalizations (65±14). Average ratings on the scales evaluating the perceived intensity of the six basic emotions were highest on the Fear scale, but they also yielded a very high score on the Surprise scale, indicating a potentially high degree of confusion between fear and surprise. This was confirmed by the proportion of maximal ratings which attained 34% on the Surprise scale (Table 4). Thus, in about one case out of three, fearful vocalizations were perceived as expressing more surprise than fear. Ratings on the Fear scale also showed a significant effect of actor's gender ($F(1,28)=12.123$, $p=0.002$), but no significant effect of judge's gender($F(1,28)=.096$, $p>.5$), indicating overall stronger ratings of perceived fear in vocalizations produced by the female (74±15) than the male

(62±17) actors. However, the interaction with judge's gender approached significance ($F(1,28)=9.140$, $p=0.005$), indicating that this difference was mostly due to the female participants as shown in Fig. 2.

*Happiness*

As expected, the happy vocalizations (laughs) were perceived as the most positive, yielding the highest average valence (Table 3). They were judged as conveying a moderate degree of arousal, intermediate between the neutral and angry vocalizations. The happy vocalizations obtained average ratings on the happiness scale that were the highest across all emotional categories and scales. They obtained a higher rating in the happiness scale than in other scales in 99% of the judgments, indicating a highly unambiguous perception. No significant effect of actors' gender (all $F(1,28)<1.711$, $p>.2$) or judge's gender (all $F(1,28)<1.013$, $p>.3$) were observed on ratings on the Valence, Arousal, or Happiness scales.

*Sadness*

Unsurprisingly, the sad vocalizations (cries) were judged as very negative, with a low average perceived valence comparable to that of the angry vocalizations. A significant effect of actor's gender on the valence ratings was observed ($F(1,28)=62.865$, $p<.001$), but no effect of judge's gender ($F(1,28)=2.056$, $p>.15$) and no interaction ($F(1,28)=2.104$, $p>.15$). As shown in Fig. 2, both male and female judges rated cries from the female actors with lower valence (11±9) than those from the male actors (28±12). The sad vocalizations were judged to convey a relatively low level of arousal, with no effect of actor's ($F(1,28)=2.169$, $p>.15$) or judge's ($F(1,28)=.186$, $p>.5$) gender. Average ratings were very high on the Sadness scale, and much lower on the other scales. A significant effect of actor's gender was also observed for the ratings on the Sadness scale ($F(1,28)=54.335$, $p<.001$), with no effect of judge's gender ($F(1,28)=.824$, $p>.3$) and no interaction ($F(1,28)=.657$, $p>.4$): both male and female judges rated cries from the female actors more strongly on the Sadness scale (87±14) than those from the male actors (68±18). The cries obtained highest ratings on the sadness scale in a very high proportion of the judgments (88%). Interestingly however, they were confounded with laughs in 10% of the cases.

*Surprise*

The surprised vocalizations were perceived as slightly negative, although much less than the angry, sad of fearful vocalizations. No significant effect of actors' ($F(1,28)=2.211$, $p>.14$) or judge's($F(1,28)=.081$, $p>.5$) gender was found for the ratings on the Valence scale. The surprised vocalizations were judged as conveying a high degree of arousal, comparable to that of the fearful and angry vocalizations. A significant effect of actor's gender was observed ($F(1,28)=38.292$, $p>.001$) but no effect of judge's gender ($F(1,28)=1.193$, $p>.25$) and no interaction ($F(1,28)=.085$, $p>.5$), indicating that both male and female judges gave higher ratings to surprised vocalizations from the female (75±11) than from the male (65±12) actors on the Arousal scale. Surprised vocalizations obtained highest average ratings on the Surprise scale, with no significant effect of actor's ($F(1,28)=7.712$, $p=0.01$) or judge's ($F(1,28)=7.073$, $p=.013$) gender, but relatively high scores on the other scales as well compared to other categories of vocalizations, particularly Fear (45). This probably indicates that the surprised vocalizations were perceived as expressing a mix of surprise and other emotions. The surprised vocalizations obtained maximal rating on the Surprise scale in 77% of the judgments, but also a fairly high degree (16%) of maximal ratings on the Fear scale. This again suggests a pattern of confusion between the fearful and the surprised vocalizations in the judges.

## DISCUSSION

Seventy affective nonlinguistic vocalizations, corresponding to the expression of six basic emotions plus a neutral one recorded in standardized conditions in ten actors, were validated based on affective ratings performed by a group of thirty judges.

All categories of affective vocalizations were well recognized by the judges, as illustrated by the significantly higher intensity ratings obtained on the scale corresponding to the portrayed emotion than on scales corresponding to other emotions (Table 3). The very high decoding accuracy of the judges observed for most categories (Table 4) confirms this good recognition. The strongest pattern of confusion was observed between the fearful and surprised vocalizations: nearly a third of the ratings for the fearful vocalizations yielded higher ratings on the Surprise scale than on the Fearful scale. This confusion between fear and surprise is also frequently observed for human faces in some cultures (P. Ekman, Friesen, & Ellsworth, 1972) as well as by an ideal observer classifying affective faces (Smith et al., 2005), suggesting it might not be specific to vocal expressions.

Ratings of perceived valence, arousal and intensity along the six basic emotions were highly consistent between the groups of male and female judges (Fig. 2). No significant effect of judge gender was observed in any vocalization category, nor any significant interaction with actor's gender, on ratings of perceived valence, arousal, or intensity in the corresponding rating scale. This suggests that the obtained ratings were consistently related to affective properties of the vocalizations.

However, a significant effect of actor's gender was found for several ratings indicating that for some vocalization categories, male and female vocalizations were perceived differently. As shown in Figure 2, this was the case for the disgusted vocalizations - rated as more negative and intense in the female actors-, the fearful vocalizations - rated as more intense and expressing more arousal in female vocalizations-, the cries - more negative in female vocalizations- and the surprised vocalizations - again more intense and conveying more arousal in the female vocalizations. It is striking that in all cases the gender effect was observed in the same direction of stronger expressions in the female than in the male actors. Whether this effect is particular to the selected group of actors or reflects a more general gender difference in vocal emotional expression remains to be explored.

Using nonlinguistic affective vocalizations in studies of auditory emotional processing presents several advantages. First these interjections do not contain any semantic information, and so they are not subject to the problems of emotional speech described above. They are not limited by linguistic barriers and so can be used to compare results in different countries and test for cross-cul;tural differences. Moreover, they are more primitive expressions of emotion, closer to affect expressions of animals or human babies than emotional speech, thus potentially allowing better cross-species or human developmental comparisons. They are also much more similar than emotional speech to stimuli used in the study of affective processing in the visual modality, such as the "Ekman faces", thus allowing better comparisons across modalities as well as studies of cross-modal emotional integration.

These reasons have led researchers to increasingly use nonlinguistic affective interjections in recent cognitive or neuroimaging studies (Morris et al., 1999; Sander & Scheich, 2001, 2005). These studies reflect a valuable effort to study auditory affective processing outside the linguistic domain. Yet these researchers typically used their own custom material, limited by the use of only few emotions- one or two- recorded from only one or few speakers, which potentially leading to problems such as interactions with speaker's identity, or mood induction caused by the repeated presentation of the same emotional expression. The "Montreal affective voices", as the auditory equivalent of the Ekman faces, do not present these limitations because they consist of several emotions recorded in standardized conditions in several actors. We believe they constitute a useful tool for research in the cognitive neuroscience of auditory affective processing.

**Table 1. Physical characteristics of the actors**

| actor | gender | age | height | weight |
|---|---|---|---|---|
| 6 | m | 38 | 6'1 | 178 |
| 42 | m | 22 | 5'10" | 190 |
| 45 | f | 22 | 5'8" | 145 |
| 46 | f | 27 | 5'3" | 125 |
| 53 | f | 20 | 5'6" | 120 |
| 55 | m | 22 | 5'10" | 160 |
| 58 | f | 25 | 5'7" | 115 |
| 59 | m | 27 | 5'8" | 175 |
| 60 | f | 24 | 5' | 95 |
| 61 | M | 28 | 5'9" | 160 |

Physical characteristics are reported for the ten selected actors. The first column indicates the actor's label, which is found at the beginning of the filename for each of the seven sounds produced by an actor.  M; male; f: female. Weight is indicated in pounds.

**Table 2: Acoustic characteristics**

| Vocalizations | Neutral | Anger | Disgust | Fear | Happiness | Sadness | Surprise |
|---|---|---|---|---|---|---|---|
| min f0 (Hz) | 149 | 150 | 108 | 266 | 181 | 185 | 228 |
| max f0 (Hz) | 184 | 413 | 295 | 642 | 421 | 508 | 453 |
| median f0 (Hz) | 168 | 317 | 200 | 508 | 278 | 323 | 373 |
| st.dev. f0 (Hz) | 4 | 80 | 58 | 97 | 58 | 73 | 69 |
| Duration (msec) | 1024 | 924 | 977 | 603 | 1446 | 2229 | 385 |
| median intensity (dB) | 81.4 | 77.9 | 74.8 | 81.1 | 59.5 | 63.3 | 75.8 |
| st.dev intensity (dB) | 6.0 | 13.9 | 11.6 | 11.8 | 13.6 | 12.9 | 13.4 |

Summary of the acoustic characteristics of each vocalizations category (averaged across the 10 actors). f0 values are expressed in Hertz, duration in milliseconds. RMS energy is in arbitrary units, with peak value of each sound equal to 0.9.

**Table 3: Average emotional ratings**

| | Portrayed emotion | | | | | | |
|---|---|---|---|---|---|---|---|
| Rating scale | Neutral | Anger | Disgust | Fear | Happiness | Sadness | Surprise |
| Valence | 47 | 16 | 24 | 24 | 85 | 18 | 39 |
| Arousal | 32 | 72 | 36 | 72 | 57 | 44 | 70 |
| Anger | 8 | 75 ** | 13 | 18 | 3 | 8 | 16 |
| Disgust | 9 | 22 | 69 ** | 20 | 4 | 8 | 23 |
| Fear | 8 | 15 | 11 | 67 * | 3 | 9 | 45 |
| Happiness | 14 | 5 | 9 | 6 | 81 ** | 10 | 15 |
| Sadness | 10 | 12 | 9 | 12 | 2 | 77 ** | 10 |
| Surprise | 9 | 24 | 25 | 57 | 18 | 10 | 77 ** |

Cells indicate ratings (on a 0-100 visual analogue scale, cf. Methods) averaged across actors (n=10) and judges (n=30) for each rating scale (lines) and for each of the seven affective vocalizations (rows). Shaded areas show that maximal average rating for each scale are indeed obtained for the corresponding portrayed emotions. Means with stars in the shaded areas indicate that the portrayed emotions received a rating on the corresponding scale that was significantly different (Fisher's protected least significance difference test) from the ratings on the other scales. * p<.01, ** p<.0001

**Table 4: Decoding accuracy**

| | Portrayed emotion | | | | | |
|---|---|---|---|---|---|---|
| Rating scale | Anger | Disgust | Fear | Happiness | Sadness | Surprise |
| Anger | 80 | 4 | 5 | 0 | 0 | 4 |
| Disgust | 6 | 81 | 4 | 1 | 0 | 5 |
| Fear | 3 | 0 | 57 | 0 | 1 | 14 |
| Happiness | 1 | 6 | 1 | 98 | 10 | 2 |
| Sadness | 0 | 1 | 2 | 0 | 88 | 1 |
| Surprise | 10 | 8 | 31 | 0 | 1 | 74 |

Cells indicate decoding accuracy, i.e., proportion of judgments with maximum rating in each scale, for each vocalization category (neutral vocalization excluded). See text for a detailed description of the procedure.

**Figure 1**

Waveforms and spectrograms (0-8000 Hz) of the 70 vocalizations.

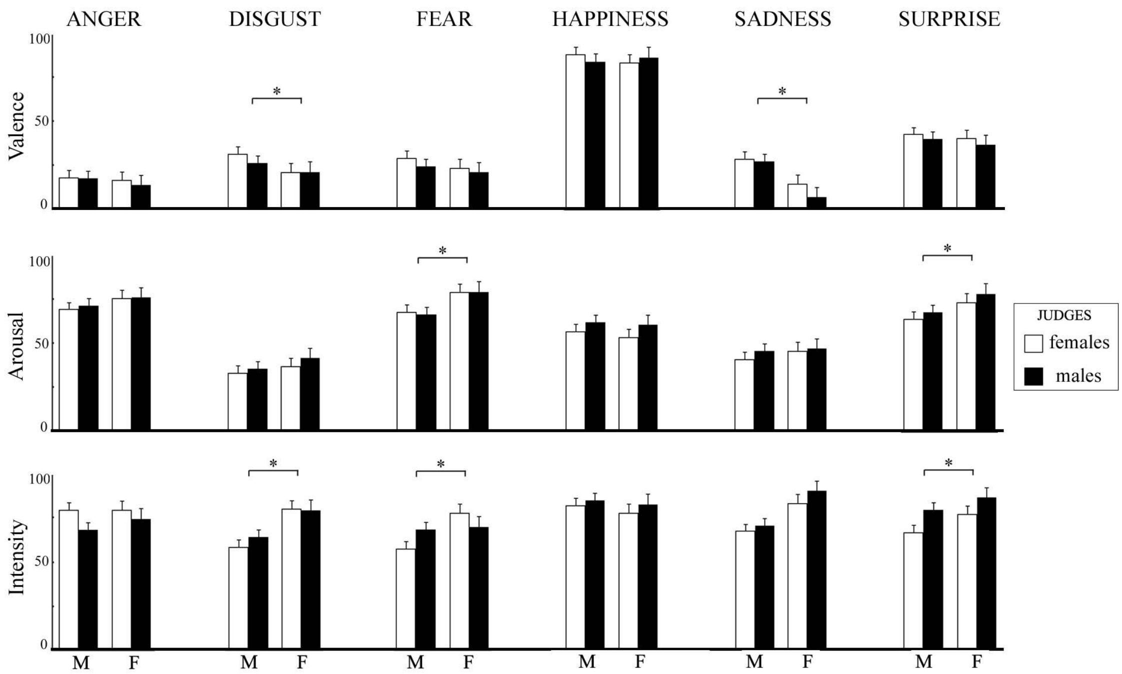


**Figure 2**

Average ratings for each vocalizations category on the corresponding scale, split by actor's gender (F: female; M: male) and by judge gender (dark bars: female judges; white bars: male judges). Stars indicate significant rating differences between the vocalizations produced by the male and female actors.

| Sound filename | min f0 | max f0 | median f0 | sd f0 | duration | power | sd power |
|---|---|---|---|---|---|---|---|
| '6_anger.wav' | 73 | 290 | 151 | 67 | 1142 | 77.7 | 13.9 |
| '6_disgust.wav' | 115 | 196 | 158 | 25 | 1051 | 63.1 | 13.4 |
| '6_fear.wav' | 129 | 537 | 317 | 93 | 761 | 80.2 | 10.5 |
| '6_happiness.wav' | 144 | 343 | 225 | 42 | 1742 | 47.5 | 15.9 |
| '6_neutral.wav' | 91 | 116 | 113 | 4 | 896 | 80.9 | 5.6 |
| '6_sadness.wav' | 201 | 336 | 262 | 25 | 1643 | 57.9 | 12.6 |
| '6_surprise.wav' | 203 | 303 | 275 | 30 | 265 | 71.5 | 16.2 |
| '42_anger.wav' | 74 | 267 | 138 | 63 | 888 | 64.7 | 15.0 |
| '42_disgust.wav' | 88 | 195 | 161 | 33 | 1045 | 77.4 | 13.6 |
| '42_fear.wav' | 149 | 313 | 289 | 50 | 405 | 81.3 | 14.8 |
| '42_happiness.wav' | 139 | 223 | 157 | 20 | 1445 | 52.6 | 13.5 |
| '42_neutral.wav' | 102 | 116 | 112 | 2 | 1312 | 78.6 | 4.5 |
| '42_sadness.wav' | 132 | 233 | 181 | 30 | 1667 | 51.1 | 13.9 |
| '42_surprise.wav' | 108 | 307 | 252 | 61 | 583 | 72.7 | 14.7 |
| '45_anger.wav' | 150 | 498 | 402 | 100 | 949 | 79.5 | 12.5 |
| '45_disgust.wav' | 185 | 545 | 391 | 115 | 607 | 78.9 | 8.8 |
| '45_fear.wav' | 300 | 653 | 629 | 87 | 628 | 80.9 | 10.7 |
| '45_happiness.wav' | 312 | 497 | 355 | 51 | 1563 | 50.6 | 13.7 |
| '45_neutral.wav' | 222 | 253 | 228 | 3 | 992 | 83.7 | 5.5 |
| '45_sadness.wav' | 251 | 815 | 519 | 171 | 1780 | 65.7 | 9.8 |
| '45_surprise.wav' | 452 | 913 | 826 | 150 | 284 | 76.4 | 16.0 |
| '46_anger.wav' | 357 | 589 | 532 | 57 | 421 | 83.0 | 16.1 |
| '46_disgust.wav' | 93 | 358 | 221 | 96 | 1566 | 70.6 | 11.3 |
| '46_fear.wav' | 375 | 1658 | 926 | 344 | 815 | 82.3 | 12.8 |
| '46_happiness.wav' | 189 | 584 | 231 | 101 | 1009 | 61.6 | 13.8 |
| '46_neutral.wav' | 209 | 289 | 260 | 11 | 240 | 83.6 | 14.8 |
| '46_sadness.wav' | 331 | 661 | 433 | 88 | 1956 | 70.4 | 11.2 |
| '46_surprise.wav' | 446 | 469 | 463 | 6 | 404 | 79.2 | 10.9 |
| '53_anger.wav' | 166 | 517 | 417 | 113 | 1518 | 83.1 | 17.5 |
| '53_disgust.wav' | 143 | 253 | 213 | 31 | 1714 | 77.1 | 12.4 |
| '53_fear.wav' | 274 | 477 | 467 | 49 | 835 | 84.5 | 9.9 |
| '53_happiness.wav' | 169 | 325 | 248 | 42 | 960 | 65.9 | 12.2 |
| '53_neutral.wav' | 160 | 196 | 190 | 3 | 946 | 83.2 | 4.6 |
| '53_sadness.wav' | 160 | 537 | 302 | 37 | 2877 | 73.8 | 13.6 |
| '53_surprise.wav' | 208 | 405 | 329 | 62 | 382 | 73.9 | 14.7 |
| 55_anger.wav' | 100 | 259 | 222 | 48 | 527 | 80.8 | 10.2 |
| '55_disgust.wav' | 63 | 252 | 169 | 57 | 672 | 80.2 | 10.3 |
| 55_fear.wav' | 204 | 302 | 284 | 23 | 614 | 80.8 | 10.0 |
| '55_happiness.wav' | 146 | 280 | 217 | 34 | 1100 | 67.6 | 12.5 |
| '55_neutral.wav' | 106 | 130 | 109 | 2 | 1236 | 77.3 | 3.7 |
| '55_sadness.wav' | 150 | 309 | 249 | 39 | 1830 | 69.4 | 13.8 |
| '55_surprise.wav' | 73 | 281 | 228 | 61 | 263 | 78.5 | 13.0 |
| '58_anger.wav' | 160 | 468 | 407 | 103 | 715 | 80.6 | 15.0 |
| '58_disgust.wav' | 143 | 295 | 214 | 47 | 978 | 72.5 | 13.8 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| '58_fear.wav' | 333 | 452 | 418 | 23 | 489 | 75.3 | 14.1 |
| '58_happiness.wav' | 197 | 523 | 299 | 61 | 1046 | 66.8 | 9.6 |
| '58_neutral.wav' | 184 | 222 | 211 | 9 | 511 | 82.5 | 4.5 |
| '58_sadness.wav' | 186 | 542 | 379 | 90 | 1416 | 65.4 | 13.5 |
| '58_surprise.wav' | 235 | 441 | 382 | 56 | 329 | 78.9 | 11.8 |
| '59_anger.wav' | 131 | 377 | 336 | 64 | 1184 | 83.3 | 9.3 |
| '59_disgust.wav' | 72 | 243 | 152 | 52 | 710 | 78.2 | 9.7 |
| '59_fear.wav' | 118 | 359 | 324 | 53 | 719 | 85.0 | 12.2 |
| '59_happiness.wav' | 179 | 594 | 466 | 95 | 1831 | 64.5 | 16.7 |
| '59_neutral.wav' | 139 | 197 | 143 | 5 | 645 | 84.3 | 5.5 |
| '59_sadness.wav' | 198 | 773 | 404 | 132 | 4310 | 53.7 | 13.8 |
| '59_surprise.wav' | 129 | 475 | 304 | 109 | 574 | 75.8 | 15.9 |
| 60_anger.wav' | 159 | 516 | 301 | 113 | 1082 | 75.4 | 13.3 |
| '60_disgust.wav' | 136 | 422 | 217 | 90 | 838 | 79.0 | 10.8 |
| 60_fear.wav' | 625 | 1158 | 1067 | 168 | 440 | 82.6 | 11.4 |
| '60_happiness.wav' | 253 | 665 | 430 | 106 | 1159 | 68.0 | 15.4 |
| '60_neutral.wav' | 193 | 222 | 214 | 3 | 1597 | 81.0 | 5.8 |
| '60_sadness.wav' | 154 | 662 | 345 | 95 | 2376 | 67.2 | 9.9 |
| '60_surprise.wav' | 343 | 707 | 485 | 116 | 253 | 78.3 | 10.4 |
| 61_anger.wav' | 130 | 352 | 262 | 68 | 815 | 71.2 | 16.4 |
| '61_disgust.wav' | 44 | 192 | 109 | 37 | 584 | 70.5 | 11.3 |
| 61_fear.wav' | 152 | 514 | 358 | 80 | 319 | 78.0 | 11.6 |
| '61_happiness.wav' | 78 | 178 | 153 | 28 | 2605 | 50.4 | 13.2 |
| '61_neutral.wav' | 85 | 101 | 95 | 1 | 1861 | 78.9 | 5.6 |
| '61_sadness.wav' | 84 | 211 | 159 | 25 | 2438 | 58.3 | 17.2 |
| '61_surprise.wav' | 82 | 225 | 185 | 43 | 514 | 73.0 | 10.8 |

**Appendix**: acoustic characteristics of all 70 sounds. Power indicates median power (dB)

**REFERENCES**


Adolphs, R., Tranel, D., & Damasio, H. (2001). Emotion recognition from faces and prosody following temporal lobectomy. *Neuropsychology, 15*((3)), 396-404.

Buchanan, T. W., Lutz, K., Mirzazade, S., Specht, K., Shah, N. J., Zilles, K., et al. (2000). Recognition of emotional prosody and verbal components of spoken language: an fMRI study. *COGNITIVE BRAIN RESEARCH, 9*((3)), 227-238.

Calder, A. J., Burton, A. M., Miller, P., Young, A. W., & Akamatsu, S. (2001). A principal component analysis of facial expressions. *Vision Res, 41*((9)), 1179-1208.

Ekman, P., & Friesen, W. V. (1978). Facial Action Coding: Consulting Psychologists Press Inc.

Ekman, P., Friesen, W. V., & Ellsworth, P. (1972). *Emotion in the human face: Guidelines for research and an integration of findings*. Oxford: Pergamon Press.

Friend, M. (2000). Developmental changes in sensitivity to vocal paralanguage. *Developmental Science*, 148-162.

Grandjean, D., Sander, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., et al. (2005). The voices of wrath: brain responses to angry prosody in meaningless speech. *Nat Neurosci, 8*, 145-146.

Imaizumi, S., Mori, K., Kiritani, S., Kawashima, R., Sugiura, M., Fukuda, H., et al. (1997). Vocal identification of speaker and emotion activates different brain regions. *Neuroreport, 8*((12)), 2809-2812.

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: different channels, same code? *Psychol Bull, 129*, 770-814.

Kawahara, H., Katayose, H., de Cheveigne, A., & Patterson, R. D. (1999). *Fixed Point Analysis of Frequency to Instantaneous Frequency Mapping for Accurate Estimation of F0 and Periodicity.* Paper presented at the EUROSPEECH'99.

Kotz, S. A., Meyer, M., Alter, K., Besson, M., von Cramon, D. Y., & Friederici, A. D. (2003). On the lateralization of emotional prosody: an event-related functional MR investigation. *Brain Lang, 86*, 366-376.

Krolak-Salmon, P., Fischer, C., Vighetto, A., & Mauguiere, F. (2001). Processing of facial emotional expression: spatio-temporal data as assessed by scalp event-related potentials. *Eur J Neurosci, 13*((5)), 987-994.

Lang, P., Öhman, A., & Vaitl, D. (1988). *The international affective picture system*. Gainsville: The Center for Research in Phycophysiology, University of Florida.

Lang, P. J. (1995). The emotion probe. Studies of motivation and attention. *Am Psychol, 50*, 372-385.

Laukka, P. (2005). Categorical perception of vocal emotion expressions. *Emotion., 5*, 277-295.

McNally, R. J., Otto, M. W., & Hornig, C. D. (2001). The voice of emotional memory: content-filtered speech in panic disorder, social phobia, and major depressive disorder. *Behav Res Ther, 39*, 1329-1337.

Mitchell, R. L., Elliott, R., Barry, M., Cruttenden, A., & Woodruff, P. W. (2003). The neural response to emotional prosody, as revealed by functional magnetic resonance imaging. *Neuropsychologia, 41*, 1410-1421.
Monrad-Krohn, G. H. (1963). The third element of speech: prosody and its disorders. In L. Halpern (Ed.), *Problems of dynamic neurology* (pp. 101-117). Jerusalem: Hebrew University Press.
Morris, J. S., Frith, C. D., Perrett, D. I., Rowland, D., Young, A. W., Calder, A. J., et al. (1996). A differential neural response in the human amygdala to fearful and happy facial expressions. *Nature, 383*, 812-815.
Morris, J. S., Scott, S. K., & Dolan, R. J. (1999). Saying it with feeling: neural responses to emotional vocalizations. *NEUROPSYCHOLOGIA, 37*((10)), 1155-1163.
Pell, M. D. (2005). Cerebral mechanisms for understanding emotional prosody in speech. *Brain Lang, e-pub ahead of print*.
Ross, E. D. (2000). Affective prosody and the aprosodias. In M.-M. Mesulam (Ed.), *Principles of behavioral and cognitive neurology* (pp. 316-331). New York, NY, US: Oxford University Press.
Russell, J. A., Bachorowski, J.-A., & Fernandez-Dols, J.-M. (2003). Facial and vocal expressions of emotions. *Annual Review of Psychology, 54*, 359-349.
Sander, K., & Scheich, H. (2001). Auditory perception of laughing and crying activates human amygdala regardless of attentional state. *Brain Res Cog Brain Res, 12*((2)), 181-198.
Sander, K., & Scheich, H. (2005). Left Auditory Cortex and Amygdala, But Right Insula Dominance for Human Laughing and Crying. *Journal of Cognitive Neuroscience, 17*, 1519 - 1531.
Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin, 99*, 143-165.
Scherer, K. R. (1995). Expression of emotion in voice and music. *J Voice, 9*, 235-248.
Schirmer, A., Kotz, S. A., & Friederici, A. D. (2005). On the role of attention for the processing of emotions in speech: sex differences revisited. *Brain Res Cogn Brain Res, 24*, 442-452.
Schlosberg, H. (1954). Three dimensions of emotion. *Psychol Rev, 61*, 81-88.
Smith, M. L., Cottrell, G. W., Gosselin, F., & Schyns, P. G. (2005). Transmitting and Decoding Facial Expressions. *Psychological Science, 16*, 184-189.
Vuilleumier, P., Armony, J. L., Driver, J., & Dolan, R. J. (2001). Effects of attention and emotion on face processing in the human brain: An event-related fMRI study. *NEURON, 30*(3), 829-841.
Young, A. W., Rowland, D., Calder, A. J., Etcoff, N. L., Seth, A., & Perrett, D. I. (1997). Facial expression megamix: tests of dimensional and category accounts of emotion recognition. *Cognition, 63*((3)), 271-313.