Behavior Research Methods
2008, 40 (2), 531-539
doi: 10.3758/BRM.40.2.531

The Montreal Affective Voices: A validated set of nonverbal affect bursts for research on auditory affective processing

Pascal Belin
University of Glasgow, Glasgow, Scotland
and Université de Montréal, Montréal, Québec, Canada

and

Sarah Fillion-Bilodeau and Frédéric Gosselin
Université de Montréal, Montréal, Québec, Canada

The Montreal Affective Voices consist of 90 nonverbal affect bursts corresponding to the emotions of anger, disgust, fear, pain, sadness, surprise, happiness, and pleasure (plus a neutral expression), recorded by 10 different actors (5 of them male and 5 female). Ratings of valence, arousal, and intensity for eight emotions were collected for each vocalization from 30 participants. Analyses revealed high recognition accuracies for most of the emotional categories (mean of 68%). They also revealed significant effects of both the actors' and the participants' gender: The highest hit rates (75%) were obtained for female participants rating female vocalizations, and the lowest hit rates (60%) for male participants rating male vocalizations. Interestingly, the *mixed* situations—that is, male participants rating female vocalizations or female participants rating male vocalizations—yielded similar, intermediate ratings. The Montreal Affective Voices are available for download at vnl.psy.gla.ac.uk/ (Resources section).

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, studies of vocal emotion perception in speech use test materials consisting of speech (words, sentences) spoken with various emotional tones by actors (Banse & Scherer, 1996; Buchanan et al., 2000; Kotz et al., 2003; Mitchell, Elliott, Barry, Cruttenden, & Woodruff, 2003; Pell, 2006; Schirmer, Kotz, & Friederici, 2005). A comparatively less studied but equally important way of expressing affect vocally (and facially) is by means of nonverbal interjections, or *affect bursts* (Schröder, 2003). Nonverbal affect bursts, such as laughter or screams of fear, are vocal expressions that usually accompany intense emotional feelings, along with the corresponding facial expressions. They are closely parallel to animal affect vocalizations (Scherer, 1995). Affect bursts can be defined as "short, emotional nonspeech expressions, comprising both clear nonspeech sounds (e.g., laughter) and interjections with a phonemic structure (e.g., 'Wow!'), but excluding 'verbal' interjections that can occur as a different part of speech (like 'Heaven!,' 'No!,' etc.)" (Schröder, 2003, p. 103).

Studies of affective perception in speech prosody are made complex, in particular, by the potential interactions between the affective and the linguistic functions of prosody, on the one hand, and between the affective value carried by prosody and the one that may be carried by its semantic content, on the other (Scherer, Ladd, & Silverman, 1984). A recent model of voice processing (Belin, Fecteau, & Bédard, 2004) suggested that speech processing and affective processing can be processed along partially independent functional pathways, as has been proposed for faces (Bruce & Young, 1986). This is clearly the case in our close evolutionary relatives. It seems important, then, to be able to study affective processing with minimal interactions with linguistic processing. Different strategies have been used to attempt to minimize these interactions: controlling the affective value of the semantic content by using semantically neutral sentences (Imaizumi et al., 1997; Kotz et al., 2003; Laukka, 2005) or pseudosentences composed of logatomes (Banse & Scherer, 1996; Grandjean et al., 2005) or through acoustic manipulations such as low-pass filtering (Friend, 2000; McNally, Otto, & Hornig, 2001). Other researchers have developed batteries of pleasant and unpleasant auditory stimuli consisting of a mix of vocal and nonvocal sounds (the International Affective Digitized Sounds; Bradley & Lang, 1999). Another difficulty associated with the study of speech prosody is that stimuli are necessarily language specific and, thus, cannot be directly used across different countries—for example, to test for cross-cultural effects

P. Belin, p.belin@psy.gla.ac.uk

in auditory affective processing (but see Scherer, Banse, & Wallbott, 2001).

In contrast, research on affective processing in the visual modality is not subject to the same problems, since studies often use nonverbal visual stimuli, such as the International Affective Picture System (Lang, Öhman, & Vaitl, 1988), or sets of affective faces (e.g., Dailey, Cottrell, & Reilly, 2001; Ekman & Friesen, 1978; Ekman, Friesen, & Hager, 2002). The so-called *Ekman faces* consist of a standardized, validated set of photographs of the face of several actors portraying six "basic," or "modal" (Scherer et al., 1984), emotions (anger, disgust, fear, happiness, sadness, and surprise) plus a neutral expression (Ekman & Friesen, 1978), and they do not convey any linguistic information. The Ekman faces were introduced nearly 3 decades ago; they consist of grayscale, static stimuli that are relatively unecological, as compared with the colored, dynamic visual stimuli of the real world. Moreover, they sample a somewhat restricted set of emotions, based on a categorical account of the facial expression of emotions that is still largely debated (Lang, 1995; Russell, Bachorowski, & Fernández-Dols, 2003; Schlosberg, 1954). Nevertheless, the Ekman faces are still widely used in cognitive neuroscience research (e.g., Calder, Burton, Miller, Young, & Akamatsu, 2001; Krolak-Salmon, Fischer, Vighetto, & Mauguière, 2001; Morris et al., 1996; Vuilleumier, Armony, Driver, & Dolan, 2001; Young et al., 1997). Two important advantages may contribute to the popularity of the Ekman faces as research materials. They are nonlinguistic and, thus, can be used in several different countries, potentially allowing cross-cultural comparisons. And several different actors portray the same emotions, allowing use of several different stimuli for each discrete emotion and avoiding the potential confound of actor identity.

The goal of the present contribution is to make available to the research community a validated set of auditory stimuli designed as an auditory counterpart of the Ekman faces. The Montreal Affective Voices (MAV) consist of 90 short, nonverbal affect bursts corresponding to portrayed expressions of anger, disgust, fear, pain, sadness, surprise, happiness, and sensual pleasure (plus a neutral expression) recorded by 10 different actors. Emotional categories of pain and sensual pleasure were added to the six basic emotions of the Ekman faces because they are especially relevant on an evolutionary or welfare level. The variability in segmental structure was minimized by asking the actors to produce these vocalizations on the vowel /ɑ/ (as in *apple*). This set of vocalizations was validated on the basis of ratings of valence, arousal, and perceived intensity along the eight discrete emotions provided by a group of 30 decoders (participants).

We hypothesized that these affect bursts would convey the intended emotional meaning when presented in isolation and only auditorily, as indicated by high recognition accuracy (Schröder, 2003). Moreover, we tested whether the gender effects observed at the level of both affective production and perception in the context of speech prosody would also be observed with the MAV.

METHOD

Recording

Participants. Twenty-two amateur or professional Francophone actors participated in the recording sessions after giving written informed consent. They received a compensation of C$20/h of recording.

Procedure. The actors were instructed to produce short emotional interjections, using the French vowel *ah* (/ɑ/, similar to the English *a* in *apple*), and were played an auditory demonstration of the expressions that 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, and disgust, as well as a neutral expression. Each category of vocalizations was performed several times until our qualitative criterion was reached—that is, 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 Montreal), using a UMT800 condenser microphone (Microtech Gefell) at a distance of approximately 30 cm. Recordings were preamplified using a Millennia Media HV-3B preamplifier and were digitized at a 96-kHz sampling rate and 16-bit resolution, using an Audiophile 2496 PCI soundcard (M-Audio). They were then edited in short, meaningful segments and normalized peak value (90% of maximum amplitude) and were downsampled at 44.1 kHz, using Adobe Audition (Adobe Systems, Inc.). For each actor and vocalization category, only the best occurrence, rated as *successful display* by the experimenter, was kept for the validation stage.

Validation

Participants. Thirty Francophone participants (15 of them male, 15 female) were recruited (average age, 23.3 ± 3 years) through notices posted at the University of Montreal. Each participant gave written informed consent and filled out a sociodemographic information sheet prior to the judgment phase. They were compensated C$10/h for their participation. Data from 1 participant were excluded from the analysis because of technical problems.

Procedure. Each participant was instructed to evaluate each of the 198 vocalizations (22 actors $\times$ 9 categories) on 10 different rating scales: perceived valence of the actor's emotion (from *extremely negative* to *extremely positive*), perceived actor's arousal (from *not at all aroused* to *extremely aroused*), and perceived intensity of the actor's emotion, rated on each of the eight rating scales corresponding to the eight targeted affective states: happiness, sadness, fear, anger, surprise, disgust, pleasure, and pain (e.g., from *not at all angry* to *extremely angry*). For each sound they had to judge, the participants were played the sound, and a judgment board was displayed on a computer screen, consisting of a small speaker icon at the top of the screen and 10 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). The participants could hear at will the sound they were to judge by clicking on the speaker icon. Each of the 10 ratings was performed by clicking, with the computer mouse, on the point of the scale corresponding to the intended judgment. When all 10 judgments had been made, the next sound was played. Ratings along the visual analogue scales were linearly converted to an integer number ranging from 0 to 100.

The stimuli were presented in a pseudorandomized order, in order to control for fatigue and order effects, in four blocks (two blocks of 50 stimuli and two blocks of 49 stimuli) at a self-adjusted comfortable level over DT770 Beyerdynamic headphones. During the session, the participants could take breaks at will between blocks.

Table 1
Acoustic Characteristics for Each Affective Category of the Montreal Affective Voices

Category	$f0$ (Hz) Minimum	Maximum	Median	*SD*	Duration (msec)	Power (dB) Median	*SD*
Angry	150	413	317	80	924	78	14
Disgusted	108	295	200	58	977	75	12
Fearful	266	642	508	97	603	81	12
Happy	181	421	278	58	1,446	60	14
Neutral	149	184	168	4	1,024	81	6
Painful	134	435	351	87	839	77	13
Pleased	120	261	192	43	1,350	70	13
Sad	185	508	323	73	2,229	63	13
Surprised	228	453	373	69	385	76	13

Note—Values are averaged across the 10 actors.

Selection

The 5 male and 5 female actors who produced the most successful displays, based on judgments from the 29 remaining participants, were selected out of the 22 actors. Vocalizations from those 10 actors were then included in further acoustical and statistical analyses.

Acoustical Analyses

Acoustic characteristics of the vocalizations were measured using Straight (Kawahara, Katayose, de Cheveigne, & Patterson, 1999) and Praat (www.praat.org). They included the minimum, maximum, median, and standard deviation of the fundamental frequency ($f0$) measured over the voiced portions (in Hertz); the sound duration (in milliseconds); and the median and standard deviation of power (in decibels). These characteristics, averaged across the 10 actors, are given in Table 1 for each vocalization category. The complete set of acoustic measures can be found in the Appendix. Figure 1 shows waveforms and spectrograms of the 90 vocalizations.

RESULTS AND DISCUSSION

The Montreal Affective Voices (MAV) consist of 90 nonverbal affect bursts corresponding to emotions of anger, disgust, fear, pain, sadness, surprise, happiness, and pleasure (plus a neutral expression), recorded by 10 different actors (5 of them male and 5 female). The set is available for download at the vnl.psy.gla.ac.uk/ (Resource section).

Affective Ratings

Interparticipant reliability in ratings was very high (Cronbach's $\alpha = .978$), so the ratings were averaged across all 29 participants. A first analysis used mixed ANOVAs to investigate the effects of the actor's gender and the participant's gender on ratings of intensity (on the scale corresponding to each portrayed emotion, except the neutral, which had no corresponding scale), valence, and arousal averaged across emotion categories. Actor's gender was found to have a significant effect on all three ratings [intensity, $F(1,27) = 22.0$, $p < .001$; valence, $F(1,27) = 8.4$, $p < .01$; arousal, $F(1,27) = 88.9$, $p < .001$], with greater intensity and arousal and smaller valence ratings for affect bursts produced by female actors. Participant's gender had a significant effect only for intensity ratings [intensity, $F(1,27) = 5.2$, $p < .05$; valence, $F(1,27) = 2.7$, $p > .1$; arousal, $F(1,27) = 1.2$, $p > .2$], with overall higher intensity ratings for male participants. Participant's gender × actor's gender interactions were not significant for any rating (all Fs < 1.1, $p > .3$). Table 2 shows the intensity, valence, and arousal ratings averaged across participants and split by actor's gender, for each emotion category.

The next analyses consisted of comparing intensity ratings across the different rating scales and emotion categories. Table 3 shows that the maximum average rating for each emotion was always obtained on the corresponding rating scale (diagonal cells in bold). We first asked, for each portrayed emotion (except the neutral), whether there would be a rating scale with ratings higher than those on the other scales; this was the case for all the categories of affect bursts except the happy and pleased vocalizations (Fisher's protected LSD; Table 3, columns). We then asked, for each rating scale, whether there would be one

Table 2
Intensity, Valence, and Arousal Ratings for Vocalizations From Male and Female Actors

Vocal Expression	Intensity of Vocal Expression: Male Actors *M*	*SEM*	Female Actors *M*	*SEM*	Valence of Vocal Expression: Male Actors *M*	*SEM*	Female Actors *M*	*SEM*	Arousal of Vocal Expression: Male Actors *M*	*SEM*	Female Actors *M*	*SEM*
Neutral	–	–	–	–	47	2.1	48	1.6	31	3.3	35	2.5
Angry	73	3.4	76	3.6	18	2.0	15	2.1	70	2.1	75	2.5
Disgusted	61	4.0	79	3.0	28	2.5	21	3.0	34	2.6	39	3.0
Fearful	62	3.3	73	2.7	27	1.5	22	1.4	67	2.6	79	2.0
Painful	48	3.9	68	3.3	21	1.9	26	2.8	61	2.8	72	2.3
Sad	68	3.3	86	2.6	28	2.2	11	1.6	44	3.3	46	3.9
Surprised	73	3.5	82	2.1	40	1.6	38	1.2	66	2.3	76	1.9
Happy	83	2.9	80	3.2	86	1.9	85	2.6	59	2.8	57	3.7
Pleased	55	3.4	69	3.0	65	2.7	76	2.1	30	3.1	42	3.3

Note—Ratings (visual analogue scales converted to 0–100) are averaged across judges and split by actor's gender.

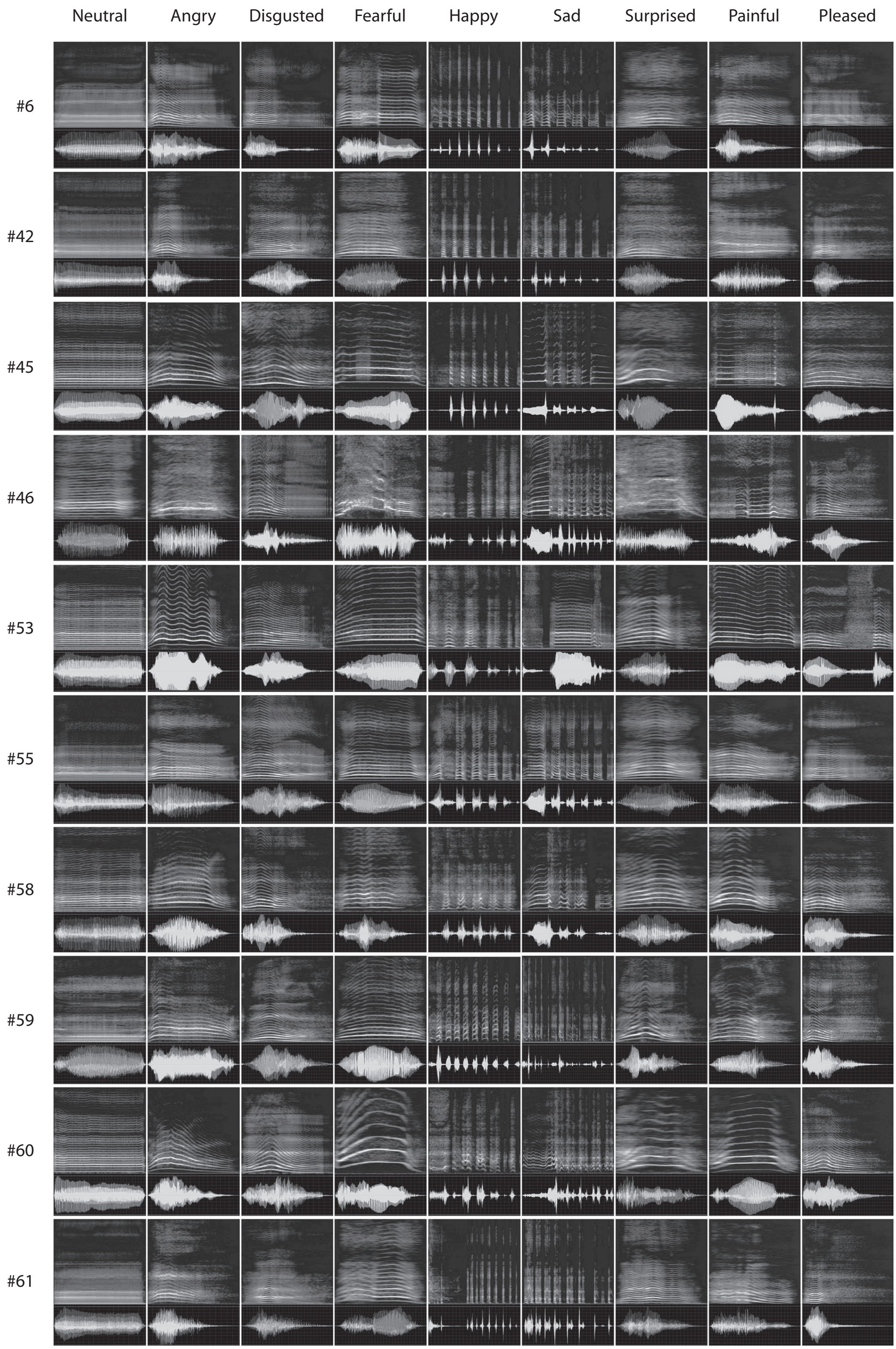

Figure 1. The Montreal Affective Voices. Waveforms and spectrograms (0–8000 Hz) of the 90 vocalizations. *x*-axis: time, scaled to the duration of each vocalization. Each row corresponds to one actor (actor's number is given on the left-hand side), and each column to a category of affect burst.

Table 3
Intensity Ratings, Sensitivity (Hit Rates), and Specificity (Correct Rejection Rates), by Expression Category

Intensity Rating Scale	Portrayed Expression																		Correct Rejection Rate	
	Neutral		Angry		Disgusted		Fearful		Painful		Sad		Surprised		Happy		Pleased			
	M	*SEM*	*M*	*SEM*	*M*	*SEM*	*M*	*SEM*	*M*	*SEM*	*M*	*SEM*	*M*	*SEM*	*M*	*SEM*	*M*	*SEM*	*M*	*SEM*
Anger	9	0.5	**75^{ac}**	**2.4**	14	1.1	19	2.1	33	3.8	9	0.7	17	0.8	3	0.3	7	0.8	77	4.4
Disgust	10	0.5	23	0.7	**70^{ac}**	**2.7**	21	1.6	26	2.7	9	0.6	24	1.4	4	0.3	8	0.7	73	4.5
Fear	9	0.5	16	2.0	11	0.6	**68^{bc}**	**2.5**	21	2.0	10	0.7	45	2.6	3	0.2	6	1.0	69	3.0
Pain	9	0.9	24	1.6	11	1.1	31	3.1	**58^{ac}**	**3.6**	26	1.8	21	1.0	3	0.2	7	0.4	62	4.0
Sadness	11	0.8	13	1.2	9	0.8	13	1.0	15	1.5	**77^{ac}**	**3.6**	11	0.4	3	0.2	5	0.3	89	2.5
Surprise	9	0.6	26	1.8	26	1.8	57	3.0	35	3.0	11	0.5	**77^{ac}**	**2.0**	18	1.1	25	2.2	64	2.7
Happiness	14	0.5	6	0.4	9	0.8	7	0.4	10	1.1	11	2.7	15	1.3	**81^{c}**	**1.2**	54	3.3	76	3.0
Pleasure	13	0.3	6	0.4	9	0.9	6	0.4	11	1.9	10	2.4	12	1.0	76	1.1	**62**	**3.8**	39	4.0
Hit rate			78	5.0	81	3.7	56	3.0	51	3.0	86	2.0	75	2.9	60	4.5	59	3.8		

Note—Cells indicate intensity ratings (0–100) averaged across all actors and judges for each portrayed emotion and intensity ratings scale. Boldface indicates maximum average rating. Note the high hit rates for most affective categories. $^{a}p < .001$. $^{b}p < .05$, strongest rating on the scale corresponding to the portrayed emotion (columns). $^{c}p < .001$, strongest rating for the portrayed emotion corresponding to the rating scale (rows) (Fisher's protected least significance test).

category of affective vocalizations yielding higher ratings than did the other vocalizations: This was the case for all the rating scales except the surprise and pleasure scales (Fisher's protected LSD; Table 3 rows).

Sensitivity and Specificity

We then examined the intensity ratings for their sensitivity (hit rate, by emotion) and specificity (correct rejection rate, by rating scale). In order to evaluate recognition accuracy, the intensity ratings along the eight scales were considered as an eight-alternative forced choice classification: For each vocalization, a maximum intensity rating in the scale corresponding to the portrayed emotion was considered as a *hit*; otherwise, as a *miss*. The bottom row of Table 3 shows averaged hit rates for each emotion category, ranging from 51% for painful vocalizations to 81% for disgusted vocalizations. Similarly, for each rating scale, a maximum score for the corresponding vocalization in a given actor/participant was considered as a *correct rejection*; otherwise, as a *false alarm*. The rightmost column of Table 3 shows correct rejection rates for each rating scale, ranging from 39% for pleasure to 89% for sadness.

Finally, we examined the effects of participant's and actor's gender on hit rates (Figure 2). A mixed ANOVA showed significant effects of both actor's gender [$F(1,27) = 29.3$, $p < .001$] and participant's gender [$F(1,27) = 5.0$, $p < .05$], with no interaction [$F(1,27) < 1$].

Recognition Accuracy

The recognition accuracy (sensitivity) for the 8 emotions expressed through affect bursts was quite high (mean of 68.2% for the chance level at 12.5%), with very little ambiguity for several categories of negative affect bursts: angry (78%), disgusted (81%), or sad (86%) vocalizations. This is in line with the very high accuracies (81% mean for 10 categories) observed in another study of affect bursts recognition (Schröder, 2003). As was noted in this study, these recognition accuracy scores are relatively higher than those observed in the context of speech prosody (e.g., 55% for 10 emotions in Banse & Scherer, 1996), suggesting that affect bursts are a highly effective means of expressing vocal emotion.

The strongest pattern of confusion was observed between the happy and the pleased vocalizations: Both vocalization categories yielded high, statistically indistinguishable ratings on the happiness and pleasure scales. Yet the two types of vocalizations were characterized by markedly different acoustical structures, with the syllabic pattern characteristic of laughter completely absent from pleasure vocalizations. Thus, this difficulty in separating happy from pleased vocalizations probably arose from the use of the rating scales, indicating that the participants probably did not have clearly separated concepts associated with the labels *pleased* and *happy*. This, in turn, may be taken as evidence that these two categories of vocalizations do not really correspond to two different basic emo-

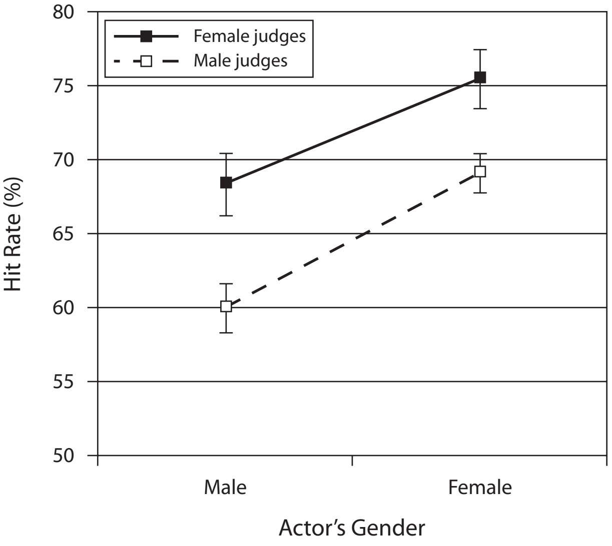

Figure 2. Gender effects on emotion recognition using the Montreal Affective Voices; hit rates (percentage of test items with maximal rating on the scale corresponding to the portrayed emotion) split by actor's and participant's gender. Note equivalent accuracies for male vocalizations judged by females and for female vocalizations judged by males, owing to similar magnitudes of the production and perception gender effects.

tions. The very similar ratings, overall, on the pleasure and happiness rating scales and the low correct rejection rate on the pleasure scale (39%) suggest that the latter could be simply removed from the rating procedure in future uses.

A second pattern of confusion, clearly observable in Table 3, is between fear and surprise: Fearful vocalizations yielded high ratings on both the fear and surprise scales, resulting in low hit rates (56%). Surprised vocalizations yielded a higher hit rate (75%), yet they often yielded high scores on the fear scale. Contrary to the case with the pleasure and happiness expressions, the confusion between fear and surprise is not likely to have arisen from unclear separation of *fearful* and *surprised* labels, which correspond to clearly separated emotional contexts. Rather, the two categories of vocalizations were acoustically quite similar: Table 1 shows that both the fearful and the surprised vocalizations were characterized by the shortest duration and the highest median $f0$. Interestingly, however, the confusion between fear and surprise is also frequently observed when these emotions are expressed by human faces in some cultures (Ekman, Friesen, & Ellsworth, 1972) or when an ideal observer classifies affective faces (Smith, Cottrell, Gosselin, & Schyns, 2005).

Finally, the lowest recognition accuracy was obtained for painful vocalizations (51%), which were frequently confounded with angry or surprised vocalizations. This pattern of confusion is relatively surprising, given the evolutionary importance of recognizing and avoiding painful contexts, and has not been observed in a recent study of emotion recognition based on dynamic facial expressions (Simon, Craig, Gosselin, Belin, & Rainville, 2008). The acoustical similarity of pain vocalizations, which, like the angry or surprised vocalizations, are relatively brief and have a high median $f0$ (Table 1), probably accounts for this confusion.

Gender Effects

Affective ratings were strongly influenced by the gender of the actor producing the vocalizations. Averaged across emotions, ratings of perceived arousal and intensity (on the scale corresponding to the portrayed emotion) were consistently higher, and valence ratings lower (reflecting the greatest contribution of the numerically superior negative emotion categories), for vocalizations produced by the female actors (Table 2). This finding is in line with the notion that in Western cultures, women are believed to be more emotionally expressive, in general, than are men. Specifically, they are expected to smile more, as well as to show more sadness, fear, and guilt. In contrast, men are believed to show more overt emotional displays only in terms of physically aggressive anger (Brody & Hall, 1993; Fischer, 1993). Whether this effect is specific to the group of actors selected (with selection biases reflecting the choice of actors with the most successful displays) or reflects a more general gender difference in vocal emotional expression, perception, or both remains to be explored.

Participant's gender was also found to have a significant effect for intensity ratings: Male participants gave slightly higher intensity ratings than did female participants overall. But the clearest picture was given by the analysis of hit rates (proportion of correct recognition): The effects of both actor's gender and participant's gender were significant and additive (no interaction), both in the direction of higher hit rates for female participants or actors and with similar magnitudes. Figure 2 shows that the highest hit rates were obtained for female participants rating female vocalizations, and the lowest hit rates for male participants rating male vocalizations. That is to say, women are generally more accurate in judging affective bursts by other females (although 25% fail to), whereas men are considerably worse at judging other men (they fail 40% of the time). Interestingly, the *mixed* situations—that is, male participants rating female vocalizations or female participants rating male vocalizations—yielded similar, intermediate ratings (Figure 2). This result is in line with evidence of better accuracy in emotion recognition (Hall, 1978) and of faster recognition and enhanced cortical responses (Schirmer & Kotz, 2006) by female than by male participants, in the context of speech prosody.

Acoustical Measures

Acoustical measures (see Table 1 and the Appendix) are characterized by a substantial degree of variation, but values averaged across actors show important but consistent differences between emotional categories. These differences are generally in good agreement with the results in the existing literature on speech prosody perception (Banse & Scherer, 1996; Juslin & Laukka, 2003; Murray & Arnott, 1993; Scherer, 1986). For example, Table 1 shows that the neutral vocalization was associated with the lowest median pitch and very small pitch variation, whereas fearful vocalizations were associated with the highest median pitch and widest pitch range, exactly as in Murray and Arnott (1993).

Using the MAV

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

We suggest that the MAV stimuli can be used for several different purposes. They can be used to test for possible deficits in auditory affect recognition, using rating scales similar to those used in the present validation phase (perhaps without the pleasure scale, which proved to overlap too much with the happiness scale). The fact that there is no verbal content in these stimuli is an advantage when it comes to testing patients with lesions that impair speech comprehension.

The MAV stimuli can also be used in cognitive psychology experiments. It is important to note that the duration of the vocalizations varies widely across the different categories: The average duration across the 10 actors varies from 385 msec for the cries of surprise, to 1,446 msec for the laughs (happiness), up to 2,229 msec for the cries (sadness). Although this important variation may raise problems in some experimental settings (e.g., reaction time studies of evoked response potential studies), it reflects the natural properties of the different classes of vocalizations and is unavoidable if one wants to preserve the ecological validity of the stimuli. Editing the cries, for example, in an attempt to make them shorter and more compatible with the duration of other sounds could potentially affect their naturalness and would probably lead to important confusions with the laughs. Thus, although the MAV stimuli do present problems for use in TR experiments (due to the duration of the vocalizations), they may be useful if used in fMRI experiments in which small differences in stimulus duration raise fewer problems than in EEG/MEG experiments. For example, a recent event-related fMRI study using a preliminary version of the MAV revealed clear bilateral activation of the amygdala, by comparison with affectively neutral vocalizations, for negatively valenced, as well as positively valenced, vocalizations (Fecteau, Belin, Joanette, & Armony, 2007).

AUTHOR NOTE

This research was supported by grants from the Canadian Institutes of Health Research, the Canadian Foundation for Innovation, and France-Telecom to P.B. We thank anonymous reviewers for their constructive comments. Correspondence concerning this article should be addressed to P. Belin, Department of Psychology, University of Glasgow, 58 Hillheed Street, Glasgow G12 8QB, Scotland (e-mail: p.belin@psy.gla.ac.uk).

REFERENCES

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality & Social Psychology*, **70**, 614-636.

Belin, P., Fecteau, S., & Bédard, C. (2004). Thinking the voice: Neural correlates of voice perception. *Trends in Cognitive Sciences*, **8**, 129-135.

Bradley, M. M., & Lang, P. J. (1999). *International affective digitized sounds (IADS): Stimuli, instruction manual and affective ratings* (Tech. Rep. B-2). Gainesville: University of Florida, Center for Research in Psychophysiology.

Brody, L. R., & Hall, J. A. (1993). Gender and emotion. In M. Lewis & J. M. Haviland (Eds.), *Handbook of emotions* (pp. 447-460). New York: Guilford.

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology*, **77**, 305-327.

Buchanan, T. W., Lutz, K., Mirzazade, S., Specht, K., Shah, N. J., Zilles, K., & Jäncke, L. (2000). Recognition of emotional prosody and verbal components of spoken language: An fMRI study. *Cognitive Brain Research*, **9**, 227-238.

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

Dailey, M., Cottrell, G. W., & Reilly, J. (2001). *California facial expressions, CAFE: Unpublished digital images*. San Diego: University of California, Computer Science and Engineering Department.

Ekman, P., & Friesen, W. V. (1978). *Facial action coding system: Investigator's guide*. Palo Alto, CA: Consulting Psychologists Press.

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

Ekman, P., Friesen, W. V., & Hager, J. C. (2002). *Facial Action Coding System investigator's guide*. Salt Lake City, UT: A Human Face.

Fecteau, S., Belin, P., Joanette, Y., & Armony, J. L. (2007). Amygdala responses to nonlinguistic emotional vocalizations. *NeuroImage*, **36**, 480-487.

Fischer, A. (1993). Sex differences in emotionality: Fact or stereotype? *Feminism & Psychology*, **3**, 303-318.

Friend, M. (2000). Developmental changes in sensitivity to vocal paralanguage. *Developmental Science*, **3**, 148-162.

Grandjean, D., Sander, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., & Vuilleumier, P. (2005). The voices of wrath: Brain responses to angry prosody in meaningless speech. *Nature Neuroscience*, **8**, 145-146.

Hall, J. A. (1978). Gender effects in decoding nonverbal cues. *Psychological Bulletin*, **85**, 845-857.

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**, 2809-2812.

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

Kawahara, H., Katayose, H., de Cheveigne, A., & Patterson, R. D. (1999, September). *Fixed point analysis of frequency to instantaneous frequency mapping for accurate estimation of f0 and periodicity*. Paper presented at the 6th European Conference on Speech Communication and Technology (Eurospeech'99), Budapest.

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 & Language*, **86**, 366-376.

Krolak-Salmon, P., Fischer, C., Vighetto, A., & Mauguière, F. (2001). Processing of facial emotional expression: Spatio-temporal data as assessed by scalp event-related potentials. *European Journal of Neuroscience*, **13**, 987-994.

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

Lang, P. [J.], Öhman, A., & Vaitl, D. (1988). *The international affective picture system*. Gainesville: University of Florida, Center for Research in Psychophysiology.

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. *Behaviour Research & Therapy*, **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., & Dolan, R. J. (1996). A differential neural response in the human amygdala to fearful and happy facial expressions. *Nature*, **383**, 812-815.

Murray, I. R., & Arnott, J. L. (1993). Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *Journal of the Acoustical Society of America*, **93**, 1097-1108.

Pell, M. D. (2006). Cerebral mechanisms for understanding emotional prosody in speech. *Brain & Language*, **96**, 221-234.

Russell, J. A., Bachorowski, J.-A., & Fernández-Dols, J.-M. (2003). Facial and vocal expressions of emotions. *Annual Review of Psychology*, **54**, 329-349.

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. *Journal of Voice*, **9**, 235-248.

SCHERER, K. R., BANSE, R., & WALLBOTT, H. G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology*, **32**, 76-92.

SCHERER, K. R., LADD, D. R., & SILVERMAN, K. E. A. (1984). Vocal cues to speaker affect: Testing two models. *Journal of the Acoustical Society of America*, **76**, 1346-1356.

SCHIRMER, A., & KOTZ, S. A. (2006). Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences*, **10**, 24-30.

SCHIRMER, A., KOTZ, S. A., & FRIEDERICI, A. D. (2005). On the role of attention for the processing of emotions in speech: Sex differences revisited. *Cognitive Brain Research*, **24**, 442-452.

SCHLOSBERG, H. (1954). Three dimensions of emotion. *Psychological Review*, **61**, 81-88.

SCHRÖDER, M. (2003). Experimental study of affect bursts. *Speech Communication*, **40**, 99-116.

SIMON, D., CRAIG, K. D., GOSSELIN, F., BELIN, P., & RAINVILLE, P. (2008). Recognition and discrimination of prototypical dynamic expressions of pain and emotions. *Pain*, **135**, 55-64.

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**, 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**, 271-313.

APPENDIX
Acoustic Characteristics of All 90 Stimuli of the MAV

	$f0$ (Hz)				Duration	Power (dB)	
Stimulus	Minimum	Maximum	Median	*SD*	(msec)	Median	*SD*
6_anger.wav	73	290	151	67	1,142	77.7	13.9
6_disgust.wav	115	196	158	25	1,051	63.1	13.4
6_fear.wav	129	537	317	93	761	80.2	10.5
6_happiness.wav	144	343	225	42	1,742	47.5	15.9
6_neutral.wav	91	116	113	4	896	80.9	5.6
6_pain.wav	108	313	238	64	745	71.4	12.9
6_pleasure.wav	71	151	110	24	1,001	75.0	11.4
6_sadness.wav	201	336	262	25	1,643	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	1,045	77.4	13.6
42_fear.wav	149	313	289	50	405	81.3	14.8
42_happiness.wav	139	223	157	20	1,445	52.6	13.5
42_neutral.wav	102	116	112	2	1,312	78.6	4.5
42_pain.wav	224	283	273	14	584	77.9	11.6
42_pleasure.wav	90	157	125	23	930	64.6	12.8
42_sadness.wav	132	233	181	30	1,667	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	1,563	50.6	13.7
45_neutral.wav	222	253	228	3	992	83.7	5.5
45_pain.wav	49	689	510	204	1,528	75.1	14.2
45_pleasure.wav	247	414	346	51	879	76.6	9.7
45_sadness.wav	251	815	519	171	1,780	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	1,566	70.6	11.3
46_fear.wav	375	1658	926	344	815	82.3	12.8
46_happiness.wav	189	584	231	101	1,009	61.6	13.8
46_neutral.wav	209	289	260	11	240	83.6	14.8
46_pain.wav	86	525	377	87	1,347	74.1	13.6
46_pleasure.wav	91	286	213	52	1,621	68.3	19.6
46_sadness.wav	331	661	433	88	1,956	70.4	11.2
46_surprise.wav	446	469	463	6	404	79.2	10.9
53_anger.wav	166	517	417	113	1,518	83.1	17.5
53_disgust.wav	143	253	213	31	1,714	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_pain.wav	252	451	397	36	1,324	81.6	6.5
53_pleasure.wav	177	318	245	42	1,655	75.3	7.6
53_sadness.wav	160	537	302	37	2,877	73.8	13.6

APPENDIX (Continued)
Acoustic Characteristics of All 90 Stimuli of the MAV

Stimulus	*f*0 (Hz)				Duration	Power (dB)	
	Minimum	Maximum	Median	*SD*	(msec)	Median	*SD*
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	1,100	67.6	12.5
55_neutral.wav	106	130	109	2	1,236	77.3	3.7
55_pain.wav	114	263	234	47	565	77.3	12.0
55_pleasure.wav	72	172	125	32	871	71.7	11.5
55_sadness.wav	150	309	249	39	1,830	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	1,046	66.8	9.6
58_neutral.wav	184	222	211	9	511	82.5	4.5
58_pain.wav	183	495	412	105	581	78.1	17.5
58_pleasure.wav	159	322	242	54	1,100	68.9	13.8
58_sadness.wav	186	542	379	90	1,416	65.4	13.5
58_surprise.wav	235	441	382	56	329	78.9	11.8
59_anger.wav	131	377	336	64	1,184	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	1,831	64.5	16.7
59_neutral.wav	139	197	143	5	645	84.3	5.5
59_pain.wav	89	393	292	103	707	77.0	13.9
59_pleasure.wav	75	272	167	59	2,067	70.4	14.2
59_sadness.wav	198	773	404	132	4,310	53.7	13.8
59_surprise.wav	129	475	304	109	574	75.8	15.9
60_anger.wav	159	516	301	113	1,082	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	1,159	68.0	15.4
60_neutral.wav	193	222	214	3	1,597	81.0	5.8
60_pain.wav	153	621	583	143	432	79.4	12.2
60_pleasure.wav	145	341	214	55	1,769	77.0	10.6
60_sadness.wav	154	662	345	95	2,376	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	2,605	50.4	13.2
61_neutral.wav	85	101	95	1	1,861	78.9	5.6
61_pain.wav	81	322	191	67	583	75.9	10.7
61_pleasure.wav	78	182	132	33	1,611	57.2	14.0
61_sadness.wav	84	211	159	25	2,438	58.3	17.2
61_surprise.wav	82	225	185	43	514	73.0	10.8

(Manuscript received September 7, 2007;
revision accepted for publication December 21, 2007.)