Brain Topogr (2008) 20:239–248
DOI 10.1007/s10548-008-0051-8

ORIGINAL PAPER

Emotional Pre-eminence of Human Vocalizations

Mélanie Aeschlimann · Jean-François Knebel · Micah M. Murray · Stephanie Clarke

Accepted: 11 February 2008 / Published online: 18 March 2008

Abstract Human vocalizations (HV), as well as environmental sounds, convey a wide range of information, including emotional expressions. The latter have been relatively rarely investigated, and, in particular, it is unclear if duration-controlled non-linguistic HV sequences can reliably convey both positive and negative emotional information. The aims of the present psychophysical study were: (i) to generate a battery of duration-controlled and acoustically controlled extreme valence stimuli, and (ii) to compare the emotional impact of HV with that of other environmental sounds. A set of 144 HV and other environmental sounds was selected to cover emotionally positive, negative, and neutral values. Sequences of 2 s duration were rated on Likert scales by 16 listeners along three emotional dimensions (arousal, intensity, and valence) and two non-emotional dimensions (confidence in identifying the sound source and perceived loudness). The 2 s stimuli were reliably perceived as emotionally positive, negative or neutral. We observed a linear relationship between intensity and arousal ratings and a "boomerang-shaped" intensity-valence distribution, as previously reported for longer, duration-variable stimuli. In addition, the emotional intensity ratings for HV were higher than for other environmental sounds, suggesting that HV constitute a characteristic class of emotional auditory stimuli. In addition, emotionally positive HV were more readily identified than other sounds, and emotionally negative stimuli, irrespective of their source, were perceived as louder than their positive and neutral counterparts. In conclusion, HV are a distinct emotional category of environmental sounds and they retain this emotional pre-eminence even when presented for brief periods.

Keywords Auditory · Emotion · Sound battery · Object · Vocalization

Electronic Supplementary Material The online version of this article (doi:10.1007/s10548-008-0051-8) contains supplementary material, which is available to authorized users.

M. Aeschlimann (✉) · J.-F. Knebel · M. M. Murray · S. Clarke
Service de Neuropsychologie et de Neuroréhabilitation, Centre Hospitalier Universitaire Vaudois (CHUV) and Université de Lausanne (UNIL), Av. Pierre Decker 5,
1011 Lausanne, Switzerland
e-mail: Melanie.Aeschlimann@chuv.ch

M. M. Murray
Radiology Service, Centre Hospitalier Universitaire Vaudois (CHUV) and Université de Lausanne (UNIL), Lausanne, Switzerland

M. M. Murray
EEG Brain Mapping Core,
Center for Biomedical Imaging, Centre Hospitalier Universitaire Vaudois (CHUV) and Université de Lausanne (UNIL), Lausanne, Switzerland

Introduction

The quintessential role of auditory stimuli in emotion and affective processing is immediately apparent upon viewing a frightening movie either with or without its soundtrack. While the visual modality has been studied in extensive detail, comparatively less is known concerning the auditory modality. To date, auditory research has largely focused on the emotional attributes of speech prosody and music [1–6], with only a few studies using either environmental sounds or non-linguistic vocalizations [7–12].

Non-linguistic stimuli play a key role in the communication of affective states both in humans and animals (e.g. [13]). Facial expressions are a prominent example of this,

and their importance relative to other objects is demonstrated by the fact that the processing of these stimuli relies on specific neural circuitry (e.g. [14–17]). The present study had two objectives. The first objective was to generate an auditory stimulus battery that includes positively, negatively, and neutrally rated sounds of relatively short and equal duration[1] that are appropriate for use in psychophysical and brain imaging investigations (c.f. [11] for a recent discussion of this issue). The second objective was to assess whether human vocalizations of the above emotional valences, similarly to facial expressions, constitute a distinct category of emotionally potent auditory stimuli, as has been proposed by Belin et al. [7]. To do this, it was necessary to contrast ratings from human vocalizations with their non-vocalization counterparts.

According to affective theory, the elicitation of emotion results from the interaction of two motivational systems, an appetitive and a defensive; the engagement of which can be measured by hedonic valence (from positive for pleasant states to negative for unpleasant ones) and arousal (from calm to excited; e.g. [19]). In pioneering studies, Bradley and Lang [20] and Fecteau et al. [21] had their participants rate stimuli along three dimensions: valence, intensity (or dominance) and arousal. With respect to our objectives, the battery developed by Fecteau et al. [21] is limited to human vocalizations, the duration of their stimuli was not detailed, and their focus at the time was on age-related effects. Bradley and Lang documented a bilinear relationship (or "boomerang-shaped" distribution) between arousal and valence as well as a linear relationship between intensity and arousal. However, these authors did not differentiate between object categories, in particular human vocalizations (both linguistic and non-linguistic) among other environmental sound categories. Furthermore, the 6 s duration of their stimuli would not allow for readily identifying portion(s) of the sound critical for conveying affective information. In fact, subsequent authors have been unable to reliably obtain positive ratings for sounds of environmental objects when stimuli were shortened to 350–500 ms duration [11], see also Thierry and Roberts [12] for a study using sounds of >1 s duration).

To foreshadow our results, we successfully constructed a battery of short-duration stimuli containing stimuli reliably rated as emotionally positive, negative, and neutral. This battery includes both non-linguistic human vocalizations (HV) and non-vocalizations (NV). For the HV stimuli, we followed a tactic similar to that of [21] and [22], in that we limited these positive and negative stimuli to laughs, cries, screams, and erotic exclamations (see electronic supplementary material for full list). The neutral stimuli were short utterances (e.g. /a/) based on digital editing of laboratory recordings. As such, these stimuli are distinct from prior studies that presented words or word-like utterances spoken with different intonations (e.g. [2, 3, 23]). Our psychophysical data support the proposition that human vocalizations are a distinct category of emotional auditory stimuli; they were reliably rated as more intense across all three emotional valences (positive, neutral, and negative). Following the application of a principal component analysis (PCA) algorithm to identify the extreme-most exemplars of both HV and NV for each emotional valence, we again observed higher intensity ratings for HV and additionally observed that the extreme-most positive HV conveyed the strongest confidence in source identification and that negative stimuli, irrespective of sound source category, were perceived as being louder than either neutral or positive stimuli, despite all stimuli being RMS-normalized and despite no evidence of reliable differences in a time-frequency analysis of the stimuli.

Materials and Methods

Participants

Sixteen healthy subjects (8 women, mean ± sd age: 28.8 ± 3.6 yrs) participated in the study. They were exempt of neurological or psychiatric disorders and reported normal audition. All gave informed consent to participate. All procedures were approved by the Ethics Committee of the Faculty of Biology and Medicine at the University of Lausanne. Additional feedback classification of participants' raw data (i.e. using the PCA-defined clusters to sort in a post-hoc manner the original data) allowed us to analyze ratings as a function of gender (see also [21]). As there were no reliable effects of gender, we do not discuss this aspect further here.

Stimuli

A set of 144 sounds was selected for their high affective potential from various libraries (IADS as supplied in Bradeley and Lang [24], BBC sounds effects) or were digitally recorded in our laboratory with a micro-phone (audio-technica® ATR20). A listing of the provenance of the stimuli can be found in the electronic supplementary material. All stimuli (16 bit stereo) were edited to be 2 s in

[1] A pilot investigation suggested that 2 s duration is sufficient for eliciting each of the three emotional valences. We would further note that studies of the discrimination of sounds of environmental objects have dissociated electrophysiological indices of categorical discrimination from psychophysical indices [18]. Thus, determining the 'recognition point' within a stimulus or category of stimuli based on behavioral measures may not be a direct reflection of underlying brain processes. More germane to the present study is that stimuli of equal 2 s duration are more readily controlled in terms of their acoustic features (see electronic supplementary material for details).

duration and were digitized at 44.1 kHz, using Adobe Audition 1.0 (Adobe Systems Incoporated). Amplitude enveloping was applied to the initial and final 10 ms of each sound to minimize clicks. All sounds were further normalized according to the root mean square of their amplitude (see electronic supplementary material for details). Eighty-four stimuli contained human non-linguistic vocalizations (e.g. laughs, cries, yells, neutral short vocalizations[2], etc.), and 60 stimuli were chosen that were neither human nor animal vocalizations (e.g. alarm, hands clapping, typewriter, etc.). One of the 144 sounds contained a mixture of both categories (i.e. a person screaming with a gunshot) but was unequivocally categorized as "human" by all participants and was therefore included in the human vocalization category.

Task and Procedure

Subjects were asked to rate each stimulus along three emotional dimensions: (a) affective valence (7-point Likert scale with 1 being very pleasant and 7 extremely unpleasant), (b) emotional intensity or potency (5-point Likert scale with 1 being exempt of emotional content and 5 highly emotional), and (c) arousal (5-point Likert scale with 1 being low and 5 high). Furthermore, listeners also provided (d) confidence ratings regarding source identification that generated each sound (5-point Likert scale with 1 being fully confident and 5 completely uncertain) and also (e) ratings of the perceived loudness of each sound (5-point Likert scale with 1 being too dull and 5 too loud with 3 being pleasant to hear).

Sounds were presented pseudo-randomly over three blocks of trials, each of which was comprised of 48 trials that were broadly equivalent in the number of HV and NV stimuli and also in sounds that resulted in positive, negative, or neutral emotional valence ratings pre-established by two independent judges who are not included in the current study. Each block of trials was completed in approximately 20 min. Some listeners completed all three blocks in one session (taking breaks between blocks), and the remaining listeners completed each of the three blocks on different days (maximally spanning over 4 days). Each sound presentation, via headphones (Technics RP-F550), was followed by the three emotional questions (valence, emotional intensity, and arousal), each presented visually and one at a time on a computer monitor. After indicating the ratings for the emotional questions, the sound was replayed, followed by the questions regarding source identification and judgment of perceived loudness, presented visually on the computer monitor. A white noise of 1 s was presented between trials. The order of the three emotional questions was distinct on each block to avoid serial-order effects. Each emotional question (a, b, c) was accompanied by visual cues (Fig. 1), and Likert scales for source identification (d) and perceived loudness (e) were cued by a design (a question mark and an ear, respectively). Responses on the Likert scale were performed with a computer keyboard without any time limit.

To assess whether there were serial-order effects on the emotional questions as well as an influence of the schedule of the experiment (i.e. completing all blocks in one session versus sessions spread out over several days), a repeated measures analysis of variance (ANOVA) was performed using schedule (single-day vs. multiple days) as the between-subject factor, and the order of emotional questions in a block and the specific emotional question as the within-subject factors. As neither a serial-order effect nor a schedule influence was found, these aspects are not further discussed.

Sound Classification

Raw data were first converted into *Z*-scores, based on the presumption of a normal cumulative distribution; to facilitate interpretation, *Z*-scores for valence and source identification were multiplied by minus one to give a negative value to negative valence or a lower confidence in source identification, respectively; and a positive value to positive valence or a higher confidence in source

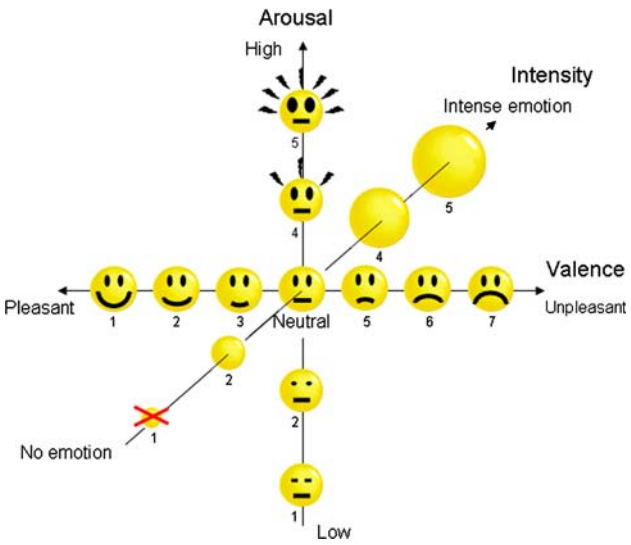


Fig. 1 Representation of the three-dimensional affective space with the symbols used for rating the three emotional questions on Likert scales

[2] These short, vowel-like stimuli were artificially edited by cutting, copying, and pasting segments from the original laboratory recordings such that, for example, a sound that was originally 1 s duration would have internal repetitions to render it 2 s duration. So doing not only rendered the stimuli completely meaningless, but also minimized their prosodic content.

identification, respectively. Three sounds were considered as outliers, their *Z*-scores being further than 2 standard deviations from the mean (c.f. Clark-Carter [25]), and were excluded from further analysis (two NV sounds were rated as significantly too loud, and one NV was unidentifiable). For the remaining 141 sounds, *Z*-score data were represented as points in the three-dimensional affective space formed by arousal, emotional intensity and valence as presented in Fig. 1. Using Matlab, the data from HV and NV, separately, were then split into two clouds of points according to their valence ratings (positive and negative human vocalizations, HV+ and HV−, as well as positive and negative non-vocalizations, NV+ and NV−). Positive *Z*-scores for valence indicated pleasantness, and negative scores unpleasantness. A principal component analysis (PCA) algorithm (e.g. Hastie et al. [26]) was applied to each cloud of data, separately, to establish the maximal covariance direction of that cloud, representing the direction of maximal extension of that subset of data.

The mean affective rating value for valence, intensity and arousal of each sound was calculated across the 16 subjects and attributed to one of four "clouds" of points within the affective space, defined by source category (HV, NV) and polarity of valence rating (positive, negative). For each cloud (i.e. HV+, HV−, NV+, and NV−), the orthogonal projection of each sound point on the principal direction identified by the PCA provided a sequential classification of the stimuli within the cloud (e.g. the most to the least pleasant). The 11 most extreme stimuli within each cloud were then selected for further analysis. To identify neutral stimuli, a "neutral principal direction" was derived from the linear combination of the positive and negative valence directions. As before, orthogonal projections of the points on this direction provided a sequential classification from which we identified the 11 most neutral sounds for HV and for NV. In summary, the PCA analysis helped to select the 11 most reliably rated sounds for each emotional valence (positive, neutral, and negative) and each sound category (HV and NV). In addition to these three sub-classes of extreme values (i.e. extremely positive, neutral, and extremely negative), the remaining stimuli were ascribed to either of two additional sub-classes according to whether their *Z*-scores for valence were positive or negative (i.e. moderately positive and moderately negative, respectively).

Results

Reliable Emotional Categorization with Short-duration Sounds

We first assessed whether sounds of 2 s duration could reliably elicit positive, neutral, and negative emotional

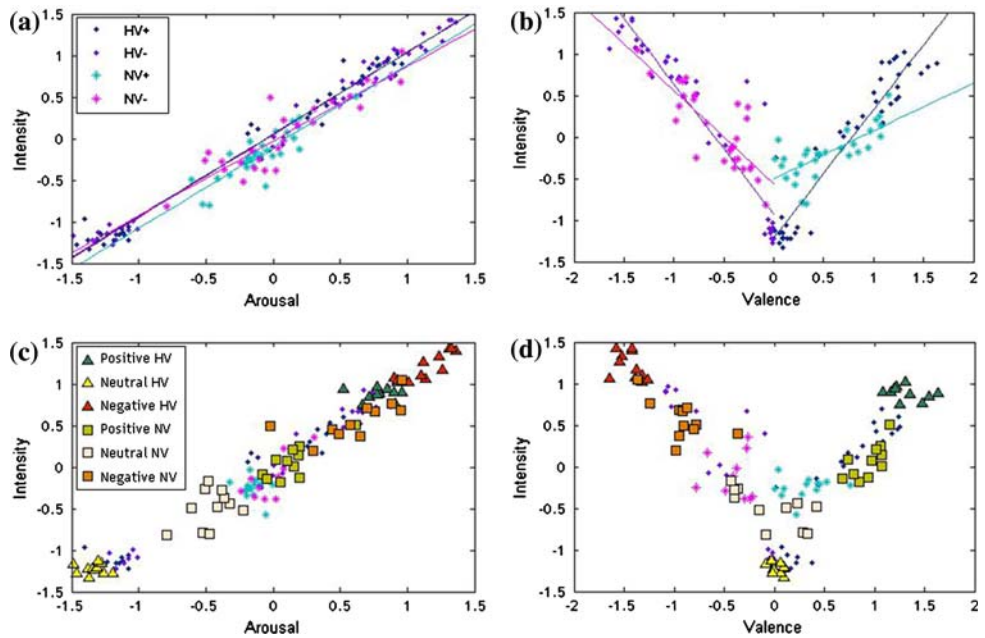


Fig. 2 Mean ratings obtained for human vocalizations (HV) and non-vocalizations (NV) judged as positive or negative for all 141 sounds (**a**, **b**) and for the battery of 66 extreme-most positive, negative, and neutral sounds (**c**, **d**). (**a**) shows a linear relationship between arousal and intensity, suggesting that these two emotional dimensions are tightly linked, irrespective of the sound source. In (**b**) a bilinear relationship between valence and intensity is observed. The 66 sounds selected with the PCA algorithm are presented in (**c**) and (**d**) and are superimposed on the plots shown in (**a**) and (**b**), respectively. Triangles represent human vocalizations (HV) and squares non-vocalizations (NV). Clusters of emotional sounds are significantly distinct from each other, both according to the three emotional dimensions and to sound source category

ratings. Two-dimensional projections of the 141 sounds tested within the affective space showed a positive linear correlation between mean ratings of arousal and emotional intensity ($r_{(139)} = 0.98$, $P < 0.01$, see Fig. 2a). The more a sound was judged as arousing the stronger was the intensity of the experienced emotion. This pattern was observed for both sound categories (HV: $r_{(82)} = 0.99$, $P < 0.01$ and NV: $r_{(55)} = 0.91$, $P < 0.01$). Plots of emotional intensity versus valence ratings reproduced the typical bilinear relationship ("boomerang-shape", see Fig. 2b) reported by Bradeley and Lang [20] for arousal and valence ratings of longer-lasting sounds. Thus, the more pleasant or unpleasant a sound was rated on the valence scale, the more arousing and emotionally intense it was rated as well. The bilinear correlation coefficients between pleasure and emotional intensity are given in Table 1 for each of the four subsets of sounds (positive and negative human vocalizations, HV+ and HV−, as well as non-vocalizations, NV+ and NV−). These results show that the emotional perceptions previously described for 6 s long sounds [20] can be reliably reproduced for sounds of 2 s duration, for both unpleasant and pleasant stimuli.

Table 1 Pearson correlation coefficients for mean ratings of the four sub-sets of emotional sounds. Positive and negative human vocalizations and non-vocalizations respectively; HV+, HV−, NV+, and NV−, respectively

Pearson correlation coeff.		Valence > 0	Valance < 0
Emo Intensity × Valence	HV	(HV+) $r_{(43)} = 0.95$ $p < 0.01$	(HV−) $r_{(37)} = 0.93$ $p < 0.01$
	NV	(NV+) $r_{(27)} = 0.72$ $p < 0.01$	(NV−) $r_{(26)} = 0.82$ $p < 0.01$
Arousal × Valence	HV	(HV+) $r_{(43)} = 0.94$ $p < 0.01$	(HV−) $r_{(37)} = 0.93$ $p < 0.01$
	NV	(NV+) $r_{(27)} = 0.75$ $p < 0.01$	(NV−) $r_{(26)} = 0.76$ $p < 0.01$

Pre-eminence of Human Vocalizations

The PCA analysis led to the subdivision of the collective 141 sounds into five emotional sub-classes that we refer to as extremely positive, moderately positive, neutral, moderately negative, and extremely negative. These ratings were submitted to a two-way multivariate analysis of variance (MANOVA) with the five rating questions as dependent variables (valence, emotional intensity, arousal, source identification, perceived loudness). Sound category (HV and NV) and the above PCA-defined sub-classes were used as between-subject factors[3].

Aside from the built-in main effect of PCA-defined emotional sub-class, we observed a main effect of sound category (HV vs. NV) in the emotional intensity ratings ($F_{(1,131)} = 4.309$, $P < 0.05$). HV were reliably perceived at a higher emotional intensity than NV (HV: 0.05 ± 0.94; NV: -0.05 ± 0.42). We also observed a main effect of sound category for source identification ($F_{(1,131)} = 16.570$, $P < 0.01$) and for perceived loudness ($F_{(1,131)} = 12.869$, $P < 0.01$). HV relative to NV resulted in generally higher confidence ratings in source identification (HV: 0.15 ± 0.32; NV: -0.12 ± 0.60) and lower perceived loudness ratings (HV: -0.10 ± 0.27; NV: 0.11 ± 0.38). By contrast, no reliable main effects were observed for either the arousal ($F_{(1,131)} = 0.400$, $P = 0.53$) or valence ratings ($F_{(1,131)} = 0.043$, $P = 0.84$). There was also a main effect of emotional sub-class for the questions regarding source identification ($F_{(4,131)} = 5.415$, $P < 0.01$) and perceived loudness ($F_{(4,131)} = 7.617$, $P < 0.01$), such that distinct ratings were observed as a function of membership in one of the five emotional sub-classes.

Significant interactions between the factors of sound category and emotional sub-class were observed for each of the three emotional questions (valence: $F_{(4,131)} = 7.792$, $P < 0.01$; emotional intensity: $F_{(4,131)} = 8.637$, $P < 0.01$; arousal: $F_{(4,131)} = 9.396$, $P < 0.01$). These interactions provide an indication that HV are rated reliably differently from their NV counterparts for each of the five sub-classes. This result can be visualized in Fig. 2b where it can be observed that the linear correlations are not superimposed for HV and NV stimuli. For perceived loudness, there was a significant interaction between the factors sound category and emotional sub-class ($F_{(4,131)} = 3.229$, $P < 0.05$), indicating that the perceived loudness of HV was rated reliably different from that of the NV counterparts for each of the PCA-defined emotional sub-classes, even though all stimuli were RMS normalized (see Materials and methods, above).

Extreme-most Affective Sounds

The PCA-defined sub-classes were used to identify an equal number (i.e. 11) of extremely positive, neutral, and extremely negative stimuli for each sound category (HV and NV compare Fig. 2a and b with Fig. 2c and d). A sound battery was comprised of the 11 extreme-most HV+, HV−, NV+ and NV− as well as 11 neutral sounds from

[3] Note that in this MANOVA, each sound is effectively treated as a unique "subject" or observation. Given that each sound was classified to one and only one PCA-defined sub-class, this was a between-subject factor (i.e. each PCA-defined sub-class is effectively a different "group"). Data from individual participants were not separately entered into this MANOVA. Rather, mean *Z*-scores were entered. We would also note that in order to present our findings in as clear a manner as possible and also given our interest in the 66 extreme-most sounds, follow-up contrasts were not conducted for this MANOVA.

each category (Fig. 2c and d). It is unambiguous with regard to the emotional classification of each sound and also contains equivalent numbers of each stimulus type that in turn could be evaluated and controlled along purely acoustic dimensions (see [27] this volume for methodological details; results are presented in the electronic supplementary materials). We assessed via MANOVA whether these extreme-most sounds also exhibited the above pattern of results observed with the original set of 141 sounds.

We observed a main effect of sound category for each question except the valence ratings (valence: $F_{(1,60)} = 0.034$, $P = 0.86$; emotional intensity: $F_{(1,60)} = 29.322$, $P < 0.01$; arousal: $F_{(1,60)} = 6.248$, $P < 0.05$; source identification: $F_{(1,60)} = 6.618$, $P < 0.05$; perceived loudness: $F_{(1,60)} = 4.802$, $P < 0.05$), again supporting the distinction between HV and NV stimuli. Apart from the built-in main effect of emotional sub-classes on the three emotional questions, there was also a main effect of emotional sub-class for both source identification ($F_{(2,60)} = 9.676$, $P < 0.01$) and perceived loudness ($F_{(2,60)} = 10.144$, $P < 0.01$), showing that the perceived emotion impacts a listener's ability to confidently recognize the sound and to judge its volume. Follow-up contrasts (independent samples two-tailed *t*-tests with unequal variance assumed) showed that the extreme-most positive sounds (from both categories) were recognized with more confidence than either the extreme-most negative sounds ($t_{(25.25)} = -2.16$, $P < 0.05$) or neutral sounds ($t_{(24.22)} = -4.47$, $P < 0.01$). Additionally, the extreme-most negative sounds were better identified than neutral sounds ($t_{(40.85)} = 2.08$, $P < 0.05$). Interestingly, as can be seen in Fig. 3a, positive HV were recognized significantly more confidently than all other sub-classes (neutral HV: $t_{(12.69)} = -8.46$, $P < 0.01$; neutral NV: $t_{(10.15)} = -2.45$, $P < 0.01$; negative HV: $t_{(14.02)} = -3.28$, $P < 0.01$; negative NV: $t_{(10.22)} = -2.48$, $P < 0.05$; positive NV: $t_{(12.84)} = -3.07$, $P < 0.01$). In contrast, both neutral HV and neutral NV separately were recognized with significantly less confidence than negative HV (neutral HV: $t_{(19.16)} = 4.75$, $P < 0.01$; neutral NV: $t_{(10.73)} = -2.73$, $P < 0.05$) and positive NV (neutral HV: $t_{(19.98)} = 4.18$, $P < 0.01$; neutral NV: $t_{(11.05)} = -2.65$, $P < 0.05$). More generally, this pattern suggests that emotion can facilitate recognition (see also [10]). Further extending this result is our observation that perceived loudness was also significantly higher for negative sounds of both sound categories. Analyses on the perceived loudness question showed (see Fig. 3b) that negative extreme-most sounds (both HV and NV altogether) were perceived slightly but significantly louder than extreme-most positive sounds ($t_{(38.96)} = -4.04$, $P < 0.01$) and neutral sounds ($t_{(40,60)} = -3.32$, $P < 0.01$). Our acoustic analyses further argue against an explanation of this result in terms of physical features (see electronic supplementary materials).

Interactions between sound category and emotional subclass were significant for each of the three emotional questions (arousal: $F_{(2,60)} = 122.471$, $P < 0.01$; emotional intensity: $F_{(2,60)} = 125.880$, $P < 0.01$; valence: $F_{(2,60)} = 27.081$, $P < 0.01$). No such interactions were observed for the two non-emotional questions (source identification: $F_{(2,60)} = 0.267$, $P = 0.77$; perceived loudness: $F_{(2,60)} = 0.209$, $P = 0.81$). In light of these interactions, a series of post-hoc analyses (independent samples two-tailed *t*-tests with unequal variance assumed) were performed; the results of which are presented in Table 2. All results were significant except for two. Arousal ratings for positive HV and negative NV samples did not significantly differ. Also,

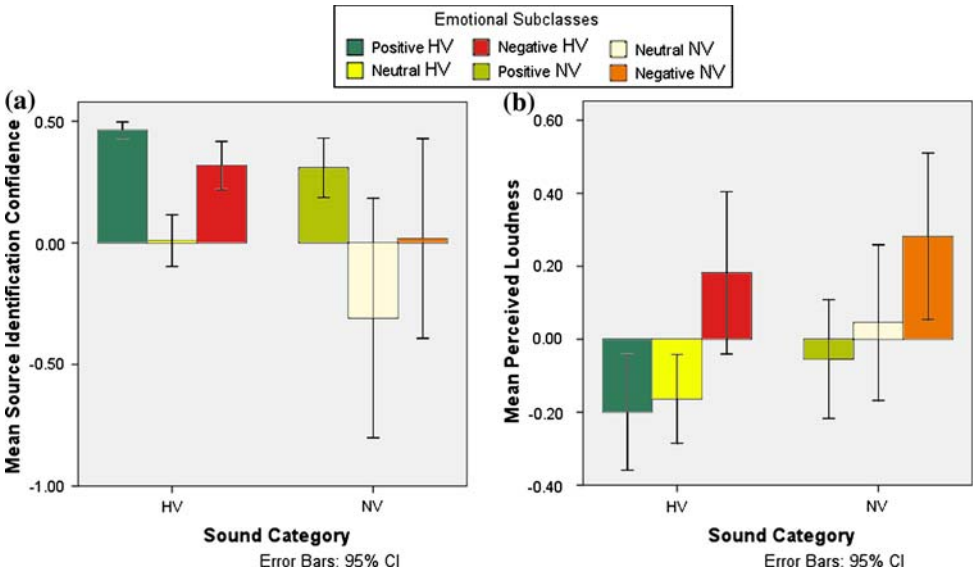
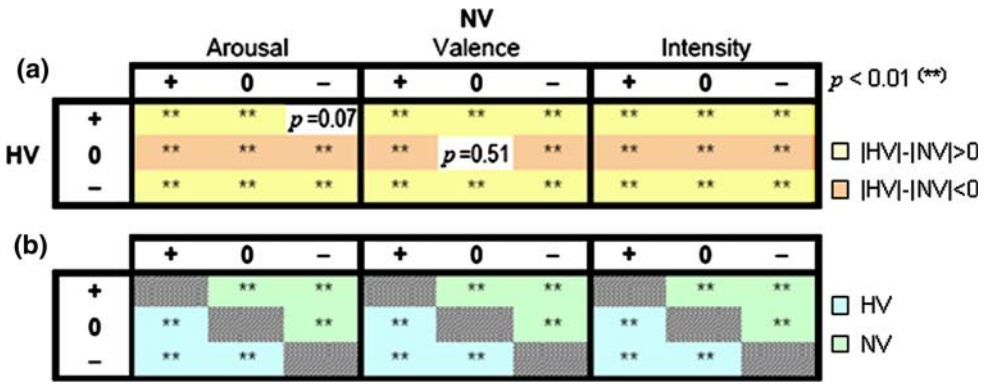


Fig. 3 Mean ratings (±confidence interval, $P < 0.05$) for (**a**) confidence in source identification, and (**b**) perceived loudness for the battery of 66 extreme-most sounds. Positive HV obtained the highest confidence ratings, conveying stronger meaning than all other sub-classes of sounds. Negative stimuli from both categories (HV and NV) were perceived as louder than their positive and neutral counterparts

Table 2 Follow-up contrasts between PCA built-in valence categories: positive (+), neutral (0) and negative (−)



(a) Between class (HV vs NV) comparison; (b) within class comparison (Blue: within HV; green: within NV)

valence ratings for both neutral samples did not significantly differ, as would be expected from the PCA analysis. In summary, when the extreme-most emotional sub-classes were analyzed, HV received more extreme ratings than NV on the three emotional questions for each valence (positive, negative, and neutral), supporting the consideration of emotional HV as a pre-eminent category of sounds.

Discussion

This study established a battery of equivalent-duration and acoustically controlled emotional sounds containing extreme-most stimuli for each valence (positive, negative, and neutral) and for two categories of environmental sounds (HV and NV). Our main findings can be summarized as follows. First, fixed-duration auditory stimuli (here 2 s long) yielded a highly similar pattern of emotional ratings as previously observed with longer, variable-duration stimuli [20]. Second, HV were emotionally pre-eminent as they yielded stronger affective ratings along the three emotional dimensions under consideration (valence, emotional intensity, and arousal) than other environmental sounds. Third, positive HV were identified with the most confidence, and neutral sounds of either category with the least. Fourth, extreme-most negative sounds of both categories (HV and NV) were generally perceived as louder than either positive or neutral sounds.

Short-duration Emotional Sound Battery

Brain imaging studies of auditory emotion processing have thus far been relatively rare in part because of the challenges in constructing a sound battery that not only is controlled psychophysically and acoustically, but also consists of sufficiently short duration stimuli such that inferences can be drawn about the temporal dynamics of emotion processing. As such, it is equally important for the stimuli within any such battery to be of equal duration to facilitate control of acoustic parameters (see electronic supplementary materials and Knebel et al., this issue). While the battery of Bradley and Lang [24] contains psychophysically controlled sounds with reliable ratings for the same emotional dimensions studied here, the stimuli are both long and of variable duration (maximally 6 s). In addition, this battery contains no controls for low-level variance in acoustic features. An important consequence is that such a battery is sub-optimal for brain imaging investigations in general and for studying the temporal dynamics of emotion processing, in particular. Two recent event-related potential (ERP) investigations have addressed the temporal dynamics of emotion processing. Thierry and Roberts [12] presented neutral and negative sounds with a mean duration of >1 s in an oddball paradigm and observed effects of emotional valence only at latencies ~300 ms post-stimulus. By contrast, Czigler et al. [11] presented listeners with short-duration (350–500 ms) sounds that were rated as either neutral or aversive and obtained ERP effects of valence at ~150 ms post-stimulus onset. It is noteworthy that in neither study were emotionally positive stimuli studied; in fact, Czigler et al. [11] state they were unable to reliably obtain positively-rated stimuli with this duration stimulus. In addition, the role of low-level acoustic features was not controlled (aside from peak volume) and thus cannot be unequivocally excluded as a confounding factor. In addition, the use of stimuli of different duration may explain some of their findings [6]. Finally and independent of the stimuli used, the ERP analyses performed by both Thierry and Roberts [12] and Czigler et al. [11] do not provide information about the likely underlying mechanism or sources of their effects (see [28] this issue for discussion). We highlight the above shortcomings to emphasize the necessity for the stimulus battery developed in the present study, as well as the continued investigation of the spatio-temporal dynamics of auditory-induced emotion processing. Specifically, our battery fulfils the criteria of relatively short and equal duration across all stimuli as well as the ability to reliably elicit emotions of each valence (see Fig. 2c and d). A further advantage of the generated battery is the availability

of multiple sound categories; in particular human vocalizations and non-vocalizations. Prior studies provide evidence that different object categories may engage distinct brain networks [7, 18, 29], including the possibility of differential responses within the amygdala as a function of emotional valence [8]. Ongoing investigations by our group are addressing the issue of categorical discrimination and the impact of emotional valence. Future studies can also address the question of age-dependent changes in the perception of emotional sounds [21, 30, 31] as well as the integrity of auditory emotion perception in clinical and developmental populations. The battery of sounds developed here can also be used for evaluating sub-classes of sounds from both sound categories that receive similar ratings in order to isolate categorical effects and/or effects of the specific questions evaluated in this study.

Pre-eminence of Human Vocalizations

There is mounting neuroimaging evidence that regions of the anterior temporal lobe, in particular the right anterior superior temporal sulcus, are specialized for the processing of human paralinguistic vocalizations (reviewed in Belin et al. [32], see also Grandjean et al. [2], Meyer et al. [23] Ethofer et al. [3, 33] Schirmer et al. [34] for evidence concerning the impact of vocal prosody). Our results extend these findings to show that the emotional content of human non-linguistic vocalizations also plays a central role in distinguishing such stimuli from other categories of environmental sounds. Our data provide evidence that even if HV and NV elicit the same perceived valence and arousal, HV are nonetheless perceived at a reliably distinct emotional intensity; the direction of which varied as a function of emotional valence. Specifically, the more emotionally extreme a given HV was perceived, the more distinct from the corresponding valence of NV it became in each of the three emotional dimensions we evaluated. Both the positive-most and negative-most HV stimuli were not only perceived as emotionally more intense than the corresponding NV stimuli, but were also judged as more pleasant/unpleasant, respectively, and gave rise to higher arousal ratings. By contrast, the neutral-most HV were rated as emotionally less intense and induced a lower arousal rating than the neutral-most NV, despite their sharing an equally null emotional valence rating.

Confidence ratings in listeners' ability to identify the sound sources provide additional support for the proposition that HV constitute a distinct category of (emotional) auditory stimuli. While all sounds were reliably recognized (mean $\pm$ sd rating prior to Z-score transformation was 1.5 ± 1.0 over the original 144 sounds), it is also apparent from Fig. 3a that the standard deviations for different sound categories and emotional subclasses were heterogeneous. Emotionally positive stimuli yielded a tight distribution in confidence ratings, whereas such was visibly wider for neutral and negative NV sub-classes, suggesting that these latter stimuli elicited a broader range of confidence in their identification. In addition, HV+ resulted in a significantly higher confidence rating than all other conditions (see Fig. 3a). We would note that this might be related to NV originating from a more varied set of sources (household objects, vehicles, etc.), whereas HV were forcibly from a less varied set (i.e. humans). However, it is not readily apparent why such lack of variability would specifically affect confidence ratings for HV+ instead of HV generally. Further investigation will be required to resolve the role of source variability in auditory object processing.

Participants were most confident in source identification when they were confronted with a positive HV. Prior related research has shown that emotionally positive human vocalizations are a particularly effective auditory stimulus for activating pre-motor networks considered part of the mirror neuron system [35]. These authors interpreted the involvement of such circuitry in passive listening to reflect the automatic preparation of emotion-appropriate vocal or facial gestures. How such activity might contribute to processes underlying object identification will require additional investigation. However, the general consistency across studies supports the proposition that human vocalizations are both a distinct perceptual category that activates a partially segregated cortical network that might itself in turn be facilitated by the emotional valence of the sounds. In contrast, participants were least confident when identifying neutral sounds of either category. However, neutral HV such as vocalizations that lack prosodic information are rarely heard in everyday life. When asked, participants admitted having recognized that the sound was made by a human voice but did not understand the context and were consequently less confident in "identifying the sound source". Although the context for NV was more realistic (typewriter, train, river, wind etc.), their identification was paradoxically more uncertain. Further investigations are necessary to determine the basis for this uncertainty and the contribution of the nature of the sound source itself and/or the emotion it conveys. For example, it will be particularly interesting to determine whether (and when) superior temporal brain regions considered specialized for the processing of vocal prosody (e.g. [2, 3, 33]) are equally well engaged in the processing of non-linguistic emotional HV and/or NV. Given the recent fMRI evidence from Fecteau et al. [8] for the involvement of a widely distributed network of temporal (including primary auditory cortex), frontal, and limbic (amygdale) structures in the differential processing of emotional non-linguistic HV stimuli, it will be important for the construction of a model

of auditory emotion processing to determine the relative timing, using techniques such as electrical neuroimaging, when each of these brain regions exhibits its differential response.

Negative Stimuli are Perceived as Louder

The rapid recognition of threat in the environment is critical for survival. Thus, emphasizing the processing of negative or aversive auditory stimuli would provide an evolutionary advantage. Although the volume of the whole set of sounds was RMS normalized (see electronic supplementary material), the extreme-most negative stimuli were perceived as significantly louder than their extreme-most positive or neutral counterparts for both sound categories (HV and NV). This effect is unlikely to be specifically linked to low-level acoustic features, because this effect was observed for both sound categories and because our time-frequency analyses revealed no reliable differences between negative and positive stimuli (see electronic supplementary material). Still, this effect was further enhanced for the negative-most HV, raising the possibility of cumulative or integrative effects of general emotion and categorical processes. We would note that these stimuli did not differ from their positive-most HV counterparts in their major acoustic features (mean F0, mean F0 variability, or physical intensity). One possibility is that such an (illusory) perceived loudness for negative stimuli might derive from the allocation of increased attentional resources that in turn facilitate the discernment of negative stimuli within auditory scenes [12, 36].

References

1. Scherer KR. Vocal communication of emotion: a review of research paradigms. Speech Commun. 2003;40:227–56
2. Grandjean D, Sander D, Pourtois G, Schwartz S, Seghier ML, Scherer KR, Vuilleumier P. The voices of wrath: brain responses to angry prosody in meaningless speech. Nat Neurosci. 2005;8(2):145–6
3. Ethofer T, Anders S, Wiethoff S, Erb M, Herbert C, Saur R, Grodd W, Wildgruber D. Effects of prosodic emotional intensity on activation of associative auditory cortex. Neuroreport 2006;17(3):249–53
4. Peretz I. The nature of music from a biological perspective. Cognition 2006;100(1):1–32
5. Zatorre RJ, Chen JL, Penhune VB. When the brain plays music: auditory-motor interactions in music perception and production. Nat Rev Neurosci. 2007;8(7):547–58
6. Wiethoff S, Wildgruber D, Kreifelts B, Becker H, Herbert C, Grodd W, Ethofer T. Cerebral processing of emotional prosody—influence of acoustic parameters and arousal. Neuroimage 2008;39(2):885–93
7. Belin P, Fecteau S, Bédard C. Thinking the voice: neural correlates of voice perception. Trends Cogn Sci. 2004;8(3):129–35
8. Fecteau S, Belin P, Joanette Y, Armony JL. Amygdala responses to nonlinguistic emotional vocalizations. NeuroImage 2007;36(2):480–7
9. Von Kriegstein K, Smith DR, Patterson RD, Ives DT, Griffiths TD. Neural representation of auditory size in the human voice and in sounds from other resonant sources. Curr Biol. 2007;17(13):1123–8
10. Armony JL, Chochol C, Fecteau S, Belin P. Laugh (or cry) and you will be remembered. Psychol Sci. 2007;18(12):1027–9
11. Czigler I, Cox TJ, Gyimesi K, Horváth J. Event-related potential study to aversive auditory stimuli. Neurosci Lett. 2007;420(3):251–6
12. Thierry G, Roberts MV. Event-related potential study of attention capture by affective sounds. Neuroreport 2007;18(3):245–8
13. Ghazanfar AA, Hauser MD. The auditory behaviour of primates: a neuroethological perspective. Curr Opin Neurobiol. 2001;11(6):712–20
14. Rolls ET. Neurophysiological mechanisms underlying face processing within and beyond the temporal cortical visual areas. Philos Trans R Soc Lond B Biol Sci. 1992;335(1273):11–20
15. Allison T, Ginter H, McCarthy G, Nobre AC, Puce A, Luby M, Spencer DD. Face recognition in human extrastriate cortex. J Neurophysiol. 1994;71(2):821–5
16. Rolls ET. The representation of information about faces in the temporal and frontal lobes. Neuropsychologia 2007;45(1):124–43
17. McKone E, Kanwisher N, Duchaine BC. Can generic expertise explain special processing for faces? Trends Cogn Sci. 2007;11(1):8–15
18. Murray MM, Camen C, Gonzalez Andino SL, Bovet P, Clarke S. Rapid brain discrimination of sounds of objects. J Neurosci. 2006;26(4):1293–302
19. Osgood C, Suci G, Tannenbaum P. The measurement of meaning. University of Illinois Press; (1957)
20. Bradley MM, Lang PJ. Affective reactions to acoustic stimuli. Psychophysiology 2000;37(2):204–15
21. Fecteau S, Armony JL, Joanette Y, Belin P. Judgment of emotional nonlinguistic vocalizations: age-related differences. Appl Neuropsychol. 2005;12(1):40–8
22. Morris JS, Scott SK, Dolan RJ. Saying it with feeling: neural responses to emotional vocalizations. Neuropsychologia 1999;37(10):1155–63
23. Meyer M, Zysset S, von Cramon DY, Alter K. Distinct fMRI responses to laughter, speech, and sounds along the human perisylvian cortex. Cogn Brain Res. 2005;24(2):291–306
24. Bradley MM and Lang PJ. International affective digitized sounds (IADS): Stimuli, instruction manual and affective ratings (Tech. Rep. No. B-2). Gainesville, FL: The Center for Research in Psychophysiology, University of Florida; 1999
25. Clark-Carter D. Doing quantitative psychological research. From design to report. East Sussex, UK: Psychology; 1997
26. Hastie T, Tibshirani R, Friedman JH. The elements of statistical learning. Springer Series in Statistics; 2004
27. Knebel JF, Toepel U, Hudry J, le Coutre J, Murray MM. Generating controlled image sets in cognitive neuroscience research. Brain Topogr. 2008. doi:10.1007/s10548-008-0046-5.
28. Murray MM, Brunet D, Michel CM. Topographic ERP analyses: A step-by-step tutorial review. Brain Topogr. 2008. doi:10.1007/s10548-008-0054-5.
29. Lewis JW, Brefczynski JA, Phinney RE, Janik JJ, DeYoe EA. Distinct cortical pathways for processing tool versus animal sounds. J Neurosci. 2005;25(21):5148–58
30. Kiss I, Ennis T. Age-related decline in perception of prosodic affect. Appl Neuropsychol. 2001;8(4):251–4
31. Phillips LH, MacLean RD, Allen R. Age and the understanding of emotions: neuropsychological and sociocognitive perspectives. J Gerontol B Psychol Sci Soc Sci. 2002;57(6):526–30
32. Belin P. Voice processing in human and non-human primates. Philos Trans R Soc Lond B Biol Sci. 2006;361(1476):2091–107

33. Ethofer T, Wiethoff S, Anders S, Kreifelts B, Grodd W, Wildgruber D. The voices of seduction: cross-gender effects in processing of erotic prosody. Soc Cog Affect Neurosci. 2007;2:334–7
34. Schirmer A, Simpson E, Escoffier N. Listen up! Processing of intensity change differs for vocal and nonvocal sounds. Brain Res. 2007;1176:103–12
35. Warren JE, Sauter DA, Eisner F, Wiland J, Dresner MA, Wise RJ, Rosen S, Scott SK. Positive emotions preferentially engage an auditory-motor "mirror" system. J Neurosci. 2006;26(50): 13067–75
36. Sander D, Grandjean D, Pourtois G, Schwartz S, Seghier ML, Scherer KR, Vuilleumier P. Emotion and attention interactions in social cognition: brain regions involved in processing anger prosody. Neuroimage 2005;28(4):848–58