Emotion

© 2022 American Psychological Association
ISSN: 1528-3542

2022, Vol. 22, No. 1, 213–225
https://doi.org/10.1037/emo0001048

The Variably Intense Vocalizations of Affect and Emotion (VIVAE) Corpus Prompts New Perspective on Nonspeech Perception

Natalie Holz[1], Pauline Larrouy-Maestri[1, 2], and David Poeppel[1, 2, 3]
[1] Department of Neuroscience, Max-Planck-Institute for Empirical Aesthetics, Frankfurt, Germany
[2] Center for Language, Music, and Emotion (CLaME), New York, New York, United States
[3] Ernst Struengmann Institute for Neuroscience, Frankfurt, Germany

The human voice is a potent source of information to signal emotion. Nonspeech vocalizations (e.g., laughter, crying, moans, or screams), in particular, can elicit compelling affective experiences. Consensus exists that the emotional intensity of such expressions matters; however *how* intensity affects such signals, and their perception remains controversial and poorly understood. One reason is the lack of appropriate data sets. We have developed a comprehensive stimulus set of nonverbal vocalizations, the first corpus to represent emotion intensity from one extreme to the other, in order to resolve the empirically underdetermined basis of emotion intensity. The *full set*, comprising 1085 stimuli, features eleven speakers expressing 3 positive (achievement/triumph, sexual pleasure, surprise) and 3 negative (anger, fear, physical pain) affective states, each varying from low to peak emotion intensity. The smaller *core set* of 480 files represents a fully crossed subsample (6 emotions × 4 intensities × 10 speakers × 2 items) selected based on judged authenticity. Perceptual validation and acoustic characterization of the stimuli are provided; the expressed emotional intensity, like expressed emotion, is reflected in listener evaluation and signal properties of nonverbal vocalizations. These carefully curated new materials can help disambiguate foundational questions on the communication of affect and emotion in the psychological and neural sciences and strengthen our theoretical understanding of this domain of emotional experience.

Keywords: voice, nonverbal vocalizations, emotion perception, emotion intensity, database

Supplemental materials: https://doi.org/10.1037/emo0001048.supp

A crucial aspect of understanding others' thoughts and feelings is to infer meaning from rich sensory signals, such as the human body, face, or voice. The inferred meaning of the expressions is, generally speaking, substantially aligned with the affective content expressed—yet meaningful variation exists. Across sensory modalities, the *emotional intensity* of the expression heavily shapes how clear-cut the inferred emotional meaning is perceived. A large body of work suggests that increasing emotional intensity facilitates emotion perception (Bänziger et al., 2012; Hess et al., 1997; Juslin & Laukka, 2001; Livingstone & Russo, 2018; Wingenbach et al., 2016). In contrast, a range of studies reports greater ambiguity in the valuation of extremely intense emotion (Anikin & Persson, 2017; Atias et al., 2019; Aviezer et al., 2012; 2017). To date, it has been complicated to adjudicate between the alternative findings: No dataset exists to formally test the relation of emotional intensity and ease of recognition, and it is unclear—and much debated—how perceptual versus signal information vary as a function of emotional intensity.

Emotional intensity is considered a central aspect of emotion under various theoretical accounts (Ekman, 1984; Ekman & Cordaro, 2011; Frijda et al., 1992; Russell & Barrett, 1999; Scherer, 2005). Empirically, the concept is employed broadly across research domains, such as cognitive and clinical psychology (Anikin, 2020a; Bernat et al., 2006; Grossman & Tager-Flusberg, 2012; Rutter et al., 2019), neuroscience (Blood & Zatorre, 2001; Bonnet et al., 2015; Ethofer et al., 2006; Frühholz et al., 2014; Kragel et al., 2018; Wang et al., 2017), and biology (Belin & Zatorre, 2015; Rendall, 2003). It is typically

Natalie Holz https://orcid.org/0000-0003-2360-9428
Pauline Larrouy-Maestri https://orcid.org/0000-0001-9245-0743
David Poeppel https://orcid.org/0000-0003-0184-163X

Data is available at https://osf.io/rwzkf/. The VIVAE corpus is available at http://doi.org/10.5281/zenodo.4066235.

We thank all speakers for their participation. We thank Susan Rogers for assistance with the implementation of the recordings at Berklee and Andres Abello for technical support during the recording session, Dominik Thiele and R. Muralikrishnan for help running the experiments, and Jakob von Petersdorff and Klaus Frieler for support with the data organization. This work was supported by The Max Planck Society.

Natalie Holz, Pauline Larrouy-Maestri, and David Poeppel designed this study. Natalie Holz and Pauline Larrouy-Maestri collected the data. Natalie Holz analyzed the data. Natalie Holz, Pauline Larrouy-Maestri, and David Poeppel wrote the article. All authors approved the final version of the article for submission.

The authors declare no competing interests.

Correspondence concerning this article should be addressed to Natalie Holz, Department of Neurosciences, Max-Planck-Institute for Empirical Aesthetics, Grüneburgweg 14, 60322 Frankfurt, Germany. Email: natalie.holz@ae.mpg.de

agreed *that*—but not *how*—intensity affects perceptual and signal properties. However, stimulus materials with varying levels of intensity are sparse, and to our knowledge, the systematic manipulation of the intensity dimension over its whole range, namely from weak to maximally high emotional intensity, has not yet been realized.

In the focus of the current work are nonverbal emotional expressions. They encompass a large variety of acoustically and perceptually distinguishable types (Anikin & Lima, 2018; Anikin et al., 2018); including screams, laughs, moans, and cries. Their role as a communicative signal is remarkable: As relatively unrestrained communication signals, likely driven by physiological effects on voice (Patel et al., 2011), nonverbal expressions are often regarded as distinct from speech and instead parallel to primate or infant vocalizations (Belin et al., 2008; Pell et al., 2015; Scott et al., 2010). Effects of increased emotional intensity and arousal have been linked to little voluntary regulation and sociocultural dependency (Anikin & Persson, 2017; Juslin, 2013). Moreover, from a phylogenetic perspective, different mechanisms have been proposed by which volitional and spontaneous vocalizations are produced (Bryant & Aktipis, 2014; on a similar conceptual account, note, as well, the *push* and *pull effects* proposed by Scherer, 1989, 1994, and *natural* and *conventional signs* discussed by Wharton, 2003). As such, nonverbal vocalizations occupy a peculiar niche in the human vocal repertoire: They comprise variable degrees of spontaneity, cognitive control, social learning, and culture (e.g., Bryant et al., 2018; Gendron et al., 2014), and we submit that the intensity dimension is a promising candidate to access gradations of each.

Given the expressive diversity, a second important objective of this work (besides the representation of the emotion intensity dimension), was to privilege naturalness over recognizability. The presumed existence of robust diagnostic patterns that allow emotion discrimination is widely debated (Bachorowski, 1999; Barrett, 2017a; Fridlund, 2017; Gendron et al., 2018), along with the concomitant cognitive and theoretical implications (Barrett, 2017b; LeDoux et al., 2016; LeDoux & Brown, 2017; Wager et al., 2015). As such, this central methodological consideration is crucial to overcome the circularity oftentimes encountered when testing properties (e.g., recognizability) of emotion portrayals that were produced and selected *to be* diagnostic and prototypical (Fridlund, 2017). In most vocal expression corpora, stimulus selection is linked to factors like typicality or recognizability, which oftentimes serve as measures of validity during materials construction (Belin et al., 2008; Cordaro et al., 2016; Hawk et al., 2009; Lima et al., 2013; Maurage et al., 2007; Sauter, Eisner, Calder, & Scott, 2010; Schröder, 2003; see Table S1 for an overview of existing stimulus sets, along with some central methodological decisions linked to their creation and validation.) For some, expressive variation is further confined by experimenter choice of a specific target sound type (Bänziger et al., 2012; Belin et al., 2008; Schröder, 2003) or by feedback concerning the decodability of the produced expressions throughout the recording session (Belin et al., 2008; Livingstone & Russo, 2018). As a side effect, natural variability is minimized, and prototypical expressions are favored.

In the current study, genuineness and expressive diversity were thus key aspects throughout the design of the materials, extending previous approaches to value authenticity and naturalness (e.g., Bänziger et al., 2012; Hawk et al., 2009; Laukka et al., 2013; Lima et al., 2013; Livingstone & Russo, 2018). Authenticity was defined as a central validation criterion. During stimulus production, encouraging expressive spontaneity rather than providing guidance or limitation on how to express a specific affective state, may aid bolster genuineness (Laukka et al., 2013; Lima et al., 2013). Induction methods, such as felt experience enacting, may facilitate more natural and variable enactment (Bänziger et al., 2012). The choice of speakers itself (i.e., actors vs. nonactors) can affect how expressions are perceived (Krahmer & Swerts, 2008; Spackman et al., 2009). A study by Jürgens et al. (2015) showed that the acoustics of nonactor expressions resembled authentic (i.e., nonstaged) expressions more than vocal materials coming from professional actors. These results support the hypothesis that professional acting training may lead to particular vocal patterns (Jürgens et al., 2011; 2015) and stereotypical expressions (Laukka et al., 2012; Scherer, 2003), and can be interpreted as evidence in favor of nonprofessional actors. Nevertheless, possible drawbacks of working with nonactors may be related to the lack of stage experience and acting training (Bänziger et al., 2012; Scherer & Bänziger, 2010). In addition, lay persons, unfamiliar with voice performance and studio settings, might experience a recording session as more stressful than professional actors (for a review on the influence of psychological stress on vocal production, see Larrouy-Maestri & Morsomme, 2014). In an (admittedly unconventional) attempt, we therefore recorded singers—nonactors with vocal performance training—to balance possible advantages and disadvantages of working with nonactors.

We report here the construction, validation, and acoustic characterization of the *Variably Intense Vocalizations of Affect and Emotion Corpus* (VIVAE; Holz et al., 2020). VIVAE consists of a broad set of human nonspeech emotion vocalizations. It is the first database to incorporate the emotion intensity dimension over its whole range and thus to provide a comprehensive set to empirically investigate affect cognition. Furthermore, we offer a novel approach to circumvent the recognizability based stimulus design, often applied in studio-produced expression corpora. Both are key to address central aspects of vocal emotion communication. The *full set*, comprising 1085 audio files, features eleven speakers (all female) expressing three positive (achievement/triumph, sexual pleasure, surprise) and three negative (anger, fear, physical pain) affective states, each varied from low to peak emotion intensity. The smaller *core set* of 480 files represents a fully crossed subsample of the full set (6 emotions × 4 intensities × 10 speakers × 2 items) selected based on judged authenticity. This smaller set offers a balanced sample of suitable size for a variety of experimental settings. This allows users to draw on two sets of different size depending on the specific research needs. We show how the database—indeed even the design itself and its construction—enrich our theoretical understanding of emotion. Furthermore, we provide evidence that emotional intensity is a robustly represented aspect in perceptual and signal properties of nonverbal vocalizations.

Method

Construction Phase

Emotions

The emotions represented in the VIVAE corpus are *achievement/triumph, anger, fear, pain, sexual pleasure,* and *positive surprise.* They constitute a selection of commonly assessed emotional states (Anikin & Lima, 2018; Anikin & Persson, 2017; Aviezer et al., 2012; Bänziger et al., 2012; Fecteau et al., 2007; Gendron et al., 2014; Kamiloğlu et al., 2020; Sauter, Eisner, Ekman, & Scott, 2010; Sauter & Scott, 2007; see Table S1 for details on emotion categories included in nonverbal vocalization corpora) and were chosen as a suitable and well-studied sample of affective states. Aiming to span a broad affective space, the VIVAE comprises a set of equally many positive and negative states, in each case ranging from minimal to maximal emotion intensity. We selected emotions for which variations in emotion intensity had previously been described (Aviezer et al., 2012; Juslin & Laukka, 2001; Lima et al., 2013; Livingstone & Russo, 2018). The represented emotions, conceptualized as feelings or mental states that can be subjectively experienced (LeDoux et al., 2016), are of interest from a theoretical perspective, given the continuum of constructs (e.g., core affect, emotion, and mood) arising from contemporary emotion theories (Ekkekakis, 2013). While all emotions under study have been examined extensively with various methodological and theoretical approaches, allowing comparability with existing emotion communication research, some of them may be seen as more 'typical' than others, and conceptualization and terminology may diverge (for discussions on the nature of anger or fear vs. sexual pleasure, pain, and surprise, see, e.g., Johnson-Laird & Oatley, 1989; Ortony & Turner, 1990; but also Barrett, 2017a).

Speakers

Eleven female speakers (mean age = 25.00 years, *SD* = 6.59) participated in the recording sessions. They were undergraduate students at the Berklee College of Music in Boston. The procedures were approved by the Berklee College of Music Review Board (IRB). All reported voice training as well as stage experience as singers. None of them underwent formal acting training or had professional acting experience. Speakers were all fluent speakers of English, and their native languages included English (*n* = 6), Spanish (*n* = 3; two of them bilingual English and Spanish), Turkish (*n* = 1), and Russian (*n* = 1). By recording speakers with stage experience but no professional acting skills we intended to obtain highly expressive recordings despite the studio setting, yet as spontaneous and genuine as possible.

Recording Procedure

Speakers were instructed to vocally express six affective states (achievement/triumph, anger, fear, pain, positive surprise, and sexual pleasure), each with four levels of intensity (low, moderate, strong, and peak).

The recording session design was optimized to foster expressive variation and genuineness. Speakers were provided with a list of emotion terms as well as with a list of short real-life scenarios, which were inspired from previous research (e.g., Cordaro et al., 2016; Lima et al., 2013; Simon-Thomas et al., 2009). Scenarios were generated with high variability regarding the typically evoked emotion intensity (e.g., mild fear to panic, annoyance to fury). Furthermore, speakers were asked to recall similar situations of their own, in which they personally experienced the specified emotion. They were asked to try to put themselves in the same state of mind of either their personally experienced situation or, if not applicable, of the proposed scenarios. Speakers were instructed to produce the vocal sound they would personally make when experiencing the emotion, as spontaneous and natural as possible. Note that the speakers' subjective feelings or physiological states were not measured, as the primary goal was not to examine underlying emotional states in *speakers*. Rather, the procedure aimed at aiding speakers to produce believable expressions, based on real, spontaneous vocalizations (Banse & Scherer, 1996; Scherer & Bänziger, 2010). The order of emotions expressed was chosen by the speaker; if no order was chosen, recordings were taken in alphabetical order of the emotion terms. For each affective state, the recording was performed with increasing emotion intensity, that is, from a very mildly sensed affective state to an extremely intense affective state. Speakers were instructed to produce at least two sounds for each emotion at each intensity (with no upper limit specified). No guidance was provided as to the specific sound speakers should produce; the only instruction provided was to produce *nonverbal* vocalizations, that is, vocalizations without verbal content as in words (e.g., "yes," "no") or interjections carrying conventionalized meaning or lexical identity (e.g., "ouch," "yippee"). For example, vocalizations include moans, grunts, screams, or laughter, and show variability in their phone structure (e.g., *aah*, *mmh*, *uuh*, *oh*, etc.) within and across different emotions. Speakers were not interrupted when expressing their affective state verbally (e.g., swearing, cheering)—they were allowed to do so in order to maximize the naturalness of expression—yet any semantic content was excluded from further analysis. Recordings took place in a soundproof studio at Berklee College of Music, Boston, using Pro Tools LE (Version 12.7.1) software and DPA 4065 Microphone (at approximately constant distance to the mouth of the speaker), and digitized at a 44.1-kHz to 48-kHz sampling rate and 16-bit resolution. One session file per speaker was collected and stored in WAV file format. A total of 1250 individual files containing one vocalization each (59 to 111 files per speaker; *M* = 99.27, 95% CI [90, 108]) were obtained manually using Audacity (Version 2.1.2). Files were downsampled to 44.1-kHz using Praat (Version 6.41) and were submitted to further validation.

Technical Validation Phase

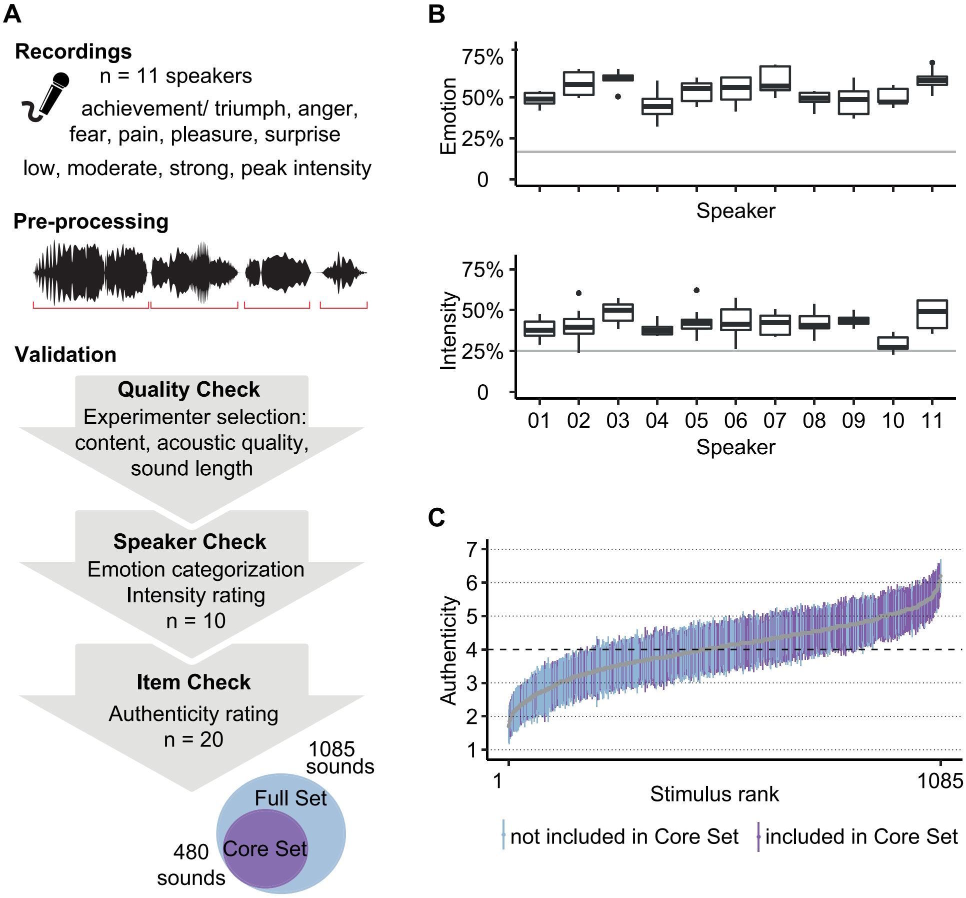
The technical validation of the corpus consisted of a three-step procedure: a *Quality check*, a *Speaker Check*, and an *Item Check* (Figure 1A). An important difference from previous studies characterizes this corpus: The validation was not based on the prototypicality or recognizability of individual stimuli. We aimed to avoid artificial inflation of recognition rates and, more critically, invalid reduction to a set of stereotypical and homogeneous expressions. Instead, in the *Quality Check*, the experimenter selection was limited to applying criteria relative to sound quality, the *Speaker Check* aimed at assaying the speaker's ability to produce nonverbal affective vocalizations, and the *Item Check* focused on perceived authenticity, and thus naturalness, independent of the accuracy of emotional content classification. In summary, the technical validation led us to discard 165 of the initial 1250 files (based on predefined criteria), and set the basis for the development of smaller, 'core' stimulus set.

Quality Check

In a first step, stimuli with a length between 400 ms to 2000 ms (*M* = 902.93 ms, *SD* = 374.43 ms) were selected, which constituted approximately 98% of the original files. Next, subsequent to the coarse chunking performed with Audacity, onset and offset of items were cut at a silence threshold using Matlab (R2017a). Some sounds (*n* = 19) with minor quality loss in the beginning or end of the file as well as sounds not entirely reaching the silence

Figure 1

Summary of the VIVAE Construction and Technical Validation



Note. (A) Pipeline describing the steps from data recording to validated materials. Recording files were chunked in individual expressions. The extracted nonspeech vocalizations were fed into a three-step validation procedure. A full set and a fully crossed subset, the core set, were obtained. (B) Results of the Speaker Check. Correct classification of emotion categories (top) and intensity levels (bottom) for each speaker affirmed validity. (C) Authenticity ratings for all Variably Intense Vocalizations of Affect and Emotion (VIVAE) stimuli. Average authenticity ratings and 95% confidence intervals for core set stimuli (purple) and remaining stimuli (blue), which together form the full set. See the online article for the color version of this figure.

threshold at 2000 ms, could be saved by applying a linear fade in/fade out ramp, which did not exceed 5% of the sound (n = 14) or 1% (n = 5). In addition, two expert judges, blind to the aim of the study, were asked to a) exclude sounds with verbal content and b) exclude sounds of poor audio quality (e.g., clippings, distortions, etc.) by both listening, and in specific cases, by visualizing the sound file (i.e., waveform and spectrogram) using Praat. On this basis, N = 1092 items passed the acoustic quality check, which on average resulted in 99 expressions per speaker, 182 expressions per emotion category, and 45 expressions per emotion × intensity level.

Speaker Check

Ten participants (mean age = 28 years, *SD* = 5.35, range = 20–39; 5 self-identified as women, 5 self-identified as men) evaluated speaker performance. Assuming an effect size (d) of .85 (a large effect for the classification of emotion: Cordaro et al., 2016; Juslin & Laukka, 2001), this sample size provides adequate power of 80% for classification tested against chance using one-sample t-tests (one-tailed), determined using G*Power (Faul et al., 2007). Participants were recruited through the Max-Planck-Institute for Empirical Aesthetics, Frankfurt. All participants reported having normal hearing, as well as no history of neurological or psychological illnesses. The experimental procedures were approved by the Ethics Council of the Max Planck Society. Participants provided informed consent before participating and received financial compensation.

Each participant was instructed to evaluate each of the 1092 vocalizations (11 Speakers × 6 Emotions × 4 Intensities) by categorizing emotion and classifying intensity. Stimulus presentation and response recording were performed using Presentation software (Version 20.0). The sound amplitude was calibrated to a maximum of 90.50 dB(A), with a range of 47.50 dB(A) resulting in 43 dB(A) for the peak amplitude in the most silent sound file.

Sounds were presented over DT 770 Pro Beyerdynamic headphones. On each trial, participants were asked to assign one out of six possible response options to the presented expression: the German emotion labels for anger (Ärger), fear (Angst), pain (Schmerz), achievement (Triumph), positive surprise (Positive Überraschung), and sexual pleasure (Sexuelle Lust), presented in random order across but fixed order within participant. Participants were instructed to pick the emotion label which best fitted the emotion expressed by the speaker. Next, a labeled 4-point Likert-scale was presented, and participants were asked to indicate how intensely they believed the speaker had experienced the emotional state, choosing between little (1–"niedrig"), moderate (2–"mäßig"), strong(3–"hoch"), and extreme (4–"extrem") emotion intensity. The next sound was played automatically. The order of tasks was counterbalanced across participants and breaks could be taken between blocks. In total, participants completed 14 blocks of 78 trials each. The order of stimuli was pseudorandomized across participants. One block lasted approximately 8.5 minutes; the averaged total session time, which included the experiment itself and a short questionnaire on sociodemographic information subsequent to the listening task, was 2.5 hr.

Item Check

Twenty participants (*M* = 25.40 years old, *SD* = 5.12 years, range = 19–36; 10 self-identified as women, 9 as men, and 1 as nonbinary) were recruited for the evaluation of authenticity. The same criteria applied as for the Speaker Check. No effect size measures were available for a priori power analysis; instead, sample size was based on previous research (Lima et al., 2013).

Participants were asked to evaluate the authenticity of each of the 1092 vocalizations on a seven-point Likert scale. They were instructed to judge perceived authenticity of the expressions of positive and negative affective states as spontaneous as possible on a rating scale ranging from little (1–"wenig") authentic to fully (7–"völlig") authentic. They were asked to indicate how much they thought the speaker had genuinely experienced an emotional state when producing the expression, regardless of their certainty of having identified the portrayed type of emotion or its emotional intensity, and regardless of how clearly or typical they thought the emotion had been expressed by the speaker. Importantly, participants were naïve to the origin of the stimuli, especially to the material being play-acted or natural, and they were not provided labels of the expressed affective state when listening to the vocalization. All other aspects of stimulus presentation and response recording were the same as in the Speaker Check, resulting in 14 blocks of 78 trials each with an average block duration of 6.5 minutes, leading to a total session time of approximately 2 hr.

Sound Analysis

Two types of analyses were performed to assess the expression of emotion and intensity in nonverbal vocalizations: (i) a descriptive evaluation of vocalization types and phonetic characteristics, and (ii) an assessment of the representation of emotion and intensity in the acoustic structure of vocalizations.

Phonetic and Vocalization Type Analysis

To assess if a certain vocalization type or vowel sound carries iconic meaning indicative of its context (beyond a vocalization's prosodic content), we catalogued these expressive properties with regard to the emotion and the intensity of their use. For each vocalization, the predominant vowel sound (e.g., "u," for [uːh], or "ə"[1], for [həhəhəhə]) was determined by two researchers who listened to all stimuli without being presented the associated affective labels. In the same procedure, and in line with the concept of different call types proposed by Anikin et al. (2018), perceptual vocalization types (e.g., moans, grunts, screams) were determined for the VIVAE core set.

Acoustic Feature Analysis

Low level acoustic descriptors were extracted using a batch-processing script in Praat (Version 6.1.14; Boersma & Weenink, 2020) and using the package soundgen in R (Anikin, 2019). For each vocalization, acoustic parameters related to pitch, loudness, and voice quality were measured, selected based on previous research on acoustic correlates of emotion and emotion intensity (Anikin, 2020a; Arnal et al., 2015; Juslin & Laukka, 2001; Laukka et al., 2010; Raine et al., 2019). Our analysis was focused on the following acoustic properties (extracted using Praat unless indicated otherwise):

F0 mean, max, SD, range (Hz): all derived from the pitch contours computed using Praat's autocorrelation method. Extracted pitch contours were systematically inspected and, if clearly erroneous (e.g., octave jumps, deterministic chaos), manually corrected. *M* fundamental frequency (F0 mean) is closely related to the auditory impression of voice pitch. F0 max denotes the maximum value of F0 in a vocalization. The standard deviation of the fundamental frequency (F0 *SD*) represents the pitch variability of a vocalization. F0 range is the difference between lowest and highest F0 in a vocalization.

F0 slope: mean absolute F0 slope in semitones.

Jitter: a measure of F0 perturbation; the average absolute difference between consecutive differences for consecutive intervals, divided by the average interval.

Int mean, max, SD, range (dB): mean, maximum, standard deviation, and range of the amplitude of a vocalization measured in dB, associated with the percept of loudness; vocal energy.

Shimmer: average absolute difference between consecutive differences for amplitudes of consecutive intervals.

F1, F2, & F3 mean: mean of formants 1–3.

F1, F2, & F3 bandwidth: median bandwidth of formants 1–3.

HF 500, HF 1000: the amount of high-frequency energy in the spectrum as the relative proportion of energy above versus below two cutoff frequencies, 500 Hz and 1000 Hz.

Voiced (%): portion of voiced opposed to unvoiced frames.

HNR (dB): harmonics-to-noise-ratio; average degree of periodicity.

COG: spectral center of gravity, or spectral centroid; the center of mass of a vocalization and a measure of timbral brightness.

Spectral slope (Soundgen): the slope of a linear regression fit to the spectrum.

[1] mid central vowel, typically denoted as "schwa".

Roughness mean *(Soundgen):* the percent of energy in the temporal "roughness range," that is, 30–150 Hz amplitude modulations. The modulation spectrum is obtained through a two-dimensional Fourier transform and captures spectro-temporal patterns in a vocalization.

For statistical analysis, pitch related variables (F0 mean, max, *SD*, range, and slope, as well as jitter and shimmer) were log-transformed; acoustic variables were normalized (*z*-transformed) per speaker (see Table S2 for descriptive statistics of each variable, and Figure S1 for intercorrelations between variables). We first performed a principal component analysis (PCA) to reduce the acoustic parameter set to a smaller number of uncorrelated factors. A PCA with no factor rotation generated six components with eigenvalues greater than 1 (Kaiser's criterion), accounting for 76% of the variance in the dataset. To improve the interpretability of the resulting principal components, the PCA was run again, retaining the first six components and using a varimax rotation. Acoustic differences between emotion categories and intensity levels were then examined using a discriminant function analysis. As homogeneity of variance could not be assumed for the components produced from the acoustic variables, we performed a leave-one-out quadratic discriminant analysis (QDA). Furthermore, we assessed how specific acoustic properties varied with increased emotional intensity. We report a series of Spearman correlation analyses between each of the 24 acoustic features and the expressed emotional intensity (Bonferroni adjusted alpha levels of .002.)

Results and Discussion

To ensure the quality and ecological validity of the VIVAE materials, the vocalizations collected in the construction phase were submitted to three validation steps (Figure 1A). All sounds which passed the *Quality Check* ($N = 1092$) were tested as stimulus materials in the subsequent validation steps, that is, the *Speaker* and *Item Checks*, following which a 'full' and a 'core' set are proposed.

Technical Validation

The *Speaker Check* served to assess speakers' ability to produce nonverbal affective vocalizations, using as criteria the classification accuracy of emotion category and emotional intensity. Overall, correct emotion classification across speakers was $M = 53.27\%$ ($SD = 5.40$), with a mean of $M = 54.21\%$ ($SD = 4.36$) for negative intended valence and $M = 52.35\%$ ($SD = 10.10$) for positive intended valence. Classification accuracy values per emotion were: $M = 45.16\%$ (*SD* across speakers = 21.22) for achievement, $M = 58.13\%$ ($SD = 14.70$) for anger, $M = 48.75\%$ ($SD = 15.80$) for fear, $M = 55.68\%$ ($SD = 7.02$) for pain, $M = 61.93\%$ ($SD = 14.74$) for pleasure, and $M = 48.26\%$ ($SD = 14.78$) for surprise. The interrater agreement on emotion classification was moderate (Fleiss Kappa $k = .47$). For each individual speaker, emotion classification was significantly greater than chance (16.67%), as confirmed by one-sample *t*-tests separately for each speaker ($ps < .001$, all significant after Bonferroni correction, Figure 1B). In addition, intended emotion intensity was successfully classified across speakers: The overall correct intensity classification rate was $M = 41.17\%$ ($SD = 4.63$, chance accuracy = 25%). Again, a series of one-sample *t*-tests confirmed that intensity classification was significantly greater than chance for each speaker (S10, $p = .01$, all other $ps < .001$, all significant after Bonferroni correction, Figure 1B). Note that, given the large sample size, parametric tests were performed despite variation in the speakers' distributions of emotion and intensity classification accuracy. Likewise, nonparametric sign tests on each speaker's data support higher than chance classification accuracy for emotion and intensity (intensity classification accuracy S02 and S10, $p = .01$, all others, $p < .001$).

Above-chance emotion and intensity classification attested that all speakers produced stimuli which listeners were able to classify in accordance with speaker intentions. These results are in line with previous research suggesting that emotion categories and emotional intensity are, generally speaking, inferred successfully from the human voice (Anikin & Persson, 2017; Bänziger et al., 2012; Juslin & Laukka, 2001; Lima et al., 2013; Livingstone & Russo, 2018). Classification accuracy was lower compared to other corpora of nonverbal vocalizations. First, classification accuracy may vary as a function of emotion intensity, and higher ambiguity has previously been described for instances of both weak and extreme emotions (Atias et al., 2019; Juslin & Laukka, 2001), represented in this corpus. (Note that a detailed analysis of the effects of intensity on perceptual evaluation are addressed in a separate study, Holz et al., 2021). Second, in contrast to corpora presupposing diagnostic emotion expression (Belin et al., 2008; Cordaro et al., 2016; Hawk et al., 2009; Lima et al., 2013; Maurage et al., 2007; Sauter, Eisner, Calder, & Scott, 2010; Schröder, 2003), individual expressions were not selected to best fit a characteristic pattern. Thus, vocalizations, despite consistent *overall* grouping, may vary in how reliably they convey a specific emotion or emotional intensity, resonating with the idea of variable and context-dependent emotion communication (Barrett, 2017a; Gendron et al., 2014).

Individual expressions were evaluated in the *Item Check* based on their perceived authenticity. The mean authenticity rating for stimuli was $M = 4.04$ ($SD = .66$). Average ratings per stimuli ranged from $M = 1.70$ for the lowest rated one to $M = 6.20$ for the highest rated expression, with more stimuli rated on the "authentic" than the "nonauthentic" end of the scale (Figure 1C). The tendency toward authenticity was supported by a significant one-sample *t*-test, confirming that the average authenticity rating for vocalizations was greater than four, the center of the Likert scale (t[1091] = 1.70, $p = .04$, $d = .05$); note, however, that the effect was very small. Additionally, a series of *t*-tests separate for each stimulus revealed that of all 1092 items, listeners rated $n = 204$ (18.68%) vocalizations as lower than four (i.e., "neutral"; two-sided, alpha defined at .05), while $n = 888$ vocalizations were either perceived as either neutral ($n = 650$, 59.52%) or rather authentic ($n = 238$, 21.79%). Interrater reliability for authenticity ratings was measured using a two-way random effects, consistency, average-measures intraclass correlation (Hallgren, 2012), using the irr package in R. Interrater reliability was high, ICC(C, 20) = .86, an ICC interpreted as excellent following Cicchetti (1994). Two participants reported having recognized verbal content in some of the stimuli (i.e., the word "no" and the interjection "woohoo"). We revised the material and discarded $n = 7$ files with linguistic content. Together, the results of the *Item Check* support that the majority of expressions was evaluated as authentic (Figure 1C). Given the high agreement on authenticity ratings, perceived naturalness constitutes a well-suited criterion to subsample the VIVAE to stimulus selections tailored to the needs of the research in which it is applied.

Data Records

The technical validation criteria were met by 1085 files. Each of the VIVAE files has a unique filename following the file naming convention: [Speaker_Emotion_Intensity_Item-ID.wav]. The filename consists of a 4-part identifier (e.g., S04_surprise_peak_10.wav) defining the stimulus characteristics:

-Speaker (S01 to S11).

-Emotion (achievement/triumph, anger, fear, pain, pleasure, and surprise).

-Emotional intensity (low, moderate, strong, peak).

-Item-ID (unique integer identifier within Speaker × Emotion × Intensity conditions)

The VIVAE is openly available from http://doi.org/10.5281/zenodo.4066235.

Full Set

The full set of the VIVAE comprises 1085 files, that is, all files that passed the three-step validation procedure.

Core Set

A core stimulus set was developed with the aim to provide a well-controlled (i.e., fully crossed) subsample of suitable size for various experimental paradigms. To this end, we applied a customized selection procedure based on perceived authenticity of individual items. For all cases in which there were more than two items for each combination of speaker and emotion and intensity ($n = 999$), the two with the highest rating in authenticity were chosen. Whenever more than two tokens were filtered by this procedure ($n = 41$), the two with the higher agreement (lower *SD*) were included. Through this procedure, a total of 480 stimuli, consisting of two exemplars each per speaker, emotion, emotion intensity combination, was selected as core set. Note that one speaker (S11), did not provide a sufficient number of stimuli per condition after the reported material revision due to noted linguistic content, so that their items ($n = 59$) were excluded from the core set. Comparison of the summary statistics (emotion classification rate, intensity classification rate, average authenticity rating) in one-sample *t*-tests before and after speaker exclusion revealed no significant differences ($ps > .05$)

We compared perceptual judgements for the selected core set with the remaining (nonselected) stimuli of the full set. Using separate independent 2-group Mann–Whitney U Tests for nonparametric data, we found no differences in emotion classification accuracy ($z = -1.81$, $p = .07$, $r = .06$) and intensity classification accuracy ($z = -.56$, $p = .57$, $r = .02$). Unsurprisingly, a significant difference in perceived authenticity ($z = -12.98$, $p < .001$, $r = .39$) was revealed. Vocalizations were perceived as quite authentic (t[479] = 11.46, $p < .001$, $d = .52$), overall, and the number of items rated significantly below four (i.e., scale center) could be reduced to $n = 41$ (8.54%), whereas 439 vocalizations were evaluated as neutral ($n = 267$, 55.63%) or authentic ($n = 172$, 35.83%), alluding to the criteria of our selection procedure (Figure 1C).

Sound Characteristics of the VIVAE

Vocalization Types

Vocalizations included a variety of perceptually distinct call types, such as cries, grunts, shrieks, or laughter. The distribution of different vocalization types across emotion categories and intensity levels is shown in Figures 2A and B. Most frequent vocalization types include moans ($n = 90$), screams ($n = 77$), and sighs ($n = 61$). Least frequent were growls ($n = 2$), expressed in the context of anger. For all other vocalization types, the number of emotion categories across which they occurred varied from at least two to all six (*Mdn* = 5). For intensities, though none of the vocalization types (except growls) was exclusive to a specific intensity level, a clear tendency is visible, such that more intense emotions (i.e., strong and peak) were more often expressed through screams or scream-like vocalizations, and less intense emotions (i.e., low and moderate) more through gasps, sighs, and moans.

Vowel Types

Perceived vowel sounds in vocalizations include /a/ ($n = 554$), /u/ ($n = 109$), /o/ ($n = 82$), /ə/ ($n = 77$), /ou/ ($n = 72$), /oa/ ($n = 17$), /au/ ($n = 9$), /e/ ($n = 9$), /ua/ ($n = 7$), /ui/ ($n = 5$), and /ai/ ($n = 5$). Their occurrence across different emotions and intensities is visualized in Figures 2C and D. Notably, and despite variation in the frequency of use associated with different emotions and intensities, no single vowel sound was indicative of the expressed emotion or intensity: Each vowel was associated with three to six emotion categories, and two to four intensity levels, respectively.

Low-Level Acoustic Feature Analysis

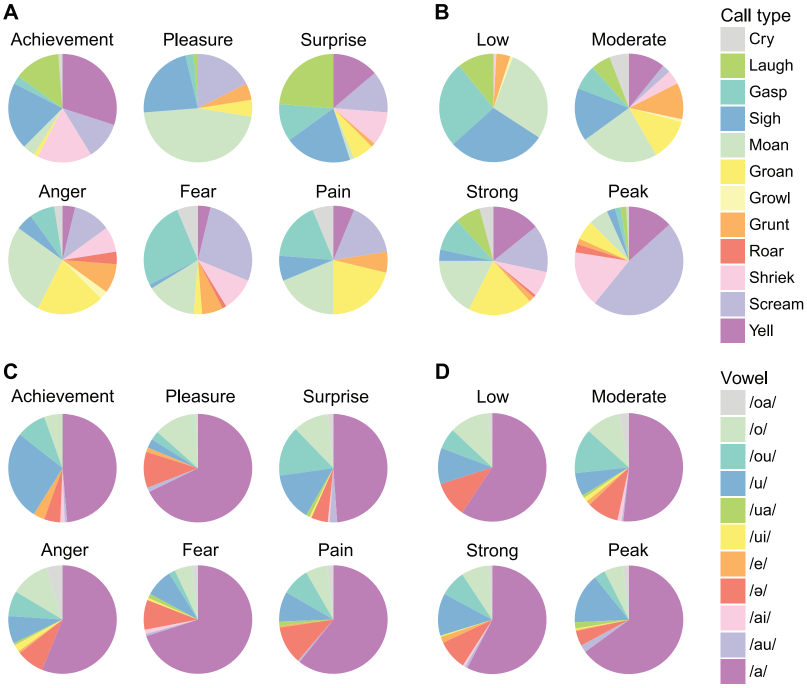
We assessed if expressed emotion categories and intensity levels differed with regards to their low-level acoustic structure. The PCA on the acoustic measures of amplitude, pitch, and spectral properties produced six components (see Table 1), which were used as independent variables in a discriminant function analysis. Acoustic variable loadings on each of the components are reported in Table S3.

Variable loadings indicated that the first component indexed *phonation frequency* and *frequency variability*: Greater Component 1 scores indicated higher pitch (F0 mean and F0 max) and greater pitch variability (F0 range and F0 *SD*). The second component had high loadings of COG, HF 500, HF 1000, and F1 mean. These features are related to *phonatory effort*, as greater high-frequency energy, produced by high subglottal pressure and strong vocal fold adduction, is linked to brighter and tense sounding voice quality (Juslin & Laukka, 2001; Patel et al., 2011). The third component included HNR, roughness, jitter, shimmer, and the F0 slope, reflecting *phonation perturbation and nonlinearity*, that is, magnitude and sharpness of frequency and amplitude changes; higher values are associated with a noisier and harsher sounding voice. The fourth component indexed the *amplitude variability* (Int *SD* and Int range). The fifth and sixth components were again related to *voice quality* and the perception of *articulation*, loading high on the spectral slope and F2 mean, and F3 bandwidth, respectively.

Figure 3 illustrates the distribution of vocalizations in the acoustic space mapped by the first three components. Despite substantial overlap, a clear differentiation is observable. More intense vocalizations are characterized by higher and more variable pitch (Component 1) as well as greater phonatory effort (Component 2)—while the picture is less clear for Component 3. We refer to the interactive version of the figure available at http://testing.musikpsychologie.de/VIVAE_emo/ for a detailed representation of the acoustic space, also for individual emotions and intensity levels. Results of the discriminant function classification, using the

Figure 2

Vowel Sounds and Call Types Across Emotions and Intensities



Note. (A) Distribution of vocalization types across emotions in the Variably Intense Vocalizations of Affect and Emotion (VIVAE) core set (N = 480 vocalizations). Top row, positive emotion; bottom row, negative emotion. (B). Proportions of vocalization types per intensity, from low to peak, core set. (C) Proportion of vowel sounds per emotion and (D) per intensity, for the VIVAE full set (N = 1085 vocalizations), respectively. See the online article for the color version of this figure.

component vectors as independent variables, are reported in Tables S4 and S5. The QDA indicated that emotions and intensities were acoustically distinguishable. The correct classification rate for emotions was M = 34.91% (chance = 16.67%), and for intensities M = 53.80% (chance = 25.00%). Together with the human evaluation data, these results support the representation of emotion and intensity in the expression and perception of the VIVAE materials. Importantly, they also reveal—to some extent—overlap and ambiguity in the classification of vocalizations.

Given the centrality of emotional intensity in VIVAE, we were interested in identifying individual acoustic features associated with varying levels of intensity. To this end, each acoustic feature was correlated with the expressed emotional intensity (see Figure 4). With increasing emotional intensity, vocalizations featured higher pitch (F0 mean and F0 max) and greater pitch variability (F0 range and F0 variability), also reflected in the intercorrelation between F0 features (Figure S1). The same was true for features of high frequency energy (COG, HF 500, and HF 1000), which, together with the spectral slope, are measures of voice quality and indicate a brighter sounding voice for more emotionally intense vocalizations. Moreover, amplitude related features (Int mean, Int max, Int *SD*, and Int range), linked to the subglottal air pressure and perceived loudness, increased with emotion intensity. This presumed increase of subglottal pressure was likewise reflected in a higher first formant frequency and narrower formant bandwidths at stronger emotional intensity, mechanistically linked to vocal

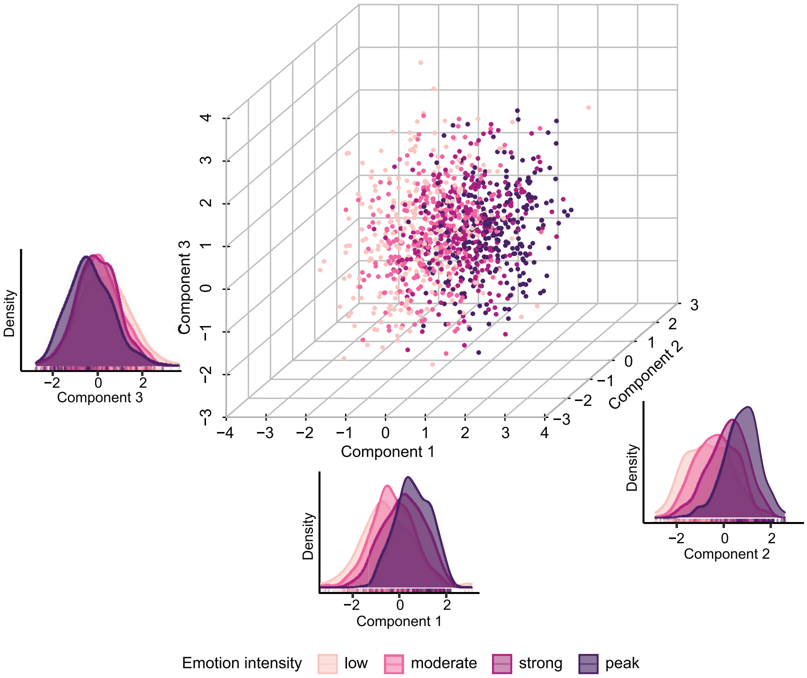
Table 1

Acoustic Dimensions Generated From Varimax-Rotated Principal Component Analysis

Component	Eigenvalue	Variance	Acoustic feature loading (> abs. 0.6)
1	4.49	19%	F0 max (.84), F0 mean (.68), F0 SD (.92), F0 range (.93)
2	4.13	17%	COG (.84), HF 500 (.80), HF 1,000 (.90), F1 mean (.70)
3	4.03	17%	jitter (.76), shimmer (.80), HNR (−.91), roughness (.69), F0 slope (.73)
4	2.17	9%	Int SD (.95), Int range (.90)
5	2.16	9%	spectral slope (.73), F2 mean (.65)
6	1.26	5%	F3 mean (.65), F3 bandwidth (−.67)

Note. Eigenvalues, percentage of explained variance, and highest loadings (>.6) are given for each component.

Figure 3
Emotion Intensity in Acoustic Space



Note. Projection of the Variably Intense Vocalizations of Affect and Emotion (VIVAE) full set onto three principal component analysis components in rotated space (grouped by intensity), along with density plots showing the differences between intensities—or lack thereof—on each dimension. The interactive version of this plot is available at http://testing.musikpsychologie.de/VIVAE_emo/. See the online article for the color version of this figure.

tract constriction and tensing—and, respectively, intercorrelations of these features as by the production process (Juslin & Laukka, 2001; Scherer, 1989). Lastly, more intense vocalizations were characterized by smaller F0 perturbation values (jitter, shimmer) and the tendency of greater harmonicity (HNR), while no effect of intensity on vocalizations' roughness was observable.

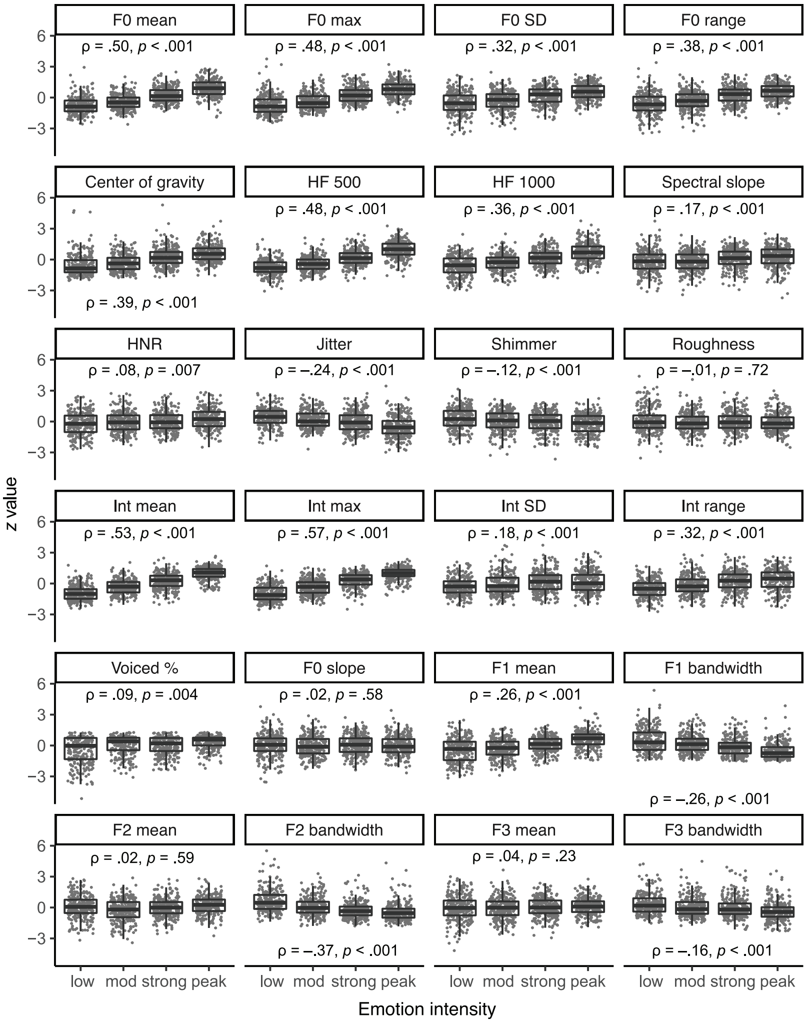
Comparing these results on the VIVAE materials with previous evidence on expressed emotional intensity in prosodic speech and nonspeech pain vocalizations, the convergence is high. Juslin and Laukka (2001) and Raine et al. (2019) also found an increase in pitch, pitch variability, voice intensity, and for speech, high frequency energy, and first formant frequency. For F0 irregularity and noisiness, the results are in line with the higher harmonicity reported for vocalizations of increasing pain intensity (Raine et al., 2019), but are opposite to the relation of jitter and emotion intensity in variably-intense speech (Juslin & Laukka, 2001). Moreover, a study by Anikin (2020a) found mixed effects of harmonicity, frequency irregularity, and modulation on the *perceived* emotional intensity of vocalizations. Possibly, these mixed results reflect the fact that multiple production mechanisms underly irregular, perturbated phonation, and chaos, more generally. One is the tension of the vocal fold adduction linked to physiological changes and vocal effort: Hypertension *and* hypotension can lead to irregular vocal fold vibration and nonlinear vocal phenomena (Anikin, 2020b; Fitch et al., 2002; Scherer, 1989). Furthermore, roughness, an overall measure of amplitude and frequency modulation, has previously been described for extremely intense vocalizations, such as screams (Arnal et al., 2015); as well as for vocalizations including unvoiced parts and a breathy voice quality (Wood, 2020), associated with lower emotional intensity (see Figures 3 and 4). A promising avenue for future work, for which the VIVAE materials are a suitable resource, is a more fine-grained analysis of nonlinear phenomena with regard to the perceived emotional intensity and the behavioral relevance of different types of 'noise' in such variably-intense vocalizations.

Conclusion

We present the construction, validation, and acoustic characterization of the Variably Intense Vocalizations of Affect and Emotion (VIVAE) Corpus, which provides material to investigate more thoroughly affect recognition, especially in light of the theoretically central but largely underspecified role of intensity. This extensive stimulus set of nonverbal vocalizations is the first corpus to systematically feature emotion intensity from one extreme to the other. The *full set*, comprising 1085 stimuli, features eleven speakers expressing three positive (achievement/triumph, sexual pleasure, surprise) and three negative (anger, fear, physical pain) affective states, each varying from low to peak emotion intensity. The smaller *core set* of 480 vocal expressions represents a fully crossed subsample (6 emotions × 4 intensities × 10 speakers × 2

Figure 4

Relation of Individual Acoustic Features and Expressed Emotion Intensity



Note. For each acoustic feature, *z*-transformed value range per intensity level and Spearman correlations for each of the acoustic cues and emotion intensity. Exact *p*-values are reported except for $p < .001$; Bonferroni adjusted alpha at .002. $N = 1080$ for F0 mean, F0 max, F0 SD, F0 range, HNR, jitter, and shimmer (all log transformed), and $N = 1085$ for all others.

items), selected based on judged authenticity. The expressed emotional intensity, like expressed emotion, is reflected in perceptual and acoustic properties of nonverbal vocalizations. We provide evidence that intensity shapes the expression of emotion: it influences the types of vocalizations used to communicate and it is represented a range of low-level acoustic features, such as signal amplitude and fundamental frequency.

These materials form the basis for a widely usable open-source database of affective vocalizations. The validation of the stimuli includes testing the reliability of the raters, evaluating the perceived authenticity of vocalizations, as well as assessing the communication of emotion and intensity between speakers and listeners. Two key features distinguish the VIVAE from existing corpora: (i) the comprehensive variation along the intensity dimension, and (ii) the design privileging naturalness over recognizability. Both considerations likely boosted expressive diversity via a broad spectrum of vocalizations and yield novel and promising materials for ongoing and future research.

The advocated design also limits possible applications of this corpus. For instance, this corpus is not ideal for studies requiring unequivocal stereotypical expressions. The dataset was not produced to depict perfectly recognizable prototypes; rather the strength of this corpus lies in its higher ecological validity and its possible contribution to better understand emotion communication in its naturalistic variability. Yet, compared to real-life expressions, it features many advantages of laboratory-produced material, such as satisfactory audio quality, well-controlled design, and coherent speaker intentions.

The use of variably intense expressions in emotion perception paradigms is a particularly interesting avenue for future work. Given the central theoretical yet by and large empirically underspecified role of intensity, this corpus serves as cornerstone for research on the effects of intensity on perceptual and acoustic properties of vocal emotion communication. The open-source stimulus material allows further investigation how the representation of emotion intensity—along with closely linked aspects like physiological arousal, perceptual salience, and biological relevance—shape the processing of human communication signals.

References

Anikin, A., & Lima, C. F. (2018). Perceptual and acoustic differences between authentic and acted nonverbal emotional vocalizations. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *71*(3), 622–641. https://doi.org/10.1080/17470218.2016.1270976

Anikin, A. (2019). Soundgen: An open-source tool for synthesizing nonverbal vocalizations. *Behavior Research Methods*, *51*(2), 778–792. https://doi.org/10.3758/s13428-018-1095-7

Anikin, A. (2020a). The link between auditory salience and emotion intensity. *Cognition and Emotion*, *34*(6), 1246–1259. https://doi.org/10.1080/02699931.2020.1736992

Anikin, A. (2020b). The perceptual effects of manipulating nonlinear phenomena in synthetic nonverbal vocalizations. *Bioacoustics*, *29*(2), 226–247. https://doi.org/10.1080/09524622.2019.1581839

Anikin, A., Bååth, R., & Persson, T. (2018). Human non-linguistic vocal repertoire: Call types and their meaning. *Journal of Nonverbal Behavior*, *42*(1), 53–80. https://doi.org/10.1007/s10919-017-0267-y

Anikin, A., & Persson, T. (2017). Nonlinguistic vocalizations from online amateur videos for emotion research: A validated corpus. *Behavior Research Methods*, *49*(2), 758–771. https://doi.org/10.3758/s13428-016-0736-y

Arnal, L. H., Flinker, A., Kleinschmidt, A., Giraud, A. L., & Poeppel, D. (2015). Human screams occupy a privileged niche in the communication soundscape. *Current Biology*, *25*(15), 2051–2056. https://doi.org/10.1016/j.cub.2015.06.043

Atias, D., Todorov, A., Liraz, S., Eidinger, A., Dror, I., Maymon, Y., & Aviezer, H. (2019). Loud and unclear: Intense real-life vocalizations during affective situations are perceptually ambiguous and contextually malleable. *Journal of Experimental Psychology: General*, *148*(10), 1842–1848. https://doi.org/10.1037/xge0000535

Aviezer, H., Ensenberg, N., & Hassin, R. R. (2017). The inherently contextualized nature of facial emotion perception. *Current Opinion in Psychology*, *17*, 47–54. https://doi.org/10.1016/j.copsyc.2017.06.006

Aviezer, H., Trope, Y., & Todorov, A. (2012). Body cues, not facial expressions, discriminate between intense positive and negative emotions. *Science*, *338*(6111), 1225–1229. https://doi.org/10.1126/science.1224313

Bachorowski, J.-A. (1999). Vocal expression and perception of emotion. *Current Directions in Psychological Science*, *8*(2), 53–57. https://doi.org/10.1111/1467-8721.00013

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, *70*(3), 614–636. https://doi.org/10.1037/0022-3514.70.3.614

Bänziger, T., Mortillaro, M., & Scherer, K. R. (2012). Introducing the Geneva Multimodal expression corpus for experimental research on emotion perception. *Emotion*, *12*(5), 1161–1179. https://doi.org/10.1037/a0025827

Barrett, L. F. (2017a). *How emotions are made: The secret life of the brain*. Houghton Mifflin Harcourt.

Barrett, L. F. (2017b). The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience*, *12*(1), 1–23. https://doi.org/10.1093/scan/nsx060

Belin, P., & Zatorre, R. J. (2015). Neurobiology: Sounding the Alarm. *Current Biology*, *25*(18), R805–R806. https://doi.org/10.1016/j.cub.2015.07.027

Belin, P., Fillion-Bilodeau, S., & Gosselin, F. (2008). The Montreal Affective Voices: A validated set of nonverbal affect bursts for research on auditory affective processing. *Behavior Research Methods*, *40*(2), 531–539. https://doi.org/10.3758/BRM.40.2.531

Bernat, E., Patrick, C. J., Benning, S. D., & Tellegen, A. (2006). Effects of picture content and intensity on affective physiological response. *Psychophysiology*, *43*(1), 93–103. https://doi.org/10.1111/j.1469-8986.2006.00380.x

Blood, A. J., & Zatorre, R. J. (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(20), 11818–11823. https://doi.org/10.1073/pnas.191355898

Boersma, P., & Weenink, D. (2020). Praat: Doing phonetics by computer (Version 6.1.14) [Computer program]. http://www.praat.org/

Bonnet, L., Comte, A., Tatu, L., Millot, J. L., Moulin, T., & Medeiros de Bustos, E. (2015). The role of the amygdala in the perception of positive emotions: An "intensity detector". *Frontiers in Behavioral Neuroscience*, *9*, 178. https://doi.org/10.3389/fnbeh.2015.00178

Bryant, G. A., & Aktipis, C. A. (2014). The animal nature of spontaneous human laughter. *Evolution and Human Behavior*, *35*(4), 327–335. https://doi.org/10.1016/j.evolhumbehav.2014.03.003

Bryant, G. A., Fessler, D. M. T., Fusaroli, R., Clint, E., Amir, D., Chávez, B., Denton, K. K., Díaz, C., Duran, L. T., Fanćovićová, J., Fux, M., Ginting, E. F., Hasan, Y., Hu, A., Kamble, S. V., Kameda, T., Kuroda, K., Li, N. P., Luberti, F. R., . . . Zhou, Y. (2018). The perception of spontaneous and volitional laughter across 21 societies. *Psychological Science*, *29*(9), 1515–1525. https://doi.org/10.1177/0956797618778235

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*(4), 284–290. https://doi.org/10.1037/1040-3590.6.4.284

Cordaro, D. T., Keltner, D., Tshering, S., Wangchuk, D., & Flynn, L. M. (2016). The voice conveys emotion in ten globalized cultures and one remote village in Bhutan. *Emotion*, *16*(1), 117–128. https://doi.org/10.1037/emo0000100

Ekkekakis, P. (2013). *The measurement of affect, mood, and emotion: a guide for health-behavioral research*. Cambridge University Press. https://doi.org/10.1017/CBO9780511820724

Ekman, P. (1984). Expression and the nature of emotion. In K. R. Scherer & P. Ekman (Eds.), *Approaches to emotion* (pp. 319–344). Erlbaum.

Ekman, P., & Cordaro, D. (2011). What is meant by calling emotions basic. *Emotion Review*, *3*(4), 364–370. https://doi.org/10.1177/1754073911410740

Ethofer, T., Anders, S., Wiethoff, S., Erb, M., Herbert, C., Saur, R., Grodd, W., & Wildgruber, D. (2006). Effects of prosodic emotional intensity on activation of associative auditory cortex. *Neuroreport*, *17*(3), 249–253. https://doi.org/10.1097/01.wnr.0000199466.32036.5d

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Fecteau, S., Belin, P., Joanette, Y., & Armony, J. L. (2007). Amygdala responses to nonlinguistic emotional vocalizations. *NeuroImage*, *36*(2), 480–487. https://doi.org/10.1016/j.neuroimage.2007.02.043

Fitch, W. T., Neubauer, J., & Herzel, H. (2002). Calls out of chaos: The adaptive significance of nonlinear phenomena in mammalian vocal production. *Animal Behaviour*, *63*(3), 407–418. https://doi.org/10.1006/anbe.2001.1912

Fridlund, A. J. (2017). The behavioral ecology view of facial displays, 25 years later. In J. M. Fernández-Dols & J. A. Russell (Eds.), *The science of facial expression* (pp. 77–92). Oxford University Press.

Frijda, N. H., Ortony, A., Sonnemans, J., & Clore, G. L. (1992). The complexity of intensity: Issues concerning the structure of emotion intensity. In M. S. Clark (Ed.), *Review of personality and social psychology: Vol. 13. Emotion* (pp. 60–89). Sage.

Frühholz, S., Trost, W., & Grandjean, D. (2014). The role of the medial temporal limbic system in processing emotions in voice and music. *Progress in Neurobiology*, *123*, 1–17. https://doi.org/10.1016/j.pneurobio.2014.09.003

Gendron, M., Crivelli, C., & Barrett, L. F. (2018). Universality reconsidered: Diversity in making meaning of facial expressions. *Current Directions in Psychological Science*, *27*(4), 211–219. https://doi.org/10.1177/0963721417746794

Gendron, M., Roberson, D., van der Vyver, J. M., & Barrett, L. F. (2014). Cultural relativity in perceiving emotion from vocalizations. *Psychological Science*, *25*(4), 911–920. https://doi.org/10.1177/0956797613517239

Grossman, R. B., & Tager-Flusberg, H. (2012). Quality matters! Differences between expressive and receptive non-verbal communication skills in adolescents with ASD. *Research in Autism Spectrum Disorders*, *6*(3), 1150–1155. https://doi.org/10.1016/j.rasd.2012.03.006

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, *8*(1), 23–34. https://doi.org/10.20982/tqmp.08.1.p023

Hawk, S. T., van Kleef, G. A., Fischer, A. H., & van der Schalk, J. (2009). Worth a thousand words": Absolute and relative decoding of nonlinguistic affect vocalizations. *Emotion*, *9*(3), 293–305. https://doi.org/10.1037/a0015178

Hess, U., Blairy, S., & Kleck, R. E. (1997). The intensity of emotional facial expressions and decoding accuracy. *Journal of Nonverbal Behavior*, *21*(4), 241–257. https://doi.org/10.1023/A:1024952730333

Holz, N., Larrouy-Maestri, P., & Poeppel, D. (2020). *The variably intense Vocalizations of Affect and Emotion Corpus (VIVAE)* [Data set]. Zenodo. https://doi.org/10.5281/zenodo.4066234

Holz, N., Larrouy-Maestri, P., & Poeppel, D. (2021). The paradoxical role of emotional intensity in the perception of vocal affect. *Scientific Reports*, *11*(1), 9663. https://doi.org/10.1038/s41598-021-88431-0

Johnson-Laird, P. N., & Oatley, K. (1989). The language of emotions: An analysis of a semantic field. *Cognition and Emotion*, *3*(2), 81–123. https://doi.org/10.1080/02699938908408075

Jürgens, R., Grass, A., Drolet, M., & Fischer, J. (2015). Effect of acting experience on emotion expression and recognition in voice: Non-actors provide better stimuli than expected. *Journal of Nonverbal Behavior*, *39*(3), 195–214. https://doi.org/10.1007/s10919-015-0209-5

Jürgens, R., Hammerschmidt, K., & Fischer, J. (2011). Authentic and play-acted vocal emotion expressions reveal acoustic differences. *Frontiers in Psychology*, *2*, 180. https://doi.org/10.3389/fpsyg.2011.00180

Juslin, P. N. (2013). Vocal affect expression: Problems and promises. In E. Altenmüller, E. Zimmermann, & S. Schmidt (Eds.), *Evolution of emotional communication* (pp. 252–273). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199583560.003.0016

Juslin, P. N., & Laukka, P. (2001). Impact of intended emotion intensity on cue utilization and decoding accuracy in vocal expression of emotion. *Emotion*, *1*(4), 381–412. https://doi.org/10.1037/1528-3542.1.4.381

Kamiloğlu, R. G., Fischer, A. H., & Sauter, D. A. (2020). Good vibrations: A review of vocal expressions of positive emotions. *Psychonomic Bulletin & Review*, *27*(2), 237–265. https://doi.org/10.3758/s13423-019-01701-x

Kragel, P. A., Koban, L., Barrett, L. F., & Wager, T. D. (2018). Representation, pattern information, and brain signatures: From neurons to neuroimaging. *Neuron*, *99*(2), 257–273. https://doi.org/10.1016/j.neuron.2018.06.009

Krahmer, E. J., & Swerts, M. G. J. (2008). On the role of acting skills for the collection of simulated emotional speech. *Proceedings of the international conference on spoken language processing (Interspeech 2008)* (pp. 261–264). ISCA.

Larrouy-Maestri, P., & Morsomme, D. (2014). The effects of stress on singing voice accuracy. *Journal of Voice*, *28*(1), 52–58. https://doi.org/10.1016/j.jvoice.2013.07.008

Laukka, P., Audibert, N., & Aubergé, V. (2012). Exploring the determinants of the graded structure of vocal emotion expressions. *Cognition and Emotion*, *26*(4), 710–719. https://doi.org/10.1080/02699931.2011.602047

Laukka, P., Elfenbein, H. A., Chui, W., Thingujam, N. S., Iraki, F. K., Rockstuhl, T., & Althoff, J. (2010). Presenting the VENEC corpus: Development of a cross-cultural corpus of vocal emotion expressions and a novel method of annotating emotion appraisals. In L. Devillers, B. Schuller, R. Cowie, E. Douglas-Cowie, & A. Batliner (Eds.), *Proceedings of the LREC 2010 workshop on corpora for research on emotion and affect* (pp. 53–57). European Language Resources Association.

Laukka, P., Elfenbein, H. A., Söder, N., Nordström, H., Althoff, J., Chui, W., Iraki, F. K., Rockstuhl, T., & Thingujam, N. S. (2013). Cross-cultural decoding of positive and negative non-linguistic emotion vocalizations. *Frontiers in Psychology*, *4*, 353. https://doi.org/10.3389/fpsyg.2013.00353

LeDoux, J. E., & Brown, R. (2017). A higher-order theory of emotional consciousness. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(10), E2016–E2025. https://doi.org/10.1073/pnas.1619316114

LeDoux, J., Phelps, L., & Alberini, C. (2016). What we talk about when we talk about emotions. *Cell*, *167*(6), 1443–1445. https://doi.org/10.1016/j.cell.2016.11.029

Lima, C. F., Castro, S. L., & Scott, S. K. (2013). When voices get emotional: A corpus of nonverbal vocalizations for research on emotion processing. *Behavior Research Methods*, *45*(4), 1234–1245. https://doi.org/10.3758/s13428-013-0324-3

Livingstone, S. R., & Russo, F. A. (2018). The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English. *PLoS ONE*, *13*(5), e0196391. https://doi.org/10.1371/journal.pone.0196391

Maurage, P., Joassin, F., Philippot, P., & Campanella, S. (2007). A validated battery of vocal emotional expressions. *Neuropsychological Trends*, *2*(1), 63–74.

Ortony, A., & Turner, T. J. (1990). What's basic about basic emotions? *Psychological Review*, *97*(3), 315–331. https://doi.org/10.1037/0033-295X.97.3.315

Patel, S., Scherer, K. R., Björkner, E., & Sundberg, J. (2011). Mapping emotions into acoustic space: The role of voice production. *Biological Psychology*, *87*(1), 93–98. https://doi.org/10.1016/j.biopsycho.2011.02.010

Pell, M. D., Rothermich, K., Liu, P., Paulmann, S., Sethi, S., & Rigoulot, S. (2015). Preferential decoding of emotion from human non-linguistic vocalizations versus speech prosody. *Biological Psychology*, *111*, 14–25. https://doi.org/10.1016/j.biopsycho.2015.08.008

Raine, J., Pisanski, K., Simner, J., & Reby, D. (2019). Vocal communication of simulated pain. *Bioacoustics*, *28*(5), 404–426. https://doi.org/10.1080/09524622.2018.1463295

Rendall, D. (2003). Acoustic correlates of caller identity and affect intensity in the vowel-like grunt vocalizations of baboons. *The Journal of the Acoustical Society of America*, *113*(6), 3390–3402. https://doi.org/10.1121/1.1568942

Russell, J. A., & Barrett, L. F. (1999). Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *Journal of Personality and Social Psychology*, *76*(5), 805–819. https://doi.org/10.1037/0022-3514.76.5.805

Rutter, L. A., Dodell-Feder, D., Vahia, I. V., Forester, B. P., Ressler, K. J., Wilmer, J. B., & Germine, L. (2019). Emotion sensitivity across the lifespan: Mapping clinical risk periods to sensitivity to facial emotion intensity. *Journal of Experimental Psychology: General*, *148*(11), 1993–2005. https://doi.org/10.1037/xge0000559

Sauter, D. A., & Scott, S. K. (2007). More than one kind of happiness: Can we recognize vocal expressions of different positive states? *Motivation and Emotion*, *31*(3), 192–199. https://doi.org/10.1007/s11031-007-9065-x

Sauter, D. A., Eisner, F., Calder, A. J., & Scott, S. K. (2010). Perceptual cues in nonverbal vocal expressions of emotion. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *63*(11), 2251–2272. https://doi.org/10.1080/17470211003721642

Sauter, D. A., Eisner, F., Ekman, P., & Scott, S. K. (2010). Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(6), 2408–2412. https://doi.org/10.1073/pnas.0908239106

Scherer, K. R. (1989). Vocal correlates of emotion. In H. Wagner & A. Manstead (Eds.), *Handbook of psycho-physiology: Emotion and social behavior* (pp. 165–197). Wiley.

Scherer, K. R. (1994). Affect bursts. In S. H. M. van Goozen, N. E. Vande Poll, & J. A. Sergeant (Eds.), *Emotions: Essays on emotion theory* (pp. 161–196). Erlbaum.

Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication*, *40*(1-2), 227–256. https://doi.org/10.1016/S0167-6393(02)00084-5

Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Sciences Information*, *44*(4), 695–729. https://doi.org/10.1177/0539018405058216

Scherer, K. R., & Bänziger, T. (2010). On the use of actor portrayals in research on emotional expression. In K. R. Scherer, T. Bänziger, & E. B. Roesch (Eds.), *Blueprint for affective computing: A sourcebook* (pp. 166–178). Oxford University Press.

Schröder, M. (2003). Experimental study of affect bursts. *Speech Communication*, *40*(1-2), 99–116. https://doi.org/10.1016/S0167-6393(02)00078-X

Scott, S. K., Sauter, D., & McGettigan, C. (2010). Brain mechanisms for processing perceived emotional vocalizations in humans. In S. M. Brudzynski (Ed.), *Handbook of behavioral neuroscience* (Vol. *19*, pp. 187–197). Elsevier; https://doi.org/10.1016/B978-0-12-374593-4.00019-X

Simon-Thomas, E. R., Keltner, D. J., Sauter, D., Sinicropi-Yao, L., & Abramson, A. (2009). The voice conveys specific emotions: Evidence from vocal burst displays. *Emotion*, *9*(6), 838–846. https://doi.org/10.1037/a0017810

Spackman, M. P., Brown, B. L., & Otto, S. (2009). Do emotions have distinct vocal profiles? A study of idiographic patterns of expression. *Cognition and Emotion*, *23*(8), 1565–1588. https://doi.org/10.1080/02699930802536268

Wager, T. D., Kang, J., Johnson, T. D., Nichols, T. E., Satpute, A. B., & Barrett, L. F. (2015). A Bayesian model of category-specific emotional brain responses. *PLoS Computational Biology*, *11*(4), Article e1004066. Advance online publication. https://doi.org/10.1371/journal.pcbi.1004066

Wang, S., Yu, R., Tyszka, J. M., Zhen, S., Kovach, C., Sun, S., Huang, Y., Hurlemann, R., Ross, I. B., Chung, J. M., Mamelak, A. N., Adolphs, R., Rutishauser, U. (2017). The human amygdala parametrically encodes the intensity of specific facial emotions and their categorical ambiguity. *Nature Communications*, *8*(1), 14821. https://doi.org/10.1038/ncomms14821

Wharton, T. (2003). Natural pragmatics and natural codes. *Mind & Language*, *18*(5), 447–477. https://doi.org/10.1111/1468-0017.00237

Wingenbach, T. S., Ashwin, C., & Brosnan, M. (2016). Validation of the Amsterdam Dynamic Facial Expression Set—Bath Intensity Variations (ADFES-BIV): A set of videos expressing low, ontermediate, and high intensity emotions. *PLoS ONE*, *11*(1), Article e0147112. https://doi.org/10.1371/journal.pone.0147112

Wood, A. (2020). Social context influences the acoustic properties of laughter. *Affective Science*, *1*(4), 247–256. https://doi.org/10.1007/s42761-020-00022-w

Received December 30, 2020
Revision received July 21, 2021
Accepted September 8, 2021 ■