Mapping 24 Emotions Conveyed by Brief Human Vocalization

Alan S. Cowen[1,*], **Hillary Anger Elfenbein**[#2], **Petri Laukka**[#3], **Dacher Keltner**[1]

Hillary Anger Elfenbein: helfenbein@wustl.edu; Petri Laukka: petri.laukka@psychology.su.se; Dacher Keltner: keltner@berkeley.edu

[1]Department of Psychology, University of California, Berkeley, 5315 Tolman Hall, Berkeley, CA 94720.

[2]Olin School of Business, Washington University, 1 Brookings Drive, Saint Louis, MO 63130.

[3]Department of Psychology, Stockholm University, 10691 Stockholm, Sweden

[#] These authors contributed equally to this work.

Abstract

Emotional vocalizations are central to human social life. Recent studies have documented that people recognize at least 13 emotions in brief vocalizations. This capacity emerges early in development, is preserved in some form across cultures, and informs how people respond emotionally to music. What is poorly understood is how emotion recognition from vocalization is structured within what we call a semantic space, the study of which addresses questions critical to the field: How many distinct kinds of emotions can be expressed? Do expressions convey emotion categories or affective appraisals (e.g., valence, arousal)? Is the recognition of emotion expressions discrete or continuous? Guided by a new theoretical approach to emotion taxonomies, we apply large-scale data collection and analysis techniques to judgments of 2032 emotional vocal bursts produced in laboratory settings (Study 1) and 48 found in the real world (Study 2) by U.S. English speakers (n = 1105). We find that vocal bursts convey at least 24 distinct kinds of emotion. Emotion categories (sympathy, awe) more so than affective appraisals (including valence and arousal) organize emotion recognition. In contrast to discrete emotion theories, the emotion categories conveyed by vocal bursts are bridged by smooth gradients with continuously varying meaning. We visualize the complex, high-dimensional space of emotion conveyed by brief human vocalization within an online interactive map: https://s3-us-west-1.amazonaws.com/vocs/map.html.

Keywords

emotion; voice; affect; computational methods; semantic space

---

It is something of an anatomical wonder how humans communicate with the voice: the contraction of muscles surrounding the diaphragm produces bursts of air particles that are

transformed into sound through vibrations of the vocal folds, and leave the mouth, depending on the position of the jaw, the tongue, and other implements of vocal control (Titze & Martin, 1998), in the form of words, laughter, playful intonation, crying, sarcastic tones, sighs, song, triumphant hollers, growls, or motherese. In essential ways, the voice makes humans human.

Recent studies are finding the human voice to be an extraordinarily rich and ubiquitous medium of the communication of emotion (Cordaro, Keltner, Tshering, Wangchuk, & Flynn, 2016; Juslin & Laukka, 2003; Kraus, 2017; Laukka et al., 2016; Provine & Fischer, 1989; Vidrascu & Devillers, 2005). In the present investigation, we apply a new theoretical model to address the following questions: How many emotions can we communicate with the voice? What drives vocal emotion recognition, emotion categories or more general affect dimensions (valence, arousal, etc.)? And is vocal emotion discrete or does it convey gradients of meaning?

The Richness of the Vocal Communication of Emotion.

Humans communicate emotion through two different kinds of vocalization (Keltner, Tracy, Sauter, Cordaro, & McNeil, 2016; Scherer, 1986). One is prosody—the non-lexical patterns of tune, rhythm, and timbre in speech. Prosody interacts with words to convey feelings and attitudes, including dispositions felt toward objects and ideas described in speech (Mitchell & Ross, 2013; Scherer & Bänziger, 2004).

In the study of emotional prosody, participants are often recorded communicating different emotions while delivering sentences with neutral content (e.g., "let me tell you something)." These recordings are then matched by new listeners to emotion words, definitions, or situations. An early review of 60 studies of this kind found that hearers can judge five different emotions in the prosody that accompanies speech—anger, fear, happiness, sadness, and tenderness—with accuracy rates approaching 70% (Juslin & Laukka, 2003; Scherer, Johnstone, & Klasmeyer, 2003). In a more recent study, listeners from four countries identified nine emotions with above chance levels of recognition accuracy (Laukka et al., 2016).

A second way that humans communicate emotion in the voice is with vocal bursts, brief non-linguistic sounds that occur in between speech incidents or in the absence of speech (Hawk et al., 2009; Scherer & Wallbott, 1994). Examples include cries, sighs, laughs, shrieks, growls, hollers, roars, oohs, and ahhs (Banse & Scherer, 1996; Cordaro et al., 2016). Vocal bursts are thought to predate language, and have precursors in mammals: for example, primates emit vocalizations specific to predators, food, affiliation, care, sex, and aggression (Snowdon, 2002).

As in the literature on emotional prosody, recent studies have sought to document the range of emotions communicated by vocal bursts (for summary, see Cordaro et al., 2016). In a paradigm typical in this endeavor, participants are given definitions (e.g., "awe is the feeling of being in the presence of something vast that you don't immediately understand") and are asked to communicate that emotion with a brief sound that contains no words (Simon-

Thomas, Keltner, Sauter, Sinicropi-Yao, & Abramson, 2009). Hearers are then presented with stories of emotion antecedents (e.g., "the person feels awestruck at viewing a large waterfall") and asked to choose from 3–4 vocal bursts the one that best matches the content of the story. Currently, it appears that 13 emotions can be identified from vocal bursts at rates substantially above chance (Cordaro et al., 2016; Laukka et al., 2013). The capacity to recognize vocal bursts emerges by two years of age (Wu, Muentener, & Schulz, 2017) and has been observed in more than 14 cultures, including two remote cultures with minimal Western influence (Cordaro et al., 2016; Sauter, Eisner, Ekman, & Scott, 2010; but see Gendron, Roberson, van der Vyver, & Barrett, 2014).

Vocal bursts are more than just fleeting ways we communicate emotion: they structure our social interactions (Keltner & Kring, 1998; Van Kleef, 2010). They convey information about features of the environment, which orients hearers' actions. For example, toddlers are four times less likely to play with a toy when a parent emits a disgust-like vocal burst—*yuch*—than a positive one (Hertenstein & Campos, 2004). They regulate relationships. From laughs, adults can infer a person's rank within a social hierarchy (Oveis, Spectre, Smith, Liu, & Keltner, 2013), and, from shared laughs, whether two individuals are friends or strangers (Smoski & Bachorowski, 2003; Bryant et al., 2016). Finally, vocal bursts evoke specific brain responses (Frühholz, Trost, & Kotz, 2016; Scott, Lavan, Chen, & McGettigan, 2014). For example, screams comprise powerfully evocative acoustic signals that selectively activate the amygdala (Arnal et al., 2015). Within 50ms, an infant's cry or laugh triggers activation in a parent's brain region—the periaqueductal grey—that promotes caring behavior (Parsons et al., 2014).

The Semantic Space of Emotion.

What is less well understood is how the recognition of emotion from vocal bursts is structured. How do people infer meaning from brief vocalizations? As emotions unfold, people rely on emotion knowledge—hundreds and even thousands of concepts, metaphors, phrases, and sayings (Russell, 1991)—to describe the emotion-related response, be it a subjective experience, a physical sensation, or, the focus here, emotion-related expressive behavior. The meaning ascribed to any emotion-related response can be mapped to what we have called a *semantic space* (Cowen & Keltner, 2017, 2018), a multidimensional space that represents all responses within a modality (e.g., experience, expression). Semantic spaces of emotion are captured in analyses of judgments of the emotion-related response (see Fig. 1A).

A semantic spaces are defined by three features. The first is its *dimensionality*—the number of distinct varieties of emotion that people represent within a response modality. In terms of the present investigation, when we perceive various emotional vocalizations, how many kinds of emotion do we recognize within this response modality?

A second property of semantic spaces is what we call *conceptualization*: how do emotion categories (e.g., "sympathy") and domain-general affective appraisals such as valence, arousal, and other themes (e.g., control, unexpectedness) detailed in appraisal and componential theories of emotion (Lazarus, 1991; Scherer, 2009; Smith & Ellsworth, 1985)

capture the individual's representation of an emotion-related response (Shuman, Clark-Polner, Meuleman, Sander, & Scherer, 2017)? When a parent hears a child cry, or a friend laugh, or people shouting in a bar in response to a buzzer beating shot, are emotion categories necessary to capture the meaning the hearer recognizes in the vocalization, or is the meaning captured by broader affective appraisals? Studies have yet to formally address this question, though it is central to contrasting positions within several theories of emotion (Barrett, 2006; Ekman & Cordaro, 2011; Russell, 2003; Scherer, 2009). Here we use statistical models to compare how emotion categories and more general affective appraisals capture the meaning people ascribe to vocal bursts.

Finally, semantic spaces of emotion are defined by how states are distributed within the space. Do the emotions people recognize in the human voice (or in facial expression, song, etc.) form clusters? What is the nature of the boundaries within that space? Are the categories really discrete, with sharp boundaries between them? Or not (Barrett, 2006)?

In a first study applying large-scale inference methods to derive a semantic space of emotion, we focused on reports of subjective emotional experience (Cowen & Keltner, 2017). Guided by recent empirical advances and relevant theory, we gathered participants' ratings (forced choice of over 30 emotions, free response, and 13 affective appraisals) of their self-reported experiences in response to 2185 evocative short videos "scraped" from the internet. New quantitative techniques allowed us to map a semantic space of subjective emotional experience, as visualized in Fig. 1B. That study yielded three findings about the semantic space of emotional experience. First, videos reliably (across participants) elicit at least 27 distinct varieties of reported emotional experience; emotional experience is greater in variety—that is, dimensionality—than previously theorized (Keltner & Lerner, 2010; Shiota et al., 2017). With respect to the conceptualization of emotion, categorical labels such as "amusement," "fear," and "desire" captured, but could not be explained by, reports of affective appraisals (e.g. valence, arousal, agency, certainty, dominance). Thus, as people represent their experiences with language, emotion categories cannot be reduced to a few broad appraisals such as valence and arousal. Finally, with respect to boundaries between emotion categories, we found little evidence of discreteness, but rather continuous gradients linking one category of experience to another.

Here, we extend our theorizing about semantic spaces to study the recognition of emotion from vocal bursts. This extension to emotion recognition is justified for empirical and theoretical reasons. Most notably, self-reports of subjective experience and emotional expression are only moderately correlated (for reviews, see Fernández-Dols & Crivelli, 2013; Matsumoto, Keltner, Shiota, O'Sullivan, & Frank, 2008). This suggests that semantic spaces of emotional experience and emotion recognition may have different properties, perhaps differing, for example, in their conceptualization or distribution. More generally, examining a semantic space of emotional expression offers new answers to central questions within emotion science: How many emotions have distinct signals? What drives the recognition of emotion, categories or affective appraisals? What is the nature of the boundaries between those categories?

The Present Investigation.

The investigation of semantic spaces of emotion recognition requires methodological departures from how emotion recognition has been studied in the past. To capture the dimensionality of emotion recognition, it is critical to study as wide an array of expressive signals as possible. One way to achieve this is by sampling structural variations in expression, for instance, by attempting to reconstruct variations in facial expression using an artificial 3D model (Jack, Sun, Delis, Garrod, & Schyns, 2016). Another approach is to sample naturalistic expressive behaviors of as many emotion concepts as possible (Cowen & Keltner, 2017). Early studies of emotional vocalization focused primarily on emotions traditionally studied in facial expression—anger, sadness, fear, disgust, and surprise (Juslin & Laukka, 2003). More recent studies have expanded in focus to upwards of 16 emotion concepts (Cordaro et al., 2016; Laukka et al., 2016). Here we build upon these discoveries to consider how a wider array of emotion concepts—30 categories and 13 appraisals—may be distinguished in vocal bursts.

To capture the distribution of recognized emotions, it is critical to account for nuanced variation in expression, anticipated in early claims about expressive behavior (Ekman, 1993) but rarely studied. Most studies have focused on 1 to 2 prototypical vocal bursts for each emotion category (Cordaro et al., 2016; Gendron et al., 2014; Simon-Thomas et al., 2009). This focus on prototypical expressions fails to capture how vocal bursts vary within an emotion category, and can yield erroneous claims about the nature of the boundaries between categories (Barrett, 2006).

With respect to conceptualization, it is critical to have independent samples of hearers rate the stimuli (vocal bursts) in terms of emotion categories and broader affective appraisals such as valence, arousal, certainty, and dominance. A more typical approach has been to match vocal bursts to discrete emotions, using words or brief stories depicting antecedents (Cordaro et al., 2016; Simon-Thomas et al., 2009). More recently, investigators have been gathering ratings not only of emotion categories, but also inferred appraisals and intentions (Nordström, Laukka, Thingujam, Schubert, & Elfenbein, 2017; Shuman et al., 2017). By combining such data with statistical approaches developed in the study of emotional experience (Cowen & Keltner, 2017), we can compare how emotion categories and affect appraisals shape emotion recognition.

In sum, to characterize the semantic space of emotion recognition of vocal bursts, what is required is the study of a wide array of emotions, many examples of each emotion, and observer reports of both emotion categories and affective appraisals. Guided by these considerations, we present two studies that examine the semantic space of emotion recognition of vocal bursts. In the first, participants rated the largest array of vocal bursts studied to date—over 2000—for the emotion categories they conveyed, for their affective appraisals, culled from dimensional and componential theories of emotion (Lazarus, 1991; Scherer, 2009; Smith & Ellsworth, 1985), and in a free response format. Building upon the discoveries of Study 1, and given concerns about the nature of vocal bursts produced in laboratory settings, in Study 2 we gathered ratings of vocal bursts culled from YouTube videos of people in naturalistic contexts. Guided by findings of candidate vocal expressions

for a wide range of emotion categories, along with taxonomic principles derived in the study of emotional experience (Cowen & Keltner, 2017), we tested the following hypotheses: (1) With respect to the dimensionality of emotion recognition in vocal bursts, people will reliably distinguish upwards of twenty distinct dimensions, or kinds, of emotion. (2) With respect to conceptualization, the recognition of vocal bursts in terms of emotion categories will explain the perception of affective appraisals such as valence and arousal, but not vice versa. (3) With respect to the distribution of emotion, emotion categories (e.g., "awe" and "surprise") will be joined by gradients corresponding to smooth variations in how they are perceived (e.g., reliable judgments of intermediate valence along the gradient).

Experiment 1: The Dimensionality, Conceptualization, and Distribution of Emotion Recognition of Voluntarily Produced Vocal Bursts.

Characterizing emotion-related semantic spaces requires: 1) the study of as wide an array of emotions as possible; 2) the study of a rich array of expressions, to expand beyond the study of expression prototypes; and 3) judgments of emotion categories and affective appraisals. Toward these ends, we captured 2032 vocal bursts in the laboratory. One set of participants then judged these vocal bursts in terms of emotion categories. A second set of participants judged the vocal bursts using a free response format. A final set of participants judged the vocal bursts for 13 scales of affect, guided by the latest advances in appraisal and componential theories, which reflect the most systematic efforts to characterize the full dimensionality of emotion using domain-general concepts. With new statistical techniques, we tested hypotheses concerning the dimensionality, conceptualization, and distribution of emotions as recognized in the richest array of vocal bursts studied to date.

Methods.

Creation of Library of Vocal Bursts.

Guided by past methods (Simon-Thomas et al., 2009), we recorded 2032 vocal bursts by asking 56 individuals (26f, ages 18–35) to express their emotions as they imagined being in scenarios that were roughly balanced in terms of 30 emotion categories and richly varying along 13 commonly measured affective appraisals (Tables S1–2). The individuals were recruited from four countries (27 USA, 9 India, 13 Kenya, 7 Singapore) and including professional actors and amateur volunteers (see SOM-I for details). The emotion categories were derived from recent studies demonstrating that a number of categories of emotion are reliably conveyed by vocal bursts (Laukka et al., 2013; Cordaro et al., 2016; Simon-Thomas et al., 2009; for summary, see Table S3 and Keltner et al., 2016) and from findings of states found to reliably occur in daily interactions (Rozin & Cohen, 2003). They included: ADORATION, AMUSEMENT, ANGER, AWE, CONFUSION, CONTEMPT, CONTENTMENT, DESIRE, DISAPPOINTMENT, DISGUST, DISTRESS, ECSTASY, ELATION, EMBARRASSMENT, FEAR, GUILT, INTEREST, LOVE, NEUTRAL, PAIN, PRIDE, REALIZATION, RELIEF, SADNESS, SERENITY, SHAME, SURPRISE (NEGATIVE), SURPRISE (POSITIVE), SYMPATHY, and TRIUMPH. The affective scales were culled from dimensional and componential theories of emotion (for summary, see Table S4). They included VALENCE, AROUSAL, APPROACH/AVOIDANCE,

ATTENTION, CERTAINTY VS. UNCERTAINTY, COMMITMENT-TO-AN-INDIVIDUAL, CONTROL, DOMINANCE VS. SUBMISSIVENESS, FAIRNESS VS. UNFAIRNESS, IDENTITY-WITH-A-GROUP, IMPROVEMENT/WORSENING, OBSTRUCTION, and SAFETY VS. UNSAFETY (Mehrabian & Russell, 1974; Russell, 2003; Scherer, 2009; Smith & Ellsworth, 1985). 425 of the stimuli were recorded as part of the VENEC corpus (Laukka et al., 2013); 1607 are introduced here. Due to potential limitations both in the expressive abilities of the individuals who produced the vocal bursts and the range of emotional scenarios used to elicit the vocal bursts, our results should not be taken as evidence against the possibility of other emotions that might be expressed with vocal bursts—that is, we do not claim that the expressions studied here are fully exhaustive of human vocal expression. Our intention was to create as complex an array of emotional vocal bursts as warranted by the current scientific literature.

Categorical, Affective Appraisal, and Free Response Judgments of Vocal Bursts.

In our study of emotional experience, collecting judgments from approximately 10 participants has proven sufficient to closely approximate the population mean in terms of reliability (Cowen & Keltner, 2017). Presuming that this would extend to vocal bursts, we used Amazon Mechanical Turk to obtain repeated (9–12) judgments of each vocal burst in three randomly assigned formats. A total of 1017 U.S. English-speaking raters ages 18–76 (545 female, mean age = 36) participated. See Figure S1 for a breakdown of participant age and gender by response format. Participants were randomly assigned to one of three emotion recognition conditions. In a first, participants judged each sound in terms of the 30 aforementioned emotion categories, choosing the category that best matched the sound that they heard from a list of 30 presented alphabetically, alongside relevant synonyms. A second group of participants rated each sound in terms of the 13 affective scales, covering not only valence and arousal but other affective appraisals required to differentiate more complex arrays of emotions (Lazarus, 1991; Scherer, 2009; Smith & Ellsworth, 1985). Given the limitations of forced choice formats (DiGirolamo & Russell, 2017; Russell, 1994), a third group of participants provided free response descriptions of the emotion conveyed by each sound. For this, they were instructed to type in whatever interpretation of the vocal burst they felt appropriate. After typing each letter, a set of all possible completions from a comprehensive list of 600 emotion terms appeared (Dataset S1). This search-based format is not entirely unconstrained, but it addresses the critique that multiple choice formats prime participants with a list of concepts and has the advantage of ruling out ambiguities from multiple spellings and conjugations of each term. Examples of each judgment format are given in Fig. 2A-C. The experimental procedures were approved by the Institutional Review Board at the University of California, Berkeley. All participants gave their informed consent.

On average, a participant judged 279.8 vocal bursts. With these methods, a total of 284,512 individual judgments (24,384 forced-choice categorical judgments, 24,384 free response judgments, and 237,744 nine-point affective scale judgments) were gathered to enable us to characterize the semantic space of vocal emotion recognition. These methods were found to capture an estimated 90.8% and 91.0% of the variance in the population mean categorical and affective scale judgments, respectively, indicating that they accurately characterized the population average responses to each individual stimulus, and demonstrating that they did so

with roughly equivalent precision for each judgment type. (See *SOM-I* for details regarding explainable variance calculation. Our methods captured 58.2% of the variance in population mean free response judgments, due to the multiplicity of distinct free responses, but none of our analyses rely on the estimation of mean free response judgments of individual stimuli.)

Results.

How Many Emotions Can Vocal Bursts Communicate?

Our first hypothesis predicted that people recognize upwards of 20 distinct emotions in the 2032 vocal bursts. We tested this hypothesis in two ways: by assessing the degree to which participants agreed in their labeling of the vocal bursts, and by analyzing commonalities in forced choice and free response labeling of the vocal bursts to uncover the number of emotions recognized with both methods.

Forced Choice Labeling.

We first analyzed how many emotion categories were recognized at above-chance levels in judgments of the 2032 vocal bursts. Most typically, emotion recognition has been assessed in the degree to which participants' judgments conform to experimenters' expectations regarding the emotions conveyed by each expression (see Laukka et al., 2013 for such results from an overlapping set of vocal bursts). We depart from this confirmatory approach for two reasons: it assumes that each scenario that prompted the production of a vocal burst can elicit only one emotion (rather than blends of emotions and/or emotions different from experimenters' predictions); and (2) the vocal bursts recorded during each session were not always attributable to a single scenario (see *SOM-I* for details). To capture emotion recognition, we instead focus on the reliable perception (recognition) of emotion in each vocal burst, operationalized as interrater agreement—the number of raters who chose the most frequently chosen emotion category for each vocal burst. This approach is motivated by arguments that the signal value of an expression is determined by how it is normally perceived (Jack et al., 2016).

In this preliminary test of our first hypothesis, we found that 26 of the 30 emotion categories were recognized at above-chance rates from vocal bursts (mean 57.6 significant vocal bursts per category; see Fig. 2D). Overall, 77% of the 2032 vocal bursts were reliably identified with at least one category (false discovery rate [FDR] $q < .05$, simulation test, detailed in *SOM-I*). For interrater agreement levels per vocal burst, as well as the number of significantly recognized vocal bursts in each category, see Fig. 2D. In terms more typically used in the field, the average rate of interrater agreement—that is, the average number of raters who chose the maximally chosen category for each vocal burst—was 47.7%, comparable to recognition levels observed in past vocal burst studies (Elfenbein & Ambady, 2002; Cordaro et al., 2016; Simon-Thomas et al., 2009; Sauter et al., 2010; Laukka et al., 2013). (Note that interrater agreement rates did not differ substantially by culture in which vocal bursts were recorded; see Fig. S2A).

Relating Categories to Free Response Judgments.

Our first finding, then, is that as many as 26 emotion categories are recognized at above-chance levels in vocal bursts in a forced choice format. Such univariate accuracy metrics, however, <u>do not</u> determine the number of categories that are truly distinct in the recognition of emotion, despite their centrality to past studies of emotion recognition. For example, any two categories, such as "sadness" and "disappointment," may both have been reported reliably even if they were used interchangeably (e.g., as synonyms). Given this issue, and potential limitations of forced choice judgments (Russell, 1994; DiGirolamo & Russell, 2017), we tested our first hypothesis in a second way, by determining how many distinct dimensions, or patterns of responses, within the category judgments were also reliably associated with distinct patterns of responses within the free response judgments.

To assess the number of dimensions of emotion, or kinds of emotion, that participants reliably *distinguished* from the 2032 vocal bursts across the two methods, we used canonical correlations analysis (CCA) (Hardoon, Szedmak, & Shawe-Taylor, 2004) (see SOM-I for discussion of limitations of factor analytic approaches used by past studies). This analysis uncovered dimensions in the category judgments that correlated significantly with dimensions in the free response judgments. As one can see in Fig. 3A, we found that 24 distinct dimensions, or kinds of emotion ($p = .014$), were required to explain the reliable correlations between the category and free response judgments of the 2032 vocal bursts. What this means is that there were 24 distinct kinds of emotion evident in the patterns of covariation between forced choice and free response judgments, a finding that goes beyond the interrater agreement findings in showing that people not only reliably identify but also reliably distinguish a wide variety of emotions from the 2032 vocal bursts, and that is in keeping with our first hypothesis. Vocal bursts can communicate at least 24 distinct emotions, nearly doubling the number of emotions recognized from vocal bursts documented in past studies (e.g., Cordaro et al., 2016).

In our next analysis, we determined the *meaning* of the 24 dimensions of emotion conveyed by the vocal bursts. We did so with a technique that reduces the patterns of category judgments to their underlying dimensions, or what one might think of as core meaning. Although different techniques extract rotations of the same dimensional space, we chose non-negative matrix factorization (NNMF) (Lee & Seung, 1999) because it produces dimensions that are easily interpretable, having only positive loadings on each category. We applied NNMF with an increasing number of dimensions, ranging from 1 to the maximum, 29 (the number of emotion categories after excluding "Neutral", which is redundant given the forced-choice nature of the judgments). We then subsequently performed CCA between the resulting dimensions of the emotion category judgments and the free response judgments. The results of this analysis (Fig. 3A) reveal that NNMF extracted 24 dimensions that were reliably shared with the free response judgments ($p = .011$). In other words, when NNMF extracted 24 dimensions from participants' emotion category judgments of the vocal bursts, it captured the 24 distinct kinds of emotion required to explain correlations between category and free response judgments.

To interpret the 24 kinds of emotion conveyed by the vocal bursts, we assessed the weights of each dimension extracted by NNMF on the 30 categories of emotion with which we

began. This analysis determines how each emotion category judgment contributes to each of the 24 dimensions we have uncovered in patterns of covariation across the forced choice and free response methods. Which categories do each of the 24 dimensions represent? The answer to this question is represented in Fig. 3B. In this chart, the colored letters on the horizontal axis refer to the 24 dimensions, and the category names are presented on the vertical axis. The dimensions closely corresponded to: ADORATION, AMUSEMENT, ANGER, AWE, CONFUSION, CONTEMPT, CONTENTMENT, DESIRE, DISAPPOINTMENT, DISGUST, DISTRESS, ECSTASY, ELATION, EMBARRASSMENT, FEAR, INTEREST, PAIN, REALIZATION, RELIEF, SADNESS, SURPRISE (NEGATIVE), SURPRISE (POSITIVE), SYMPATHY and TRIUMPH. (The categories that did not correspond to distinct dimensions—GUILT, PRIDE, ROMANTIC LOVE, SERENITY, and SHAME—were among the least frequently chosen, as shown in Fig. S2B. Hence, the categories people chose more regularly tended to be used reliably.)

To verify that the 24 dimensions of emotion conveyed by vocal bursts can be accurately interpreted in terms of the 24 categories above, we examined whether participants used similar emotion terms in the free response format to describe each vocal burst. The results of this analysis are presented in Fig. 3C, which shows how the dimensions, labeled in colored terms to the left of each row, correspond to the free response terms, in black, chosen to label the 2032 vocal bursts, listed in order of descending correlation. These results indicate that each of the 24 dimensions, or kinds of emotion, was closely correlated with the use of the free response term identical to its maximally loading category found in the forced choice judgments, along with synonyms and related terms. To illustrate, awe emerged as a kind of emotion conveyed with vocal bursts, and in free response format, people were most likely to label vocal bursts conveying awe with "awestruck, amazement, and awe." Thus, the dimensions accurately capture the meaning attributed to each vocal burst, as opposed to reflecting methodological artifacts of forced choice, such as process of elimination (DiGirolamo & Russell, 2017).

Taken together, the results thus far lend support to our first hypothesis: 24 dimensions of emotion were reliably identified in categorical and free response judgments of 2032 vocal bursts.

The Conceptualization of Emotion: Are Emotion Categories, Affective Appraisals, or Both Necessary to Capture the Recognition of Emotion?

What captures the recognition of emotion from vocal bursts: emotion categories, more general affective appraisals (such as valence, arousal, certainty, effort, and dominance), or both? Our second hypothesis derived from our past study of emotional experience (Cowen & Keltner, 2017), and held that participants' categorical judgments of vocal bursts would explain their judgments along the 13 affective scales, but not vice versa. Emotion categories, this hypothesis holds, offer a broader conceptualization of emotion recognition than the most widely studied affective appraisals.

This prediction can be formally tested using statistical models. With the present data we used cross-validated predictive models, which determined the extent to which participants' affective scale judgments (of valence, arousal and so on) explained participants' categorical

judgments, and vice versa. In keeping with our second hypothesis, we found that using both linear (ordinary least squares) and nonlinear (nearest neighbors) models, the categories explain almost all of the explainable variance in the scales of affect (90% and 97%), whereas the scales of affect explain at most around half of the explainable variance in the categories (32% and 52%). (See *SOM-I* for more details of this analysis.) In Fig. 4A we illustrate how the information carried in participants' ratings using the 13 affective scales—valence, arousal, dominance, etc. as represented in green—is a subspace of the information carried by the emotion categories, as represented in orange. What this means is that how people conceptualize emotion in vocal bursts in terms of valence, arousal, certainty, and so on can be explained by how they categorize the emotion with emotion categories, but there is variance in the emotions conveyed by vocal bursts that is not explained by valence, arousal, dominance, and other affective appraisals. This finding supports our second hypothesis, that emotion categories have greater explanatory value than the affective scales in representing how people recognize emotion in vocal bursts.

Are Emotion Categories Discrete or Not? The Distribution of Emotion Categories.

To test our third hypothesis, that we would find smooth gradients between the different categories of emotion recognized in vocal bursts, we first visualized the distribution of the category judgments of vocal bursts within the 24-dimensional space we have uncovered. We did so using a method called t-distributed stochastic neighbor embedding (t-SNE) (Maaten & Hinton, 2008). This method projects high-dimensional data into a two-dimensional space in a way that attempts to preserve the local distances between data points, in this case separate vocal bursts, as much as possible, but allows for distortions in longer distances between data points, with the goal of illustrating the shape of the distribution of the data. We applied t-SNE to the projections of the vocal burst categorical judgments into the 24-dimensional space extracted using NNMF (Fig. 4). As expected, t-SNE distorted larger distances between vocal bursts in the 24-dimensional space but preserved more fine-grained similarities between vocal bursts, such as continuous gradients of perceived emotional states, illustrated by smooth variations in color in Fig. 4. In an online interactive version of Fig. 4 (https://s3-us-west-1.amazonaws.com/vocs/map.html), each of the 2032 sounds can be played in view of their relative positions within the 24-dimensional space of vocal emotion. This visualization simultaneously represents the distribution of the 24 distinct emotion categories uncovered in prior analyses within a semantic space, the variants of vocal bursts within each emotion category, and the gradients that bridge categories of emotion.

Visual and auditory inspection of this representation of the semantic space of emotion recognition in vocal bursts yields several insights that converge with claims in the science of emotion: that the emotions of moral disapproval—anger, disgust, and contempt (see bottom left of figure) are in a similar general space but distinct (Rozin, Lowery, & Haidt, 1999), that one set of emotions—interest, awe, realization, and confusion (see middle, to the right of the figure)—track the individual's understanding of the world (Shiota et al., 2017), that expressions of love (adoration) and desire are distinct, as anticipated by attachment theory (Gonzaga, Keltner, Londahl, & Smith, 2001), that embarrassment and amusement (see upper left of figure) are close in meaning (Keltner, 1996) as are surprise and fear (Ekman, 1993).

As evident in these figures, and in keeping with our third hypothesis, the semantic space of emotion recognition from the human voice is a complex, non-discrete structure with smooth within-category variations in the relative location of each individual vocal burst. For example, some vocal bursts most frequently labeled "negative surprise" were placed further away from vocal bursts labeled fear, while others were placed adjacent to fear. Listening to these different surprise vocal bursts gives the impression that the ones placed closer to fear signal a greater potential for danger than those placed further away, suggesting that surprise and fear may occupy a continuous gradient despite typically being considered discrete categories. Similar gradients were revealed between a variety of other category pairs, bridging anger with pain, elation with triumph, relief with contentment, desire with ecstasy, and more (Fig. 4, left).

One explanation for these results is that many of the vocalizations were perceived one way by some participants (e.g., as anger) and differently by others (e.g., as pain). An alternative possibility is that while some vocal bursts clearly signal one emotion, others reliably convey intermediate or mixed emotional states. To address these contrasting possibilities, we compared proximity to category boundaries (operationalized as the standard deviation in category ratings) to the standard deviation in ratings of the affective scales. If the smooth gradients between categories were caused by ambiguity in meaning, then vocal bursts that fell near the boundary between two categories that consistently differed in their affective appraisals would also have received highly variable affective scale judgments, across participants.

We find that gradients between categories correspond to smooth variations in meaning. In many cases, neighboring categories (e.g., adoration, sympathy) differed dramatically in their affective appraisals, but intermediacy between categories did not relate to *ambiguity* in affective appraisals such as valence, as seen in Fig. 4, right. Overall, the correlation between category and affective scale ambiguity—the sum of the standard deviations of each attribute —was quite weak (Pearson's $r = .14$, Spearman's $r = .13$). These results suggest that the smooth gradients between categories largely reflect what might be thought of as emotional blends of categories, as opposed to ambiguity in meaning. These findings align with the recent interest in blended emotions captured in subjective experience (Cowen & Keltner, 2017; Watson & Stanton, 2017) and facial expression (Du, Tao, & Martinez, 2014). For further analyses confirming that smooth gradients between specific categories correspond to smooth variations in meaning, see Figure S3.

Experiment 2: The conceptualization of emotion in real-world vocal bursts Methods.

A shortcoming in our first study, common to the study of emotional vocalization, is that the vocal bursts were produced in laboratory settings. This raises the question of whether our findings extend to vocalizations found in the real world (Anikin & Lima, 2017; Juslin, Laukka, & Bänziger, 2018; Sauter & Fischer, 2018; Scherer, 2013). With respect to the *conceptualization* of emotion recognition in vocal bursts, concerns regarding ecological validity may limit our finding that emotion categories explain the signaling of affective scales such as valence and arousal, given that signals of valence, arousal, and so on may vary more in ecological contexts (e.g., Russell, 1994). To address this possibility, we obtained

judgments of a set of 48 vocal bursts recorded in more naturalistic settings and used these data to further examine whether categories judgments reliably capture scales of affect in real-life vocal bursts.

Guided by the 24 kinds of emotion recognized in vocal bursts documented in our first study, we extracted a library of 48 vocal bursts from richly varying emotional situations captured in naturalistic YouTube videos. These contexts were intended to span the 24 dimensions of emotion observed in Study 1 and to richly vary along the 13 affective scales (see Table S6 for description of the contexts in which the sounds were extracted, predicted categorical and affective appraisals, as well as links to each sound). The vocal bursts were found in emotionally evocative situations—watching a baby's pratfalls, viewing a magic trick, disclosures of infidelity or the loss of virginity, observing physical pain, eating good food or receiving a massage. Videos were located using search terms associated with of emotional content (e.g., "puppy", "hugged", "falling", "magic trick") and vocal bursts were extracted from emotional situations within each video. We obtained repeated (18) judgments of each vocal burst in terms of the 30 categories and the 13 affective scales, as in Experiment 1, using Amazon Mechanical Turk. A total of 88 additional U.S. English-speaking raters (46 female, mean age = 36) participated in these surveys.

Results.

Interrater agreement for the naturalistic vocal bursts averaged 42.6%, comparable to those observed for laboratory-induced vocal bursts in Experiment 1 (Fig. S2A, S4), despite potential limitations in the quality of recordings found on YouTube. Using predictive models estimated on data from Experiment 1 and tested on data from Experiment 2, we determined the extent to which categorical judgments explained affective appraisal judgments and vice versa in the naturalistic vocal bursts. Replicating results from Experiment 1, and in keeping with Hypothesis 2, we found that the categories explained almost all of the explainable variance in the affective scales (89% and 92%) whereas the affective scales explained around half of the explainable variance in the categories (41% and 59%), as shown in Fig. 4B. These results extend our findings from vocal bursts collected in the laboratory to samples from the real world, confirming that emotion categories capture, but cannot be captured by, affective scales representing core affect themes such as valence, arousal, dominance, and so on.

Discussion.

The human voice is a complex medium of emotional communication that richly structures social interactions. Recent studies have made progress in understanding which emotions are recognized in the voice, how this capacity emerges in development and is preserved across cultures, and parallels between emotional communication in the voice and feelings induced by music. Given past methodological tendencies (e.g., the focus on prototypical vocalizations of a limited number of emotions), what is only beginning to be understood is how the emotions conveyed by the voice are represented within a multidimensional space.

Toward this end, we have proposed a theoretical approach to capturing how people represent emotion-related experience and expression within a semantic space (Cowen & Keltner,

2017, 2018). Semantic spaces are defined by their dimensionality, or number of emotions, their conceptualization of emotional states (e.g., specific emotion categories or domain-general affective appraisals), and the distribution of emotion-related responses, in all their variation, within that space. Building upon a study of emotional experience (Cowen & Keltner, 2017), in the present study we examined the semantic space of emotion recognition of vocal bursts. We gathered vocal bursts from 56 individuals imagining over a hundred scenarios richly varying in terms of 30 categories of emotion and 13 affective scales, yielding the widest array of emotion-related vocalizations studied to date. Participants judged these 2032 vocal bursts with items derived from categorical and appraisal/constructionist approaches, as well as in free response format. These methods captured a semantic space of emotion recognition in the human voice, revealing the kinds of emotion signaled and their organization within a multidimensional space.

Taken together, our results yield data-driven insights into important questions within emotion science. Vocal bursts are richer and more nuanced than typically thought, reliably conveying 24 dimensions of emotion that can be conceptualized in terms of emotion categories. In contrast to many constructivist and appraisal theories, these dimensions cannot be explained in terms of a set of domain-general appraisals, most notably valence and arousal, that are posited to underlie the recognition of emotion and commonly used in the measurement of emotion-related response. However, in contrast to discrete emotion theories, the emotions conveyed by vocal bursts vary continuously in meaning along gradients between categories. These results converge with doubts that emotion categories "carve nature at its joints" (Barrett, 2006). Visualizing the distribution of 2032 vocal bursts along 24 continuous semantic dimensions (https://s3-us-west-1.amazonaws.com/vocs/map.html) demonstrates the variety and nuance of emotions they signal.

Vocal bursts signal myriad positive emotions that have recently garnered scientific attention, such as adoration, amusement, awe, contentment, desire, ecstasy, elation, interest, and triumph (Shiota et al., 2017; Tracy & Matsumoto, 2008). Evidence for expressions distinguishing these states validates the recent expansion in the range of emotions investigated empirically. So, too, do findings of distinctions between nuanced states relevant to more specific theoretical claims: adoration/sympathy, anger/disappointment, distress/fear, and negative/positive surprise (Egner, 2011; Johnson & Connelly, 2014; Reiss, 1991; Shiota et al., 2017).

That many categories were bridged by smooth gradients converges with recent studies of emotional blends in subjective experience and facial expression (Cowen & Keltner, 2017; Watson & Stanton, 2017; Du et al., 2014). Across gradients between categories, vocal signals vary continuously in meaning. For example, as they traverse the gradient from "disappointment" to "sympathy" to "love", vocal bursts increasingly signal the desire to approach (Fig. S3). These findings point to a rich landscape of emotional blends in vocalization warranting further study.

The present findings also inform progress in related fields, including the neural basis of emotion recognition and the training of machines to decode emotion. Future neuroscience studies will need to expand in focus beyond discrete prototypes of a few emotion categories

to explain how the rich variety of categories are distinguished by the brain and how continuous variation between categories may be represented in continuously varying patterns of brain activity (for indications of this, see (Harris, Young, & Andrews, 2012). Similarly, machine learning efforts to decode emotion from audio will need to expand in focus beyond a small set of discrete categories or two affective appraisals (Scherer, Schüller, & Elkins, 2017) to account for how a wide array of categories and continuous gradients between them are conveyed by acoustic features (for relationship with duration, fundamental frequency, and harmonic-to-noise ratio, see Figure S5).

In considering these findings, it is worth noting that they are based on vocal emotion recognition by U.S. participants. Further work is needed to examine variation in the semantic space of recognition in other cultures, though preliminary studies have suggested that upwards of 20 emotion categories may be recognized even in remote cultures (Cordaro et al., 2016), supporting possible universals in the semantic space of vocal emotion. Likewise, given reports of developmental change and age-related decline in vocal emotion recognition (Lima, Alves, Scott, & Castro, 2014; Sauter, Panattoni, & Happé, 2013), it would also be interesting to examine variation in the semantic space of vocal emotion recognition across the life span. Finally, given the acoustic similarities between emotion recognition in vocalization and in music (Juslin & Laukka, 2003; Laukka & Juslin, 2007), it would be interesting to compare these semantic spaces.

The present findings dovetail with recent inquiries into the semantic space of emotional experience (Cowen & Keltner, 2017). There is substantial overlap between the varieties of emotion recognition uncovered in the present study and those identified in reported emotional experience, including ADORATION, AMUSEMENT, ANGER, AWE, CONFUSION, CONTENTMENT / PEACEFULNESS, DESIRE, DISGUST, DISTRESS / ANXIETY, ELATION / JOY, EMBARRASSMENT /AWKWARDNESS, FEAR, INTEREST, PAIN, RELIEF, SADNESS, and SURPRISE. In both studies, categories captured a broad space of emotion recognition, and continuous gradients bridged categories such as interest and awe. Together, these results converge on a high-dimensional taxonomy of emotion defined by a rich array of categories bridged by smooth gradients.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

Anikin A, & Lima CF (2017). Perceptual and acoustic differences between authentic and acted nonverbal emotional vocalizations. Q J Exp Psychol A, 1–21.

Arnal LH, Flinker A, Kleinschmidt A, Giraud AL, & Poeppel D (2015). Human screams occupy a privileged niche in the communication soundscape. Curr Biol, 25(15), 2051–2056. [PubMed: 26190070]

Banse R, & Scherer KR (1996). Acoustic profiles in vocal emotion expression. J Pers Soc Psychol, 70(3), 614–636. [PubMed: 8851745]

Barrett LF (2006). Are emotions natural kinds? Persp Psychol Sci, 1(1), 28–58.

Cordaro DT, Keltner D, Tshering S, Wangchuk D, & Flynn L (2016). The voice conveys emotion in ten globalized cultures and one remote village in Bhutan. Emotion 16(1), 117. [PubMed: 26389648]

Bryant GA, Fessler DM, Fusaroli R, Clint E, … & De Smet D. (2016). Detecting affiliation in colaughter across 24 societies. Proc Nat Acad Sci USA, 113(17), 4682–4687. [PubMed: 27071114]

Cowen AS, & Keltner D (2017). Self-report captures 27 distinct categories of emotion bridged by continuous gradients. Proc Nat Acad Sci USA, 201702247.

Cowen AS, & Keltner D (2018). Clarifying the conceptualization, dimensionality, and structure of emotion: Response to Barrett and colleagues. Trends Cogn Sci. 22(4), 274–276. [PubMed: 29477775]

DiGirolamo MA, & Russell JA (2017). The emotion seen in a face can be a methodological artifact: The process of elimination hypothesis. Emotion, 17(3), 538. [PubMed: 27893222]

Du S, Tao Y, & Martinez AM (2014). Compound facial expressions of emotion. Proc Nat Acad Sci USA, 111(15), E1454–E1462. [PubMed: 24706770]

Egner T (2011). Surprise! A unifying model of dorsal anterior cingulate function? Nat Neurosci, 14(10), 1219–1220. [PubMed: 21952261]

Ekman P (1993). Facial expression and emotion. Am Psychol, 48(4), 384–392. [PubMed: 8512154]

Ekman P, & Cordaro D (2011). What is meant by calling emotions basic. Emot Rev, 3(4), 364–370.

Elfenbein HA, & Ambady N (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. Psychol Bull. 128(2), 203. [PubMed: 11931516]

Fernández-Dols JM, & Crivelli C (2013). Emotion and expression: Naturalistic studies. Emot Rev. 5(1), 24–29.

Frühholz S, Trost W, & Kotz SA (2016). The sound of emotions - towards a unifying neural network perspective of affective sound processing. Neurosci Biobehav Rev, 68, 1–15. [PubMed: 27179453]

Gendron M, Roberson D, van der Vyver JM, & Barrett LF (2014). Perceptions of emotion from facial expressions are not culturally universal: Evidence from a remote culture. Emotion, 14(2), 251–262. [PubMed: 24708506]

Gonzaga GC, Keltner D, Londahl EA, & Smith MD (2001). Love and the commitment problem in romantic relations and friendship. J Pers Soc Psychol, 81(2), 247–262. [PubMed: 11519930]

Hardoon D, Szedmak S, & Shawe-Taylor J (2004). Canonical correlation analysis: An overview with application to learning methods. Neural Computation, 16(12), 2639–2664. [PubMed: 15516276]

Harris RJ, Young AW, & Andrews TJ (2012). Morphing between expressions dissociates continuous from categorical representations of facial expression in the human brain. Proc Nat Acad Sci USA, 109(51), 21164–21169. [PubMed: 23213218]

Hertenstein MJ, & Campos JJ (2004). The retention effects of an adult's emotional displays on infant behavior. Child Dev, 75(2), 595–613. [PubMed: 15056208]

Jack RE, Sun W, Delis I, Garrod OGB, & Schyns PG (2016). Four not six: Revealing culturally common facial expressions of emotion. J Exp Psychol Gen, 145(6), 708–730. [PubMed: 27077757]

Juslin PN, & Laukka P (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? Psychol Bull. 129(5), 770 [PubMed: 12956543]

Juslin PN, Laukka P, & Bänziger T (2018). The mirror to our soul? Comparisons of spontaneous and posed vocal expression of emotion. J Nonverb Beh, 42(1).

Keltner D, Cordaro DT (2015). Understanding Multimodal Emotional Expression: Recent Advances in Basic Emotion Theory. Emotion Researcher.

Keltner D (1996). Evidence for the distinctness of embarrassment, shame, and guilt: A study of recalled antecedents and facial expressions of emotion. Cogn Emot, 10(2), 155–172.

Keltner D, & Kring A (1998). Emotion, social function, and psychopathology. Rev Gen Psychol, 2(3), 320–342.

Keltner D, & Lerner JS (2010). Emotion In Fiske ST, Gilbert DT, & Lindzey G (Eds.), Handbook of Social Psychology. Wiley Online Library.

Keltner D, Tracy J, Sauter DA, Cordaro DC, & McNeil G (2016). Expression of Emotion In Barrett LF, Lewis M (Ed.), Handbook of Emotions (pp. 467–482). New York, NY: Guilford Press.

Kraus MW (2017). Voice-only communication enhances empathic accuracy. Am Psychol, 72(7), 644–654. [PubMed: 29016168]

Laukka P, Elfenbein HA, Söder N, Nordström H, Althoff J, Chui W, … Thingujam NS (2013). Cross-cultural decoding of positive and negative non-linguistic emotion vocalizations. Front Psychol. 4, 353. [PubMed: 23914178]

Laukka P, Elfenbein HA, Thingujam NS, Rockstuhl T, Iraki FK, Chui W, & Althoff J (2016). The expression and recognition of emotions in the voice across five nations: A lens model analysis based on acoustic features. J Pers Soc Psychol. 111(5), 686. [PubMed: 27537275]

Laukka P, & Juslin PN (2007). Similar patterns of age-related differences in emotion recognition from speech and music. Motiv Emot, 31(3), 182–191.

Lazarus RS (1991). Progress on a cognitive-motivational-relational theory of emotion. Am Psychol, 46(8), 819. [PubMed: 1928936]

Lee DD, & Seung HS (1999). Learning the parts of objects by non-negative matrix factorization. Nature, 401(6755), 788–791. [PubMed: 10548103]

Johnson G, & Connelly S (2014). Negative emotions in informal feedback: The benefits of disappointment and drawbacks of anger. Human Relations, 67(10), 1265–1290.

Lima CF, Alves T, Scott SK, & Castro SL (2014). In the ear of the beholder: How age shapes emotion processing in nonverbal vocalizations. Emotion, 14(1), 145. [PubMed: 24219391]

Maaten L, & Hinton G (2008). Visualizing data using t-SNE. J Mach Learn Res, 9, 2579–605.

Matsumoto D, Keltner D, Shiota MN, O'Sullivan M, & Frank M (2008). Facial expressions of emotion In Lewis M, Haviland-Jones JM, & Barrett LF (Eds.), Handbook of Emotions (Vol. 3, pp. 211–234). Guilford Press New York, NY.

Mehrabian A, & Russell J (1974). An approach to environmental psychology. MIT Press.

Mitchell RLC, & Ross ED (2013). Attitudinal prosody: What we know and directions for future study. Neurosci Biobehav Rev. 37(3), 471–479. [PubMed: 23384530]

Nordström H, Laukka P, Thingujam NS, Schubert E, & Elfenbein HA (2017). Emotion appraisal dimensions inferred from vocal expressions are consistent across cultures: A comparison between Australia and India. Royal Society Open Science 4(11), 170912. [PubMed: 29291085]

Oveis C, Spectre A, Smith PK, Liu MY, & Keltner D (2013). Laughter conveys status. J Exp Soc Psychol, 65, 109–115.

Parsons CE, Young KS, Joensson M, Brattico E, Hyam JA, Stein A, … Kringelbach ML (2014). Ready for action: A role for the human midbrain in responding to infant vocalizations. Soc Cogn Affect Neurosci, 9(7), 977–984. [PubMed: 23720574]

Provine RR, & Fischer KR (1989). Laughing, Smiling, and Talking: Relation to Sleeping and Social Context in Humans. Ethology, 83(4), 295–305.

Reiss S (1991). Expectancy model of fear, anxiety,and panic. Clin Psych Rev, 11(2), 141–153.

Rozin P, & Cohen AB (2003). High frequency of facial expressions corresponding to confusion, concentration, and worry in an analysis of naturally occurring facial expressions of Americans. Emotion, 3(68), 1.

Rozin P, Lowery L, & Haidt J (1999). The CAD triad hypothesis: A mapping between three moral emotions (contempt, anger, disgust) and three moral codes (community, autonomy, divinity). J Pers Soc Psychol. 76(4):574–86. [PubMed: 10234846]

Russell JA (1991). Culture and the categorization of emotions. Psychol Bull, 110(3), 426. [PubMed: 1758918]

Russell JA (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. Psychol Bull. 115(1), 102. [PubMed: 8202574]

Russell JA (2003). Core affect and the psychological construction of emotion. Psychol Rev, 110(1), 145. [PubMed: 12529060]

Sauter DA, Eisner F, Ekman P, & Scott SK (2010). Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. Proc Nat Acad Sci. 107(6), 2408–2412. [PubMed: 20133790]

Sauter DA, & Fischer AH (2018). Can perceivers recognise emotions from spontaneous expressions? Cogn Emot 32(3), 504–515. [PubMed: 28447544]

Scherer K, Johnstone T, & Klasmeyer G (2003). Vocal expression of emotion. In Davidson RJ, Scherer KR, Goldsmith H (Eds.), Handbook of Affective Sciences.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Scherer KR (1986). Vocal affect expression. A review and a model for future research. Psychol Bull 99(2), 143. [PubMed: 3515381]

Scherer KR (2009). The dynamic architecture of emotion: Evidence for the component process model. Cogn Emot, 23(7), 1307–1351.

Scherer KR (2013). Vocal markers of emotion: Comparing induction and acting elicitation. Computer Speech & Language, 27(1), 40–58.

Scherer KR, & Wallbott HG (1994). Evidence for universality and cultural variation of differential emotion response patterning. J Per Soc Psychol, 66(2), 310–328.

Scherer K, Schüller B, & Elkins A (2017). Computational analysis of vocal expression of affect: Trends and challenges. Social Signal Processing, 56.

Scott SK, Lavan N, Chen S, & McGettigan C (2014). The social life of laughter. Trends Cogn Sci, 18(12).

Shiota MN, Campos B, Oveis C, Hertenstein MJ, Simon-Thomas E, & Keltner D (2017). Beyond happiness: Toward a science of discrete positive emotions. Am Psychol, 72(7), 617–643. [PubMed: 29016167]

Shuman V, Clark-Polner E, Meuleman B, Sander D, & Scherer KR (2017). Emotion perception from a componential perspective. Cogn Emot, 31(1), 47–56. [PubMed: 26308096]

Simon-Thomas ER, Keltner DJ, Sauter D, Sinicropi-Yao L, & Abramson A (2009). The voice conveys specific emotions: Evidence from vocal burst displays. Emotion 9(6), 838 [PubMed: 20001126]

Smith CA, & Ellsworth PC (1985). Patterns of cognitive appraisal in emotion. J Pers Soc Psychol, 48(4), 813. [PubMed: 3886875]

Smoski MJ, & Bachorowski JA (2003). Antiphonal laughter between friends and strangers. Cogn Emot 17(2), 327–340. [PubMed: 29715722]

Snowdon CT (2002). Expression of emotion in nonhuman animals In Davidson RJ, Scherer KR, Goldsmith HH (Eds.), Handbook of Affective Sciences. Oxford University Press.

Titze IR, & Martin D (1998). Principles of voice production. J Acoust Soc Am, 104(3), 1148.

Tracy JL, & Matsumoto D (2008). The spontaneous expression of pride and shame: Evidence for biologically innate nonverbal displays. Proc Nat Acad Sci USA, 105(33), 11655–11660. [PubMed: 18695237]

Van Kleef GA (2010). The emerging view of emotion as social information. Soc Personal Psychol Compass, 4(5), 331–343.

Vidrascu L, & Devillers L (2005). Real-life emotion representation and detection in call centers data. In Lecture Notes in Computer Science (Vol. 3784 LNCS, pp. 739–746).

Watson D, & Stanton K (2017). Emotion blends and mixed Emotions in the hierarchical structure of affect. Emot Rev, 9(2), 99–104.

Wu Y, Muentener P, & Schulz LE (2017). One- to four-year-olds connect diverse positive emotional vocalizations to their probable causes. Proc National Acad Sci USA, 201707715.
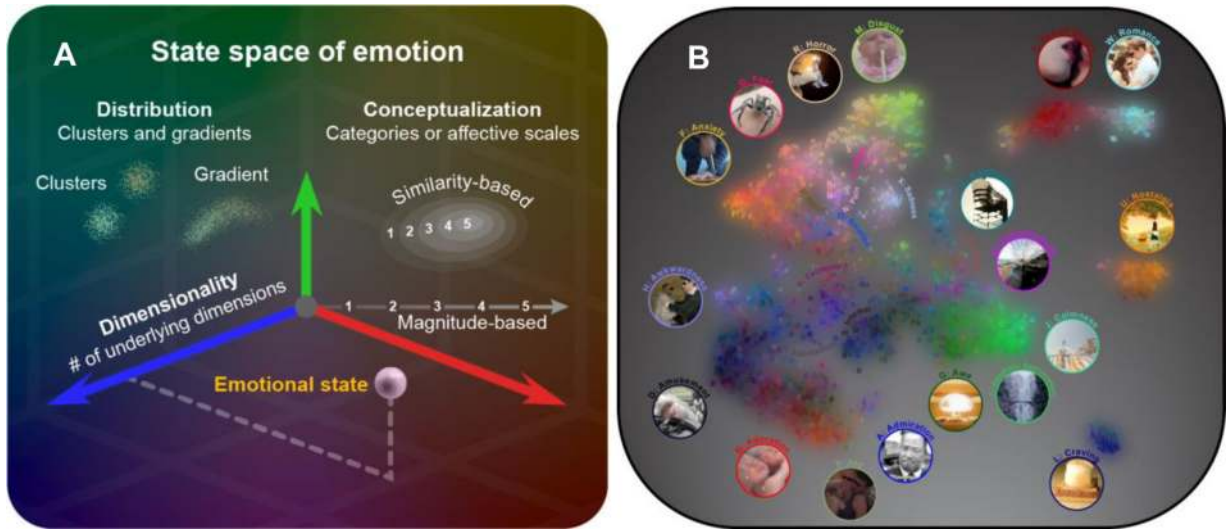
Fig. 1. A. Schematic representation of a semantic space of emotion.
A semantic space of emotion is described by (1) its dimensionality, or the number of varieties of emotion within the space; (2) the conceptualization of the emotional states in the space in terms of categories and more general affective features; that is, the concepts which tile the space (upper right); and (3) the distribution of emotional states within the space. A semantic space is the subspace that is captured by concepts (e.g., categories, affective features). For example, a semantic space of expressive signals of the face is a space of facial movements along dimensions that can be conceptualized in terms of emotion categories and affective features, and that involves clusters or gradients of expression along those dimensions. **B. 27-dimensional semantic space of reported emotional experience evoked by 2185 short videos.** Each letter represents a video. Videos are positioned and colored according to 27 distinct dimensions required to account for the emotions people reliably reported feeling in response to different videos, which corresponded to emotion categories. Within the space, we can see that there are gradients of emotion between categories traditionally thought of as discrete. Images from illustrative videos are shown for 18 of the dimensions. An interactive version of the figure is available here: https://s3-us-west-1.amazonaws.com/emogifs/map.html.
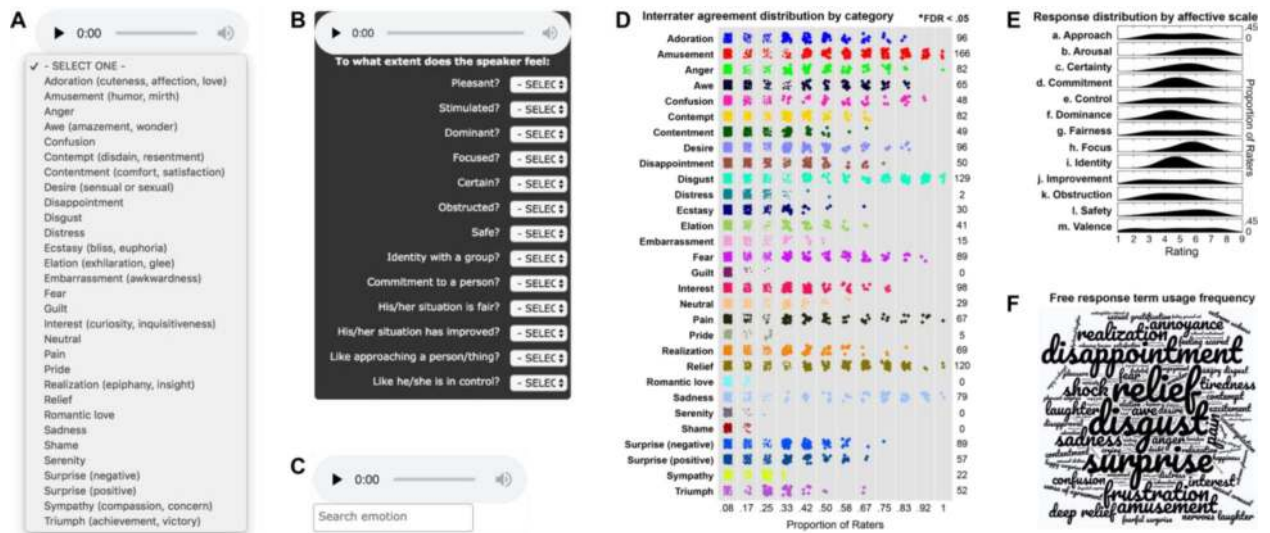
Fig. 2. Judgment formats and response distributions. A-C.
Example survey segments are shown from the (A) categorical, (B) affective scale, and (C) free response surveys. **D. Interrater agreement levels per vocal burst, per category for all 30 categories.** Non-zero interrater agreement rates are shown for each category across all 2032 vocal bursts. The number of vocal bursts in each category with significant interrater agreement rates (FDR < .05, simulation test described above) is shown to the right of each dot plot, and varies from 0 (guilt, romantic love, serenity, and shame) to 166 (amusement); mean = 57.6, SD = 43.0. Dots have been jittered for clarity. **E. Response distribution by affective scale.** Smooth histograms indicate the distribution of average judgments of the vocal bursts for each affective scale. **F. Free response term usage frequency.** The height of each term is directly proportion to its usage frequency. The most commonly used term, "disgust" was applied 1074 times.
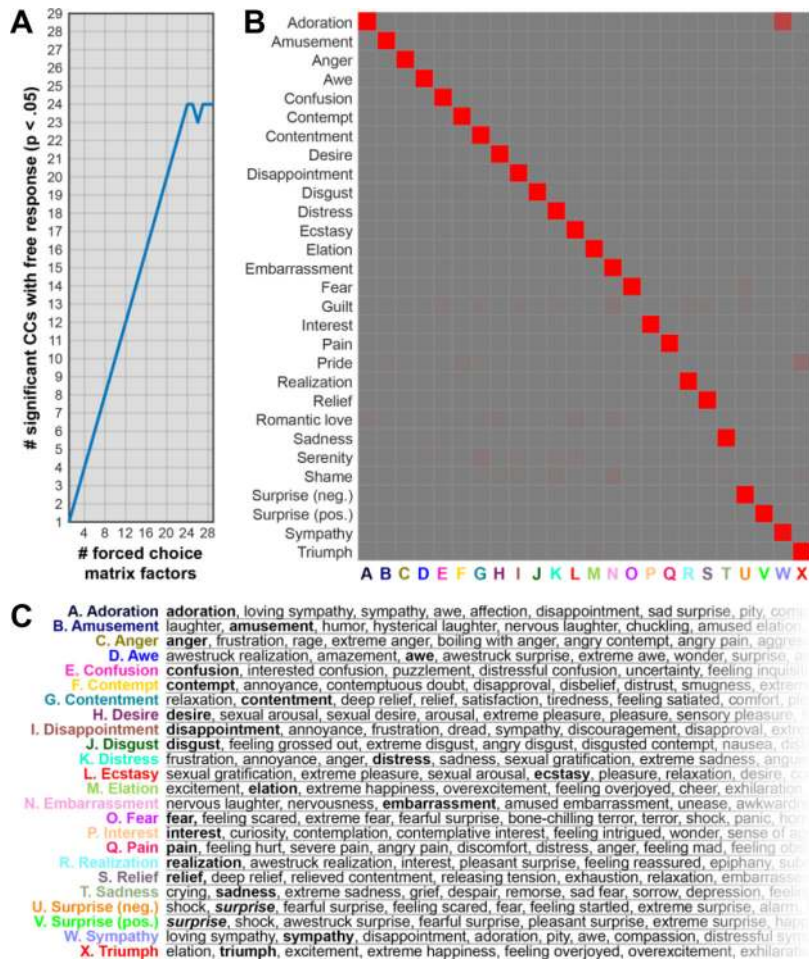
Fig. 3. Vocal bursts reliably and accurately convey 24 distinct dimensions of emotion.
A. The application of canonical correlations analysis (CCA) between the categorical and free response judgments revealed 24 shared dimensions (p<.05). Here, the results of a subsequent analysis reveal that this 24-dimensional space can be derived by applying non-negative matrix factorization (NNMF) to the categorical judgments. NNMF was applied iteratively with increasing numbers of factors, from 1 to 29. The first 24 all span separate dimensions of meaning shared with the free response judgments when subsequently applying CCA ($p < .05$). **B.** Shown here are the 24 dimensions of emotion conveyed by vocal bursts, extracted by applying NNMF to the category judgments. We can see that they each load maximally on a distinct category. **C.** The free response terms most closely correlated with each of the 24 dimensions consistently include this maximally loading category and related terms (for corresponding correlation values, see Table S5). Together, these results reveal that vocal bursts can reliably (in terms of consistency across raters) and accurately (in terms of recognized meaning) convey at least 24 semantically distinct semantic dimensions of emotion.
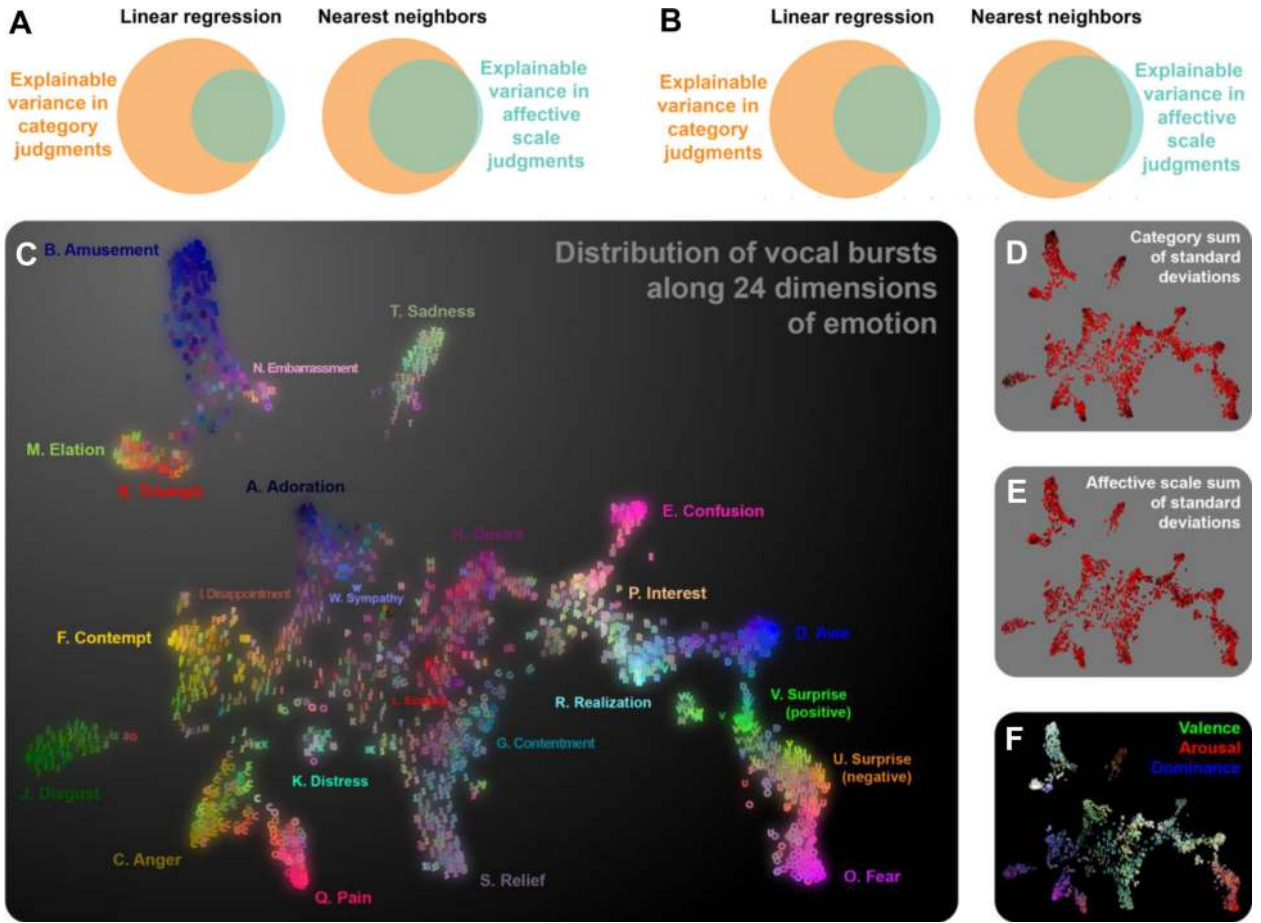
Fig. 4A. Affective scale judgments of the vocal bursts are explained by, but cannot explain, the category judgments.
Both linear and nonlinear models reveal that the explainable variance in the 13 affective scale judgments is almost entirely a subspace of the explainable variance in the forced-choice categorical judgments. (See *SOM-I* for details.) Thus, the categories have more explanatory value than the scales of affect in explicating how emotions are reliably recognized from vocal bursts. **B. These results extend to real-world expressions.** Models were trained on data from Experiment 1 and tested on ratings of the naturalistic vocal bursts gathered in Experiment 2. Both linear and nonlinear models reveal that the explainable variance in the 13 affective scale judgments is almost entirely a subspace of the explainable variance in the forced-choice categorical judgment. The categories have more explanatory value than the scales of affect in explicating how emotions are reliably recognized from vocal bursts. **C. Map of 2032 vocal bursts within a 24-dimensional space of vocal emotion generated with t-distributed stochastic neighbor embedding (t-SNE).** T-SNE approximates local distances between vocal bursts without assuming linearity (factor analysis) or discreteness (clustering). Each letter corresponds to a vocal burst and reflects its maximal loading on 24 dimensions extracted from the category judgments. Unique colors assigned to each category are interpolated to smoothly reflect the loadings of each vocal burst on each dimension. The map reveals smooth structure in the data, with variations in the proximity of vocal bursts within each category to other categories. See the interactive map

(https://s3-us-west-1.amazonaws.com/vocs/map.html) to explore every sound and view its ratings. **D & E. Standard deviations in judgments of the categories and affective scales.** A black-to-red scale represents the sum, across categories or affective scales, of the standard deviations in judgments across participants. For example, vocal bursts plotted in black had lower standard deviations in categorical judgments (i.e., high interrater agreements) and thus were further from the perceptual boundaries between categories. There is little relationship between affective scale ambiguity and proximity to category boundaries, suggesting that the categories are confused primarily because they are bridged by smooth gradients in meaning rather than due to ambiguity. See Fig. S3 for further analyses confirming this for specific gradients. **F. Cross-category gradients correlate with variations in valence (green), arousal (red), and dominance (blue).** Despite the weak relationship between affective scale ambiguity and proximity to category boundaries, categories predict smooth gradients in affective meaning, further establishing that these gradients are not a byproduct of ambiguity. We can also see why the affective scales are insufficient to explain the category judgments, given that almost every unique color—i.e. combination of valence, arousal, and dominance —is associated with multiple regions in the semantic space of emotion recognized in vocal bursts.