nature human behaviour

ARTICLES

https://doi.org/10.1038/s41562-019-0533-6

The primacy of categories in the recognition of 12 emotions in speech prosody across two cultures

Alan S. Cowen[1*], Petri Laukka[2], Hillary Anger Elfenbein[3], Runjing Liu[4] and Dacher Keltner[1]

Central to emotion science is the degree to which categories, such as Awe, or broader affective features, such as Valence, underlie the recognition of emotional expression. To explore the processes by which people recognize emotion from prosody, US and Indian participants were asked to judge the emotion categories or affective features communicated by 2,519 speech samples produced by 100 actors from 5 cultures. With large-scale statistical inference methods, we find that prosody can communicate at least 12 distinct kinds of emotion that are preserved across the 2 cultures. Analyses of the semantic and acoustic structure of the recognition of emotions reveal that emotion categories drive the recognition of emotions more so than affective features, including Valence. In contrast to discrete emotion theories, however, emotion categories are bridged by gradients representing blends of emotions. Our findings, visualized within an interactive map, reveal a complex, high-dimensional space of emotional states recognized cross-culturally in speech prosody.

The recognition of emotions is fundamental to human social interactions. Brief emotional displays in the face and voice by nearby adults guide the responses of infants and children to their environment, and are important for the ways in which adults negotiate rank and status, establish trust, discern affection and commitment, and forgive each other[1–6]. Given the centrality of emotional expression to social life, it is not surprising that the recognition of facial expression and emotion-related vocalization are processed in specific brain regions[7–11], are preserved to a considerable extent across many cultures[12–16] and have evolutionary homologies in a wide range of primate species and even other mammals[17–20].

Early in the study of emotion recognition, empirical studies focused on prototypical facial expressions of six to eight categories of emotions[13,21,22]. More recently, scientists have begun to document the varied ways in which humans communicate emotions using their voice. With short bursts of sound, known as vocal bursts, humans can communicate more than 15 emotions, a finding now replicated in over a dozen cultures, including two remote cultures with minimal contact with the West[16,23]; but see a previous study[24] for findings of greater cultural relativity. With fleeting emotional vocalizations, parents communicate to infants what is worth approaching or what warrants avoidance[25], adults infer a person's rank within a social hierarchy and singers convey specific emotions in song[26] (reviewed in a previous study[27]). By the age of two, children can readily identify at least five positive emotions from brief emotion-related vocalizations[28].

In the present study, we focus on emotional prosody—the nonlexical patterns of tune, rhythm and timbre in speech, modulated by the implements of human vocal control: air pressure from the lungs, tension in the vocal cords and filtration through the throat, tongue, palate, cheeks, lips and nasal passages[29]. (Some definitions of prosody exclude timbre, but we include it here for simplicity, as described in Supplementary Discussion 1.) Prosody interacts with spoken words to convey emotional feelings and attitudes, including dispositions felt toward the objects and ideas described in speech[30–32]. Work in this area suggests that prosodic modulation conveys more than 12 emotion categories as well as broader affective features, such as Valence and Arousal[33,34] and that these signals are to some degree understood by listeners from different cultures[32–36].

In this emerging science, what is not well-understood is how people recognize emotion in the voice. We therefore sought to analyse, first, how the variations in emotional prosody that people hear are mapped into the complex network of words and phrases that people (including scientists) rely on to represent emotion. Second, we investigated how many distinct emotions are recognized by people in the complex array of variations in prosody that they hear in their daily lives. Third, we analysed what drives the recognition of emotion, emotion categories (for example, Awe or Fear) or broader scales that capture core affect appraisals (Valence and Arousal). Finally, we investigated the structure of the emotions inferred from speech prosody, including whether they are discrete or bridged by gradients of meaning.

To investigate the above objectives, we examined how the cross-cultural recognition of prosodic modulations of the voice is explained by their organization within a semantic space of emotion recognition. A semantic space consists of the set of dimensions that capture how emotional states are perceived in relation to one another[7]. Such a space is characterized by three properties. The first is the conceptualization of emotional states in terms of emotion concepts and more general affective features, and how people use these concepts to represent emotions[37]. This property, a central focus in this investigation, informs theoretical claims about whether distinct emotion categories or affective features, such as Valence and Arousal, organize the recognition of emotions (see Supplementary Discussion 2 for definitions). The second is the dimensionality of the semantic space—or the number of independent directions in the space—the study of which yields answers to questions about the number of distinct emotions that can be signalled by expressive behaviour. And the third is the distribution of emotional states along these dimensions, which is germane to questions concerning the nature of the boundaries between emotion categories (for example, are they discrete or not).

[1]Department of Psychology, University of California, Berkeley, Berkeley, CA, USA. [2]Department of Psychology, Stockholm University, Stockholm, Sweden. [3]Olin School of Business, Washington University, Saint Louis, MO, USA. [4]Department of Statistics, University of California, Berkeley, Berkeley, CA, USA. *e-mail: alan.cowen@berkeley.edu

To capture a semantic space of any modality of emotion, empirical work should be guided by several principles. First, it is critical to study a vast array of stimuli to allow for the emergence a full dimensionality of that space, which potentially might include dozens of distinct emotions increasingly of interest in the field[12,14,33,38]. Most studies of emotional expression have focused on a narrow array of emotions, most typically six (Anger, Disgust, Fear, Sadness, Surprise and Happiness). Second, large-scale stimulus collection approaches should more reliably capture natural, within-category variation in how each emotion can be elicited or expressed[7,39]; this is in contrast to a traditional focus in the literature on the recognition of prototypical expressions or visual morphs between them[12,40–42]. A focus simply on emotion prototypes risks overestimating the degree of discreteness of emotion categories. Third, to capture the conceptualization of emotion, it is important to gather independent ratings in terms of emotion categories and affective features of the behaviour of interest—such as experience or expression. In doing so, studies need to move beyond foundational work that suggested emotions may be organized within a space[43] that is defined by its two to three broadest dimensions to include the affective features of cognitive appraisal theory[34,44,45] and componential theorizing[46]. Such theories describe how affective features other than Valence and Arousal are needed to account for the wide array of emotions that are studied today. Fourth, multidimensional reliability analysis techniques (techniques that extract dimensions on the basis of reliability across raters, rather than, for example, variance) can be used to investigate the extent to which judgments of distinct expressions can reliably be mapped into a high-dimensional space[7] (see also Supplementary Discussion 3). This large-scale statistical inference approach contrasts with the reliance of typical emotion-recognition research on either univariate recognition accuracy (reliability)[12,13,16,21,42] or factor analysis[44,47,48]. With these methodological advances, researchers can document how many distinct varieties of emotion are recognized and how these different varieties of emotion may simultaneously be organized by affective features, emotion categories and gradients of relatedness between emotion categories across different cultures (see Supplementary Discussion 4 for further details regarding methodological limitations of past studies).

A recent analysis of the semantic space of reported emotional experience[7] validated these methodological approaches, suggesting that they can be extended to understand the semantic space of emotion recognition. In this previous study, participants reported emotional responses to over 2,000 videos in terms of a wide array of emotion categories and in terms of 14 affective features that were derived from appraisal and componential theories of emotion[43–49]. These responses were analysed to derive a semantic space of reported emotional experience[7]. This previous study documented that (1) at least 27 distinct dimensions, or what one might think of as distinct kinds of emotion, were reliably elicited by different videos; (2) categorical labels were more powerful organizers of self-reported experiences than reports along well-studied scales of affect such as Valence; and (3) reported experiences fell along gradients that blurred the boundaries between categories of emotion.

Here, with further large-scale statistical inference advances, we derive a semantic space of the recognition of emotional prosody. We do so based on judgements obtained from US and Indian participants of prosodically modulated, lexically identical speech samples produced by actors from five different cultures who imagined themselves to be in an array of emotional scenarios. Samples of vocal prosody produced in this fashion have been found to resemble the spontaneous emotional modulations that occur in roughly 2% of everyday speech[50–52] and as much as a quarter of speech in emotional contexts[53], differing modestly from naturalistic vocalizations in terms of their average perceptual and acoustic features[54–57]. By comparing how participants from India and the United States interpret speech samples that richly vary only in their prosodic features, we can ascertain how the meaning of emotional prosody may be preserved across two very distinct English-speaking cultures[58,59] within a shared semantic space, including the relative primacy of emotion categories and affective features.

In the study of semantic space of emotion recognition, a number of questions of central theoretical importance remain unanswered. First, it is important to analyse how emotional expressions are best conceptualized and to what extent they convey specific categories of emotion, such as Awe and Fear[40,60–63], and information about affective features, such as Valence and Arousal[43,44,46,64,65] (and whether one of these methods of conceptualizing expression can account for the other). Second, the number of distinct varieties of emotion that are conveyed by emotional expressions and map to separate semantic dimensions should be investigated. Third, we should distinguish whether emotional expressions occupy discrete clusters—families of states, such as Awe, Interest and Surprise[40,65–70]—or whether they are distributed along continuous gradients[7,39,41,43,49,71]. And finally, it remains unclear to what extent the aforementioned properties of emotional expressions are preserved across cultures[12–16,72]. To address these issues and derive a cross-cultural semantic space of the recognition of emotion from prosody, we collected judgments from participants in the United States and India of a stimulus set of emotional prosody that includes an extensive number of emotions and diverse cultural origins of the speakers, which are ideal for deriving a semantic space of emotion recognition for the reasons that we outlined above.

The 2,345 participants from the United States and India were recruited on Amazon Mechanical Turk. Each participant was asked to judge at least 30 randomly selected speech samples from the VENEC corpus of 2,519 speech samples[33,73]. The VENEC corpus consists of two sentences ("Let me tell you something" and "That's exactly what happened") that were spoken by 100 actors from five different English-speaking cultures (United States, India, Australia, Kenya and Singapore) in tones targeting 18 categories of emotion derived from past studies of emotion-related vocalization[33]. Participants judged the speech samples in one of two randomly assigned response formats. One group of participants was asked to select a term, from 30 emotion categories, that best matched the emotion expressed in each speech sample. The emotion categories used for judgment were derived from recent studies of emotion-related prosody and vocal bursts and included: Adoration*, Amusement*, Anger*, Awe, Confusion, Contempt*, Contentment, Desire*, Disappointment, Disgust*, Distress*, Ecstasy, Elation*, Embarrassment, Fear*, Guilt*, Interest*, Pain, Pride*, Realization, Relief*, Romantic love, Sadness*, Serenity*, Shame*, Surprise (negative)*, Surprise (positive)*, Sympathy, Triumph and Neutral (categories indicated with an asterisk parallel those targeted with scenarios presented during the recording of the speech samples (see Supplementary Note 1)).

A second group of participants rated each of the 30 different speech samples that they judged in terms of 23 different affective features. These features were taken from dimensional and componential theoretical accounts of the appraisal processes that have been proposed to underlie emotion recognition and experience[43,44,46,65,74,75], and included Abruptness, Adjustability, Approach, Arousal, Attention, Certainty, Commitment, Control, Dominance, Effort, Expectedness, Fairness, Goal relevancy, Identity, Improvement, Normativity to the agent, Normativity to society, Novelty, Obstruction, Probability, Safety, Urgency and Valence. (Note that we use these labels only as shorthand for the more colloquial, literature-derived questions to which raters actually responded. See Supplementary Note 2 and Supplementary Tables 1, 2 for specific wording of each of these appraisal dimensions and their sources in the theoretical literature.) Participants judged each speech sample on nine-point Likert scales (1 = negative levels

or none of the feature, 5 = neutral or moderate levels, 9 = extreme levels of the feature).

On the basis of past estimates of reliability in observer judgment[7], for each speech sample we collected 10–15 judgments from separate participants in each of the two response formats in each culture. Thus, we gathered a total of 1,270,736 individual judgments of all speech samples (75,461 forced-choice categorical judgments and 1,195,275 nine-point scale judgments; see also Supplementary Note 2). The categories used for judgment included the emotions the actors were instructed to target[33] as well as emotions that were previously found to be conveyed by the voice[22,76,77]. The collection of a wide range of judgments for a rich array of lexically identical speech samples allows us to apply large-scale statistical inference techniques to examine the conceptualization, dimensionality and distribution of emotion recognized in prosody across two cultures.

Results

Overview. Guided by our semantic space analysis of emotion recognition, past validated methods[7] and a central design feature of this investigation—data gathered from two cultures—our data analysis proceeds as follows. First, to explore issues of conceptualization, we examine what is better preserved across the two cultures, emotion categories or affective features, and which explains the variance in the other variable better across judgments of Indian and US participants. Then, to address the dimensionality of the space, we rely on statistical techniques that uncover how many distinct varieties of emotion are required to account for cross-cultural similarities in the recognition of emotional prosody. Finally, with recently developed visualization techniques, we explore the distribution of emotions signalled by prosody within a high-dimensional space. We conclude by looking at the acoustic correlates of the emotions conveyed in prosody and address whether these acoustic features better track emotion categories or affective features across the two cultures.

Verifying the recognition of emotion categories. Past studies have often relied on accuracy rates that are derived from whether the judgements of the participants matched the expectations of the experimenters to ascertain the cross-cultural similarity in emotion recognition[12,13,16,23,24,38]. For reasons that we outline in Supplementary Discussion 5, we operationalized emotion recognition in terms of interrater agreement, a more data-driven approach to observer consensus in emotion recognition. Guided by this conceptual approach, we first analyse the combined data from Indian and US participants to verify that we had obtained reliable judgments of emotion, in order to determine how many emotions were recognized in the 2,519 speech samples. In this analysis, we found that raters were able to recognize a wide variety of emotion categories with a moderate degree of reliability. Twenty-two different emotion categories were recognized with significant interrater agreement from at least one speech sample ($q < 0.05$, Monte Carlo simulation using empirical base rates; see Supplementary Fig. 1 for the distribution of interrater agreement rates and Supplementary Note 3). Out of the 2,519 speech samples, 56% elicited significant rates of interrater agreement for at least one category. On average, 25% of raters chose the most agreed-upon category of emotion for each speech sample (chance level = 14%, Monte Carlo simulation of all category judgments matching the same overall proportions of categories that were selected by the real participants). These levels of interrater agreement are comparable to those documented in previous studies of emotion prosody[33]. Although interrater agreement rates varied across the different speech samples, highly recognized examples were found for a number of emotion categories. For instance, five emotion categories—Amusement, Anger, Desire, Fear and Sadness—were recognized in some speech samples by more than half of raters. Another five emotion categories—Adoration, Confusion, Distress, Pain and Relief—were recognized in some speech samples by at least 1/3 of raters.

The preservation of emotion categories and broader affective features across cultures. Next, we compared how emotions were recognized from prosody across cultures. In doing so, we sought to ascertain whether judgments of emotion categories or affective features were better preserved across two cultures in the recognition of emotion—to address whether the conceptualization of emotional prosody is driven more by categories or affective features. In past studies, cross-cultural similarity has typically been ascertained by comparing the rates with which members of different cultures label expressive behaviours with the same emotion terms[12,16,23,24,38]. This approach does not capture how members of different cultures also use emotion concepts (either emotion categories or affective features) to label non-target expressions in a similar manner, data critical to understanding cultural similarities in how individuals recognize emotion in expressive behaviour. Given this concern, for each emotion category and affective feature, we correlated the mean judgments of US raters with those of Indian raters across all 2,519 samples of emotional prosody. This analysis reveals the extent to which US and Indian participants use the emotion categories and affective features in a similar manner when labelling the meaning of the 2,519 samples of emotional prosody. To control for error, or noise, in the use of the emotion categories and scales of affect, we then divided this value by the within-culture explainable variances[78] of these mean judgments (see Supplementary Notes 4, 5 for further rationale and details). Dividing by the explainable variances results in an estimate of what the correlation would be if we averaged an infinite number of ratings in each culture. We refer to this estimate as a signal correlation: it captures the degree of similarity between cultures in the recognition of each emotion category and scale of affect from prosody while correcting for the sampling error arising from inconsistent judgments within each culture (see Supplementary Fig. 2 for a demonstration that these methods are effective using simulated data). The cross-cultural signal correlations in emotion judgments of each emotion category and affective feature are shown in Fig. 1.

If categories of emotion (for example, Amusement) are psychologically constructed from more basic appraisals of core affect (Valence and Arousal), one would expect the recognition of emotion in prosody along scales such as Valence and Arousal to be better preserved across cultures than the recognition of emotion categories. That is, there should be greater convergence across cultures in how affective features are recognized in emotional prosody than emotion categories, which are presumably constructed out of more basic affective appraisals[43,60,79]. As shown in Fig. 1, our results diverge from this prediction. We find that the recognition of a number of emotion categories from prosody is better preserved across cultures than that of any of the 23 affective scales that we considered, including Valence and Arousal. With cross-cultural signal correlations (r) exceeding 0.7, the recognition of Adoration, Amusement, Anger, Awe, Contentment, Desire, Fear, Interest, Pain, Realization, Romantic love, Sadness, Surprise (negative) and Surprise (positive) was better preserved across India and the United States than that of Valence ($r = 0.67$; 90% confidence interval: $0.55 < r < 0.78$). In addition, cross-cultural signal correlations were significantly greater for Anger ($r = 0.94$; $0.83 < r < 1$; $P = 0.001$, two-tailed bootstrap test; $q < 0.05$, Benjamini–Hochberg false-discovery rate (FDR) correction[80]; see Supplementary Note 5) than for Valence. Furthermore, many emotion categories were significantly better preserved across cultures than the recognition of Arousal ($r = 0.39$; $0.20 < r < 0.58$), including Amusement, Anger, Contentment, Desire, Fear, Pain, Sadness and Surprise (positive) ($r = 0.92$, 0.94, 0.94, 0.87, 0.88, 0.77, 0.96, and 0.90; 90% confidence intervals: 0.72, 0.83, 0.63, 0.63, 0.71, 0.62, 0.75, $0.61 < r < 1$, 1, 1, 1, 1, 0.90, 1, 1; all $P < 0.01$, $q < 0.05$,

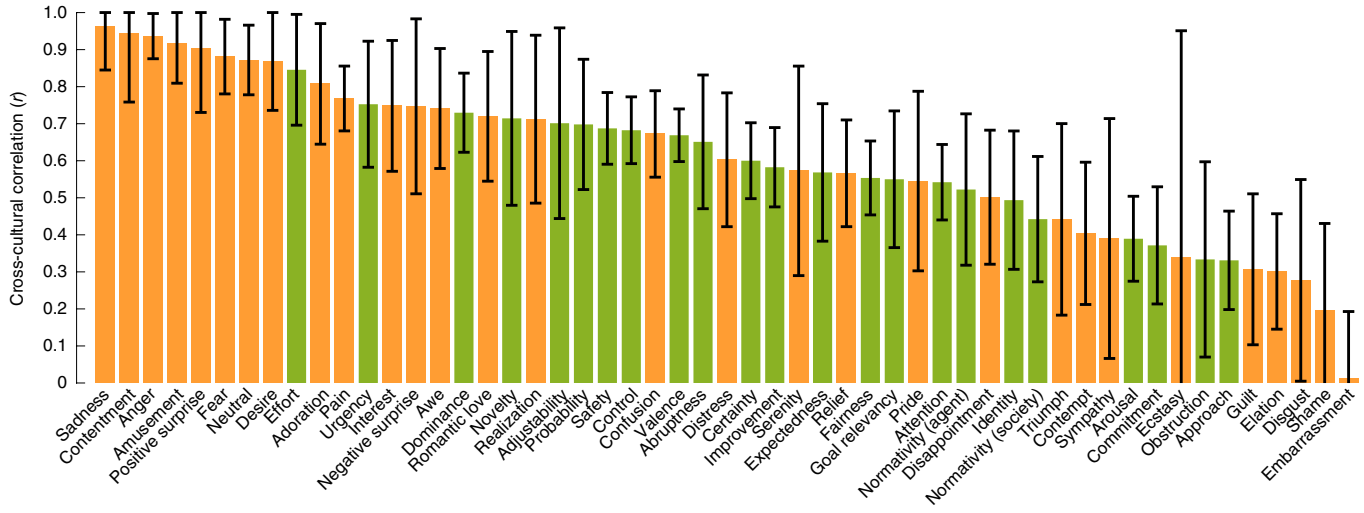

Fig. 1 | Correlations in the meaning of emotional prosody across cultures. The cross-cultural signal correlation (*r*) for each category (orange bars) and affective scale (green bars) captures the degree to which each judgment is preserved across India and the United States across all 2,519 speech samples. It is found by correlating the mean responses by Indian participants with the mean responses by US participants across the 2,519 speech samples, then dividing by the explainable variance[78] in responses from each culture. Error bars represent the standard error estimated by bootstrapping across raters. For category surveys, participant sample size, $n_{US} = 525$, $n_{India} = 152$, and for the two affective scale surveys, $n_{US} = 927$ and 827 and $n_{India} = 242$ and 205. See Supplementary Notes 4 and 5 for details regarding the explainable variance and standard error estimations, Supplementary Fig. 2 for confirmation that these results accurately recover population-level correlations and Supplementary Fig. 3 for similar results using Spearman correlations and/or binary affective scale ratings.

two-tailed bootstrap test). The finding that the recognition of certain emotion categories is better preserved across cultures than that of Valence—which is considered to be abasic building block of emotional life[79]—or Arousal, the other putative component of core affect, contrasts with the claim that the recognition of emotion categories derives from the recognition of such affective features[43,60,79,81].

The primacy of the recognition of emotion categories over affective features. That several emotion categories were more robustly recognized across cultures than Valence or Arousal raises an intriguing question about the conceptualization of emotion in prosody: perhaps affective features, such as Valence and Arousal, are psychologically constructed from categories of emotion. In other words, perhaps emotion recognition from prosody involves the immediate recognition of an emotion category (or categories), and then levels of Valence, Arousal and other affective features are inferred from these more primary categorical judgments. If so, the additional stage of inference involved in affective feature judgments might introduce additional cultural variation, which should be reflected in their lower levels of cross-cultural similarity—which we have observed—and their reduced power in predicting categorical judgments across cultures. Given this reasoning, one might expect that category judgments, rather than affective scale judgments, from one culture are better predictors of affective scale judgments from another culture. Gathering separate judgments of a full complement of emotion categories and affective scales across 2,519 samples of emotional prosody enabled a rigorous test of this possibility.

To test the hypotheses concerning the primacy of emotion categories and affective features in emotion recognition, we used linear regression analyses to derive cross-cultural signal correlations in the mapping between category and affective scale judgments of prosody (Supplementary Note 6). These analyses ascertain whether emotion category ratings are stronger predictors of affective feature judgments across cultures, or vice versa (Fig. 2). Critical to the present question of the primacy of categories or affective features in emotion recognition are certain linkages denoted by letters in Fig. 2. Notably, emotion categories are better preserved than judgments of affective features across India and the United States (C > D; correlation of 0.80 compared to 0.59, $P = 0.041$, two-tailed bootstrap test; $C - D = 0.21$, 90% confidence interval: $0.041 < C - D < 0.39$; see Supplementary Note 6 for details). In keeping with the idea that the interpretation of prosody in terms of affective features derives from the shared recognition of emotion categories, we find that (1) Indian category judgments predicted US affective scale judgments far better than Indian affective scale judgments predicted US affective scale judgments (D1 > D; correlation of 0.80 compared to 0.59, $P = 0.012$, two-tailed bootstrap test; $D1 - D = 0.21$, 90% confidence interval: $0.06 < D1 - D < 0.35$), and that (2) US category judgments also predicted Indian affective scale judgments nominally better than US affective scale judgments predicted Indian affective scale judgments (D2 > D; correlation of 0.65 compared to 0.59, $P = 0.12$, two-tailed bootstrap test; $D2 - D = 0.050$, 90% confidence interval: $-0.01 < D2 - D < 0.12$). In addition, there is little preserved variation across the two cultures in affective scale judgments that is independent of the category judgments (see term R in Fig. 2), suggesting little cross-cultural similarity in how Valence, Arousal and other affect features operate in the recognition of emotion from prosody once emotion category judgments are accounted for. Additionally, as one might expect, the affective scale judgments from each country do a poor job of predicting the category judgments from the other (see Supplementary Fig. 4). On the basis of these results, it is more plausible that judgments of general affective features (such as Valence and Arousal) are psychologically constructed from the recognition of emotion categories (such as Amusement and Fear) than vice versa, at least during the recognition of emotion in prosody.

The number of distinct varieties of emotion recognized in both cultures. Thus far, we have shown that at least 22 emotion categories were recognized at above-chance accuracy in at least one speech sample and that the cross-cultural recognition of emotion from prosody was better represented by these categories than scales of affect. We next investigated how many distinct varieties of emotion

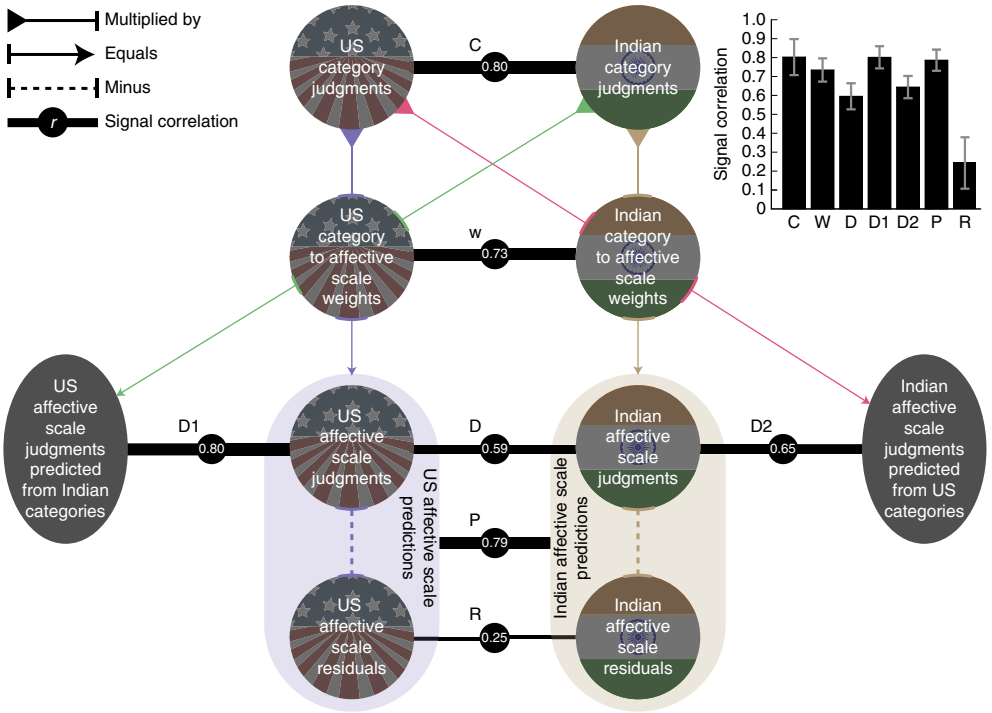

Fig. 2 | The preserved recognition of emotion categories accounts for the preservation of affective feature judgments across cultures. Each circle (or ellipse) represents a matrix of estimators relevant to the recognition of emotion attributes—categories or affective features—in the United States (on the left) and India (on the right). In the first row, for example, US mean category judgments are compared to Indian mean category judgments. Relationships between circles are described using the symbols indicated in the top left key. For example, US category judgments are multiplied by a set of category-to-affective-scale weights, estimated using ordinary least-squares regression, to predict US affective scale judgments. Signal correlations between the Indian and US matrices of emotion judgment data are given in the small black circles, signified by adjacent letters, and plotted in the top right graph. Category judgments are significantly better preserved than affective scale judgments (C > D, $P = 0.041$, two-tailed bootstrap test; C − D = 0.21, 90% confidence interval: 0.041 < C − D < 0.39). Moreover, category judgments are better than affective scale judgments at predicting affective scale judgments from another culture (D1 > D, $P = 0.012$, two-tailed bootstrap test; D2 > D, $P = 0.12$; D1 − D = 0.21, 90% confidence interval: 0.06 < D1 − D < 0.35; D2 − D = 0.050, 90% confidence interval: −0.01 < D2 − D < 0.12). The variance in the affective scale judgments left over after removing the predictions of the categories is less correlated across cultures ($R = 0.25$ between residuals, 90% confidence interval: 0.024 < R < 0.47). Taken together, these results are consistent with the hypothesis that categories of emotion are recognized from prosody, and then subsequently used to construct affective scale judgments in a more culture-specific process of inference. All signal correlations have been divided by the explainable variance[78] for each matrix (see Supplementary Note 6). Error bars in the top right plot represent the standard error estimated by bootstrapping across raters (for category surveys, $n_{US} = 525$, $n_{India} = 152$, and for the two affective scale surveys, $n_{USA} = 927$ and 827 and $n_{India} = 242$ and 205). Also note that due to limitations in model fitting, estimates D1, D2 and P are biased downward, such that the preservation of dimensional judgments is probably even better explained by the preservation of categorical judgments than indicated here.

were preserved in the recognition of emotional prosody across the two cultures—that is, what the dimensionality of the cross-cultural recognition of emotion from prosody is. More specifically, we have not yet ruled out whether some of the categories were redundant (for example, synonyms). There may have been interrater reliability in labelling emotional prosody with categories such as Awe and Fear, but perhaps these categories ultimately capture the same kinds of emotional prosody. If we reduce the US and Indian judgments of prosody to a more limited number of statistically independent dimensions, we wondered what the minimum number of dimensions that is necessary to account for commonalities in the recognition of emotion across cultures is.

To compute the total number of distinct varieties of emotion that were significantly preserved across the US and Indian emotion judgments of the speech samples, we introduce a principal preserved component analysis (PPCA) method. PPCA extracts linear combinations of attributes (here, emotion judgments) that maximally covary across two datasets that measure the same attributes (US and Indian judgment data). The resulting components are ordered in terms of their level of positive covariance across the two datasets (cultures). More technically, PPCA maximizes the objective function $\mathrm{Cov}(X\alpha_i, Y\alpha_i)$. It shares features of partial least-squares correlation analysis (PLSC)[82], canonical correlation analysis (CCA)[83] and PCA. Like PLSC and CCA, PPCA examines the cross-covariance between datasets rather than the variance–covariance matrix within a single dataset. However, whereas PLSC and CCA derive two sets of latent variables, α and β, maximizing $\mathrm{Cov}(X\alpha_i, Y\beta_i]$ or $\mathrm{Corr}[X\alpha_i, Y\beta_i]$, PPCA derives only one variable: α. The goal here is to find dimensions of recognition that are common to both datasets X and Y. Our method reduces to principal component analysis (PCA) when the two datasets to which it is applied are identical, so that the objective becomes $\mathrm{Cov}(X\alpha_i, X\alpha_i) = \mathrm{Var}(X\alpha_i)$, but PCA (and related factor-analytic methods) capture the variance within a dataset rather the covariance across datasets. Note that we also apply a second kind of PPCA—correlational PPCA—which performs a whitening transformation within each dataset and then derives a latent set of variables α that maximizes the correlation $\mathrm{Corr}(X_{wh}\alpha_i, Y_{wh}\alpha_i)$ rather than the covariance. See Supplementary Note 7 for further details and discussion.

Given that we had previously found the categories that explained the cross-cultural preservation of affective scales, we applied PPCA to the US and Indian category judgments of the 30 emotions

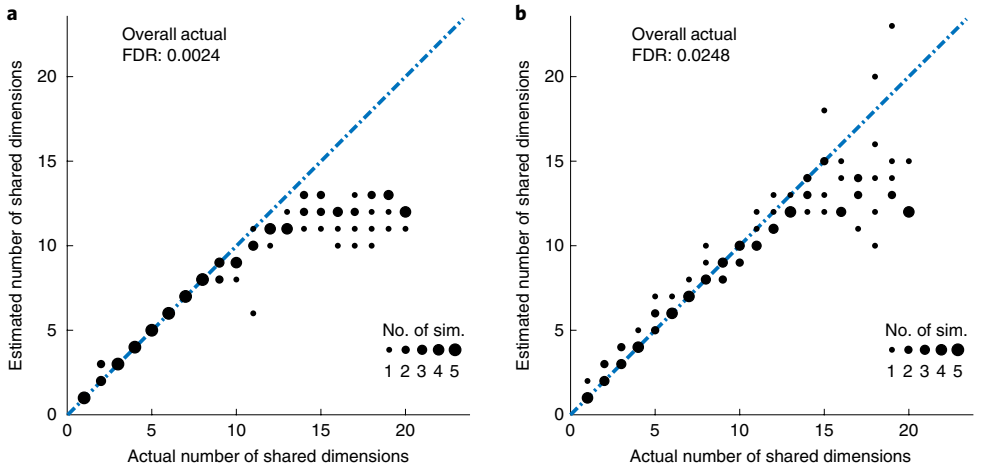

Fig. 3 | Verifying that PPCA accurately estimates the number of shared dimensions. To test whether a leave-one-rater-out PPCA would accurately estimate the number of preserved dimensions of emotion recognition across cultures, we ran Monte Carlo simulations of our experiment in which the ratings were drawn from distributions that varied systematically in their underlying dimensionality. In 100 separate simulations, five each for dimensionalities between 1 and 20, category ratings of each speech sample in each culture were drawn at random from multinomial distributions parameterized by $\widetilde{X}$, the 2,519 × 30 probabilities of selecting each category for each speech sample, and **N**, the number of raters who rated each speech sample in each country. $\widetilde{X}$ was computed by applying PPCA to the proportion of times each category was actually selected for each speech sample in each culture, $\mathbf{X}_{US}$ and $\mathbf{X}_{India}$, projecting $\mathbf{X}_{US}$ onto the first 1–20 dimensions extracted by PPCA, back-projecting these scores into the space of categories ($\alpha_{sim}{}^{T}\mathbf{X}_{USA}\alpha_{sim}$), and normalizing the result by subtracting the minimum from each row and dividing by the sum of each row. This resulted in 1–20 preserved dimensions in each simulation, each repeated five times. We used the same $\widetilde{X}$ for both cultures to maximize the similarity in ratings. **N** was set to the number of ratings actually obtained of each speech sample in each culture. Each rating was randomly assigned to a 'rater' with a probability given by the percentage of ratings each rater actually contributed. Finally, we used a leave-one-rater-out PPCA to determine the *P* values for extracted dimensions (Supplementary Note 7). **a**,**b**, The actual numbers of preserved dimensions used to generated the data in each simulation (*x* axis) versus the estimated number of dimensions (*y* axis) using two criteria (**a**: $q < 0.001$ across all held-out raters; **b**: $q < 0.05$ across held-out raters from each country individually; ForwardStop sequential FDR-corrected[84] one-tailed Wilcoxon signed-rank test[83]). We can see that leave-one-rater-out PPCA accurately estimates the number of preserved dimensions and generates conservative *q* values. Note that there was less power to detect later dimensions given that these carried less covariance; dimensions 21+ carried negative covariance (see Fig. 4) and are therefore excluded.

conveyed by the 2,519 speech samples to determine the number of independent dimensions—or kinds—of emotion that are recognized in both cultures. We applied PPCA in a leave-one-rater-out manner to determine the statistical significance of each component. More technically, we iteratively applied PPCA to extract components from the judgments of all but one of the raters, projected the held-out rater's ratings onto the components and assessed the partial Pearson correlation between the component scores derived from each held-out rater's ratings and those derived from the mean ratings from the other culture, partialing out each previous component. We then tested whether these held-out, statistically independent correlation values were consistently positive for each component using a non-parametric Wilcoxon signed-rank test[84]. We excluded the 'Neutral' category from PPCA to avoid matrix degeneracy, resulting in dimensions that can be conceived as variations from neutrality. It was not a guarantee that these methods would be effective for sparse data of the kind that we analysed here. Hence, we established using simulations that they would produce accurate results given the distribution of responses that we observed. See Fig. 3 for results of simulations.

As we show in Fig. 4, PPCA revealed that 12 distinct semantic dimensions of emotion were recognized in prosody and that these were significantly preserved across the US and Indian participant judgments (Fig. 4b; out-of-sample $r \geq 0.066$, $q < 0.001$ across all held-out raters, $q < 0.05$ across held-out raters from each country individually, ForwardStop sequential FDR-corrected[84] one-tailed Wilcoxon signed-rank test[84]—one-tailed because we are only interested in positive cross-cultural correlations). We find that the pattern of tune, rhythm and timbre in speech conveys 12 distinct varieties of emotion across the two cultures. In Fig. 4, the top left plot shows the proportion of variance explained by each dimension (principal preserved components (PPCs)) uncovered by PPCA in data from each culture; the bottom left plot shows the proportion of preserved covariance for each dimension, as well as the corresponding correlation and its significance.

Of note, the application of CCA[83] also resulted in 12 significant dimensions (out-of-sample $r \geq 0.049$, $q < 0.05$ across all held-out raters, $q < 0.05$ across held-out raters from each country individually, ForwardStop sequential FDR-corrected[84] one-tailed Wilcoxon signed-rank test[84]; see Supplementary Note 7), as did the application of correlational PPCA (out-of-sample $r \geq 0.041$, $q < 0.001$ across all held-out raters, $q < 0.05$ across held-out raters from each country individually, ForwardStop sequential FDR-corrected[84] one-tailed Wilcoxon signed-rank test[84]). Each method resulted in a latent variable solution for the first 8–12 dimensions relatively similar to that obtained using PPCA, as demonstrated in Fig. 5 and Supplementary Figs. 5–7. Differences beyond around the eight dimension emerge when using PCA methods that were not designed to extract preserved components.

The preserved categories of emotional prosody. To find the 12 patterns (dimensions) of emotion recognition within the categorical judgments that were preserved across participants from the United States and India, we applied factor rotation (varimax) to the 12 significant components extracted using PPCA. Here, factor rotation extracts a simplified representation of the space by attempting to find dimensions constituted of only a few categories each, if possible. After factor rotation, we find that each of the 12 resulting dimensions (PPCs) loaded maximally on a distinct category (see Fig. 4c). These categories include aDoration, Amusement, Anger,

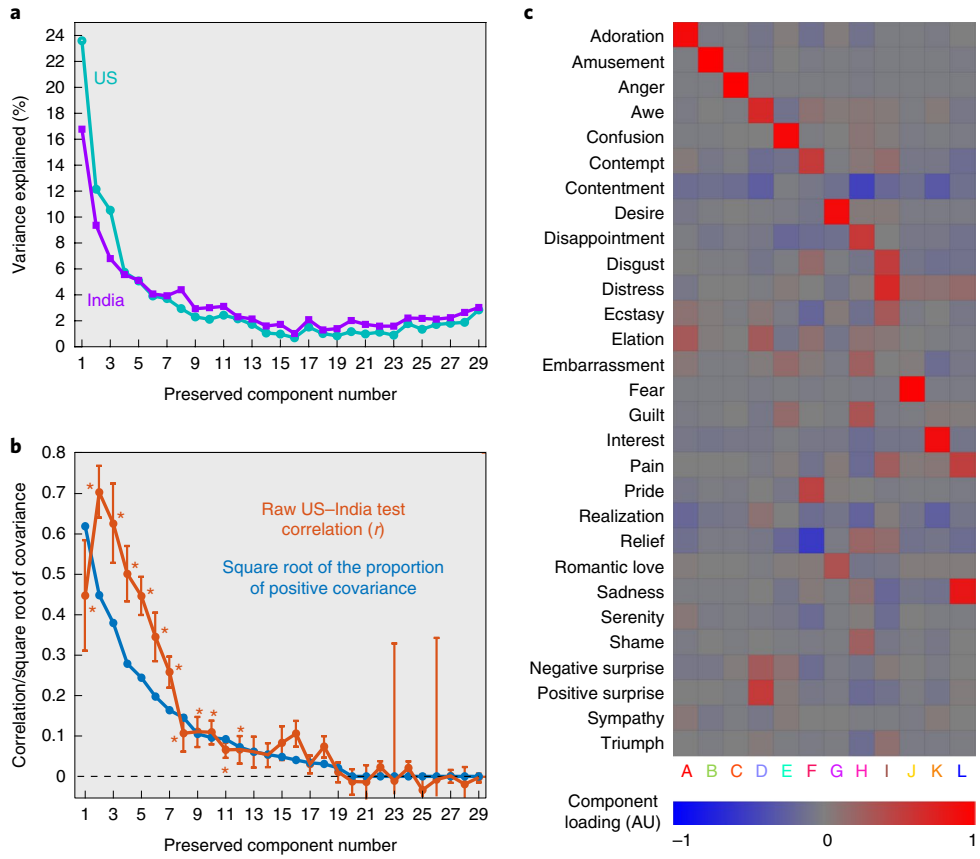

Fig. 4 | 12 distinct varieties of emotional prosody are preserved across cultures via category recognition. a, The in-sample proportion of variance explained within US and Indian ratings by the 29 principal preserved components (PPCs) of the mean categorical ratings of 30 emotions across cultures. **b**, The square root of the in-sample covariance of each PPC across cultures, scaled by total positive covariance, is plotted alongside the out-of-sample cross-cultural correlation derived from a cross-validation analysis (see Supplementary Note 7). The test correlation was significant for 12 PPCs (out-of-sample $r \geq 0.066$, $q < 0.001$ across all held-out raters, $q < 0.05$ across held-out raters from each country individually, ForwardStop sequential FDR-corrected[95] one-tailed Wilcoxon signed-rank test[84]). Error bars represent the standard error (participant sample size, $n_{USA} = 525$, $n_{India} = 152$). Note that these correlations are not adjusted for explainable variance, so it is safe to assume that the corresponding population-level correlations are substantially higher. **c**, The 12 distinct varieties of emotional prosody that are preserved across cultures correspond to 12 categories of emotion—Adoration, Amusement, Anger, Awe, Confusion, Contempt, Desire, Disappointment, Distress, Fear, Interest and Sadness. By applying factor rotation (varimax) to the 12 significant PPCs, we find 12 preserved varieties of emotional prosody that each load maximally on a distinct emotion category. Thus, we will refer to each component, labeled A–L on the *x* axis, with a single emotion category (not to be confused with the raw categorical judgments). AU, arbitrary units.

Awe, Confusion, Contempt, Desire, Disappointment, Distress, Fear, Interest and Sadness. We can infer that these 12 categories correspond to distinct prosodic modulations of speech that are preserved in India and the United States. Note that some dimensions involve multiple categories, such as Awe and Surprise (dimension D), indicating that they were used similarly across cultures to label speech samples. These findings replicate the conclusions of previous studies that several emotions can be conveyed across cultures with prosody (Anger, Contempt, Fear, Interest, Desire, Relief and Sadness[33–36]), but also reveal other emotions—Adoration, Amusement, Awe, Confusion, Disappointment and Distress—that can be reliably communicated with prosody. (It is also worth noting that three of the categories—Awe, Confusion and Disappointment—were not among those targeted with scenarios during the recording of the speech samples (see Supplementary Note 1), illustrating that the induction procedure was compatible with rich variation in emotional responses.)

The distribution of categories of emotional prosody within a semantic space. Having thus far examined the dimensionality and conceptualization of the semantic space of emotion recognition in prosody, we now ask how these categories of emotion recognized from prosody are distributed within a semantic space. We also asked whether they are found within discrete clusters, as predicted by basic emotion theories[40,65–70], or along continuous gradients between emotion categories, as recently documented in our investigation of reported emotional experience[7]. We find that the emotional states conveyed by prosody lie along continuous gradients between categories rather than within discrete clusters (as with emotional experience) (Fig. 6). These gradients between categories are evident when we visualize smooth variations in the categorical judgment profiles of the 2,519 speech samples using a method called *t*-distributed stochastic neighbour embedding(*t*-SNE)[85]. *t*-SNE projects data into a two-dimensional space that largely preserves the local distances between data points. A limitation of *t*-SNE is that it will generate a different result each time it is run. We therefore ran *t*-SNE analyses 100 times, identified the instance that resulted in the lowest loss of information (Kullback–Leibler divergence) and fine-tuned this map using more iterations of *t*-SNE (see Supplementary Note 8 for further details). In Fig. 4, *t*-SNE is used to visualize the smooth gradients between emotion categories—represented in different colours—and the extent to which they are

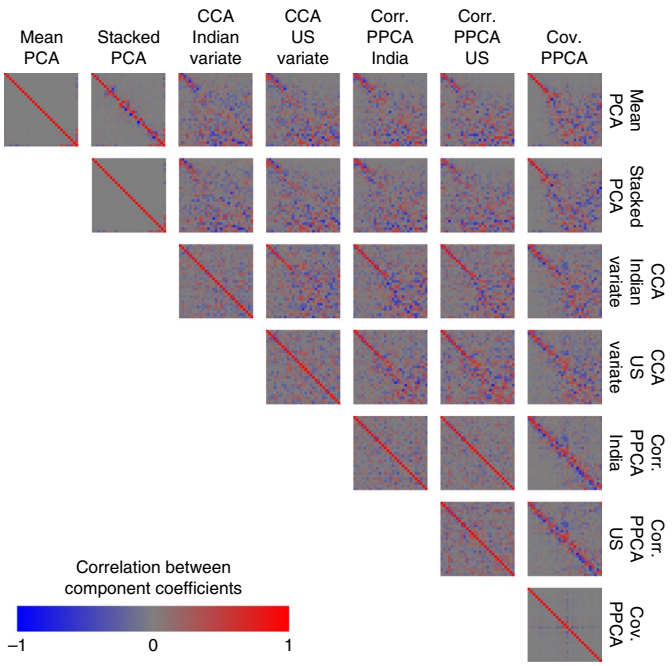

Fig. 5 | Correlations between coefficients of components extracted from US and Indian category judgments using different methods. Each method was used to extract 29 components after excluding the 'Neutral' category. The actual component coefficients of each category are shown in Supplementary Fig. 5. Here, each pixel in each plot represents a correlation between coefficients of components derived using two different methods. The components are ordered in a matter appropriate for each method: in terms of descending explained variance for stacked and mean PCA, in terms of descending canonical correlation for CCA, and in terms of descending correlation and covariance for correlational and covariational PPCAs, respectively. Note that the CCA extracts an entirely separate latent space for each culture and the correlational PPCA extracts a slightly modified latent space for each culture ($[X^TX]^{-1/2}\alpha_1$ and $[Y^TY]^{-1/2}\alpha_1$). In general, early components (the first seven to ten) are similar across the different methods. The solution derived using the covariational PPCA, our primary focus, shares similarities with those derived by both the PCA and CCA, whereas the correlational PPCA solution was more similar to the CCA solution. See also Supplementary Fig. 6 for US–India correlations between the scores of categorical judgments projected onto the components derived using each method.

preserved across cultures. To allow further exploration of the categories of emotion signalled by prosody and the smooth gradients between them, we also provide an online, interactive version of Fig. 6 in which each speech sample can be played while viewing its categorical and affective scale ratings: https://s3-us-west-1.amazonaws.com/venec/map.html.

To verify that the smooth gradients between categories correspond to smooth differences in emotional meaning, we determined whether judgments of the affective scales, such as Valence and Arousal, also varied smoothly along these gradients. First, we ascertained whether the raw proportions of category judgments assigned to each speech sample were more predictive of its affective scale judgments than its discrete, modal category assignment alone. We found that this was indeed the case, with the 12 dimensions (PPCs) predicting 86% of the variance in the affective scales (90% confidence interval: $0.82 < r^2 < 0.89$), a fully discrete model (with 12 indicator variables denoting the maximal dimension that each speech sample loaded on—that is, one variety of emotion at a time) predicting 68% of the variance in the affective scales (90% confidence interval: $0.65 < r^2 < 0.71$) and a discrete model with intensity (keeping only the top non-zero score per speech sample) predicting 76% of the variance in the affective scales (90% confidence interval: $0.72 < r^2 < 0.79$). Both discrete models performed significantly worse than the full 12-dimensional model ($P < 0.001$, two-tailed bootstrap test; see Supplementary Note 9 for details). This result confirms that the affective scales vary along category gradients rather than just being a function of the most recognized category. Furthermore, to ascertain whether these results could be explained by correlations in perceptual ambiguity across the category and affective scale judgments (for example, some subjects perceiving a sample as Awe and others perceiving it as Adoration), we correlated the mean standard deviation across participants of the category judgments with that of the affective scale judgments of each speech sample, finding that they were slightly inversely correlated (Pearson's $r = -0.21$, $P < 0.001$, two-tailed bootstrap test, 90% confidence interval: $-0.25 < r < -0.18$; Spearman's $\rho = -0.16$, $P < 0.001$, two-tailed bootstrap test, 90% confidence interval: $-0.19 < \rho < -0.12$). Hence the smooth gradients between categories most likely cannot be explained by ambiguity of recognition—rather, they point to intermediate blends of emotion categories traditionally thought of as discrete.

Figure 7 presents the number of prosody samples that loaded significantly on each of the 12 dimensions that we uncovered and on each combination of two dimensions. This analysis shows the varieties of emotion that can be blended together in prosodic modulations of a single sentence and suggests that the gradients tend to bridge distinct, conceptually related emotional states. Careful inspection of Fig. 7 reveals, for example, that prosodic modulations traverse gradients from Fear to Sadness, from Amusement to Adoration, from Confusion to Interest, and from Anger to Contempt. But these gradients were specific to particular category pairs; for example, Sadness overlapped heavily with Fear (51 speech samples) but did not overlap at all with Desire. Thus, the emotions conveyed by prosody are neither entirely discrete nor entirely independent, but are rather distributed along continuous gradients between particular pairs of emotion categories.

The preservation of acoustic correlates of emotion recognition across cultures. Theorists have long claimed that certain acoustic features drive the recognition of emotion in prosody[30,32,33,86]. Within these theories, emotion recognition from vocalization is posited to rely on lower-level processing of acoustic signals, which undergo complex, multistage neural processing to yield appraisal feature and categorical judgments such as those that we have considered thus far[10,11,46]. Past work has examined how broad, lower-level acoustic properties (for example, fundamental frequency) are associated with emotion judgments[30,32,33,87]. We therefore wondered to what extent associations between acoustic features of prosody and emotion category and appraisal feature judgments are preserved across cultures. Answers to this question trace the preserved recognition of emotion categories and affective features across two cultures to a more basic level of auditory processing, central to the analysis of the mechanisms of emotion recognition from prosody.

Broad acoustic properties, such as the fundamental frequency (F0—the lowest and loudest frequency of sound, corresponding to perceived pitch), spectral centroid (the centre of the frequency spectrum, corresponding to perceived 'brightness'), pitch saliency (corresponding to the perceived tonality and sound) and rate of speech, are known to correlate with the recognition of emotional and affective features[30,32,33,56]. To interrogate the emotional correlates of these acoustic properties in each culture, we computed them for the 2,519 speech samples, correlated them with our 12 PPCs as well as the raw category and affective scale judgments in each culture and measured the extent to which these associations were preserved across cultures (see Supplementary Note 11 for details).

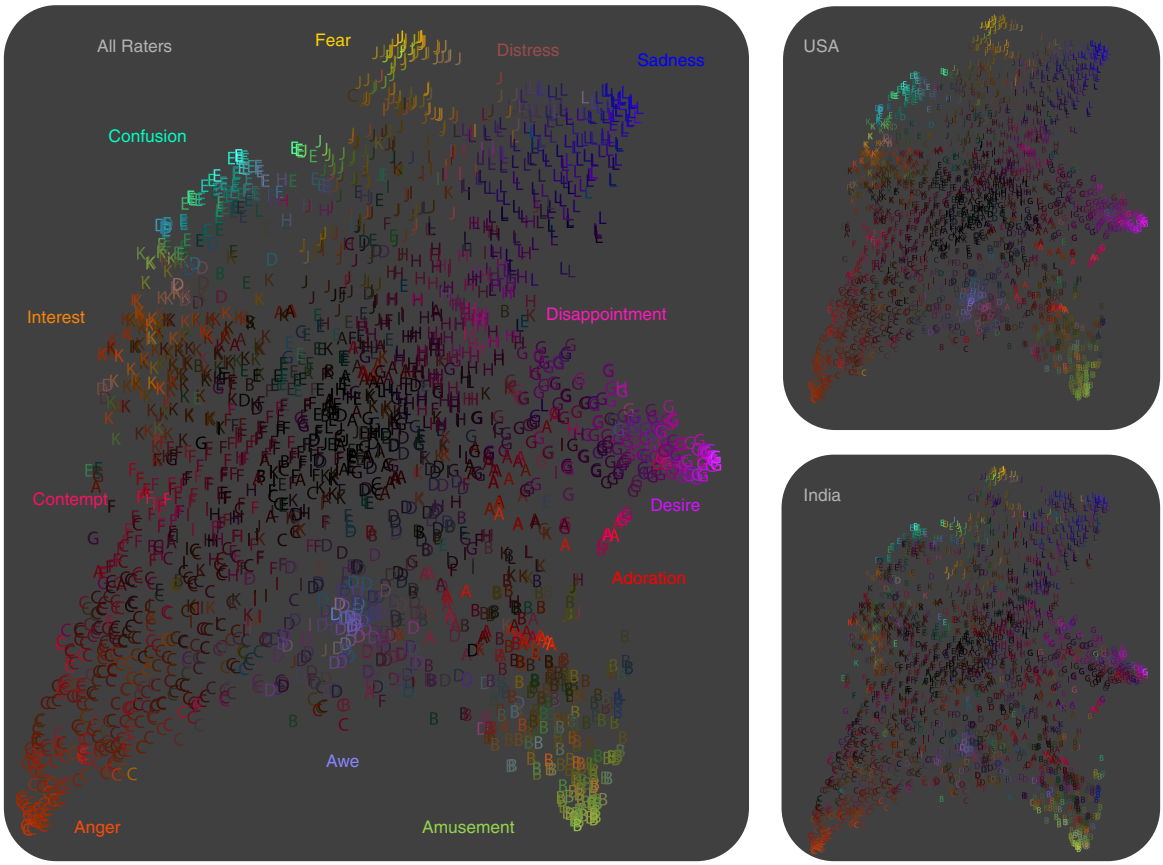

Fig. 6 | Visualizing the 12-dimensional structure of emotion conveyed by prosody. To visualize the categories of emotion conveyed by prosody, maps were generated of average emotional categorical judgments of the 2,519 speech samples within a 12-dimensional categorical space of recognized emotion. *t*-SNE—a data visualization method that accurately preserves local distances between data points while separating more distinct data points by longer, more approximate distances—was applied to the concatenated US and Indian scores of the 2,519 speech samples on the 12 categorical judgment dimensions, generating coordinates of each speech sample on two axes (this does not mean that the data is in fact two-dimensional; see Supplementary Discussion 6). The individual speech samples are plotted along these axes as letters (A–L, as assigned on the *x* axis of Fig. 4c) that correspond to their highest-loading categorical judgment dimension (with ties broken alphabetically) and are coloured using a weighted average of colours corresponding to their scores on the 12 categorical judgment dimensions (see Supplementary Note 8). The resulting map reveals gradients from Amusement to Adoration, Anger to Contempt, and so on. For an interactive version of this map, see https://s3-us-west-1.amazonaws.com/venec/map.html. The smaller maps, coloured using projections of the mean US or Indian judgments alone onto the same 12 dimensions, demonstrate that the recognition of categories and smooth gradients between them is largely preserved across the two cultures.

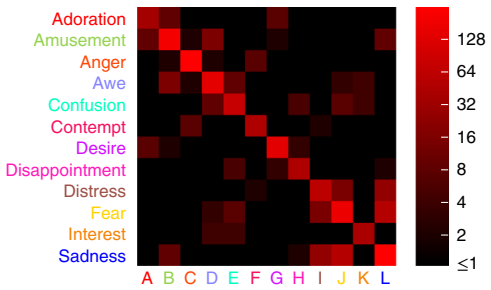

Fig. 7 | The 12 distinct categories can be blended together in a number of ways. Represented here are the number of speech samples that loaded significantly on each dimension, or kind, of emotion (diagonal) and on pairwise combinations of dimensions ($q < 0.05$, Monte Carlo simulation using rates of each category judgment, Benjamini–Hochberg FDR-corrected analysis[80]; see Supplementary Note 10 for details). Categories are often blended together, combining Adoration with Amusement, Anger with Contempt, Awe with Interest, Sadness with Fear.

The top row of Fig. 8 shows the correlations between low-level acoustic features and emotion category or affective scale judgments. The bottom row of Fig. 8 shows the extent to which the associations between low-level acoustic features and emotion category or affective scale judgments are preserved across India and the United States. The low-level acoustic correlates of the 12 PPCs—the 12 emotions that Indian and US participants reliably recognized in the speech samples—were very well preserved across cultures. The cross-cultural Spearman correlation in acoustic correlates exceeded 0.95 for 5 of the PPCs and exceeded 0.8 for all but Distress and Contempt. By contrast, the correlations between acoustic features and the recognition of Valence—which is considered a basic building block of emotional life[79]—were considerably less well preserved across cultures ($\rho = 0.40$, 90% confidence interval: $0.02 < \rho < 0.76$; significantly lower than ρ for 5 of the 12 PPCs (Amusement, Anger, Awe, Desire and Disappointment; $\rho = 1$, 0.99, 0.99, 0.99 and 0.95; $P = 0.002$, 0.004, 0.002, 0.004 and 0.008), $q < 0.05$, two-tailed bootstrap test; see Supplementary Note 11). Acoustic correlates of many of the raw category judgments were also better preserved than Valence, as were the acoustic correlates of several less typically studied affective features; see Supplementary Fig. 8 for breakdown by category and affective scale. These findings reveal that acoustic parameters that are thought to contribute to emotion recognition are more robustly associated with emotion category judgments than with Valence judgments in terms of how they are preserved across the United States and India.

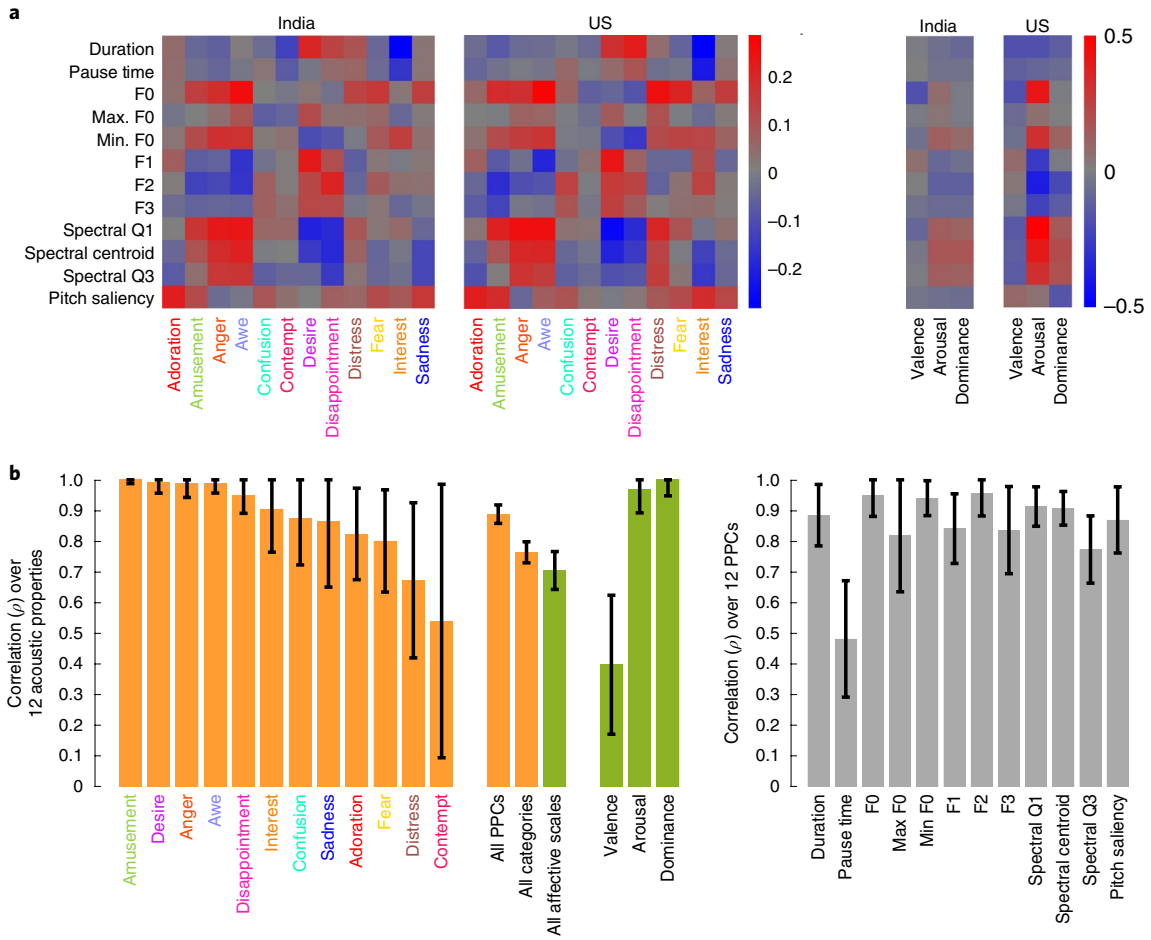

Fig. 8 | Low-level acoustic correlates of emotion recognition and their preservation across cultures. a, Correlations (ρ) between each acoustic property and judgments from each culture across the 2,519 speech samples. Acoustic correlates of the 12 emotion dimensions (PPCs) are similar in both cultures. For example, judgments of Awe correlate with fundamental frequency (F0) in both cultures. However, the acoustic correlates of the affective scale judgments are less similar across cultures. Supplementary Figure 8 shows results for every category and dimension. **b**, Cross-cultural signal correlations in acoustic correlates of emotion category and affective feature recognition. Each coloured bar represents the Spearman correlation between a given column of the above acoustic correlation matrices across cultures (individual dimensions A–F as well as Valence, Arousal and Dominance), between entire matrices (all PPCs, all categories, all affective scales; see full category/affective scale matrices in Supplementary Fig. 8) or between rows (duration, F0 and the 10 other acoustic properties). The acoustic correlates of many of the emotion dimensions, or distinct kinds, are very well-preserved across cultures, whereas those of Valence are considerably less well-preserved. Error bars represent the standard error (participant sample size, $n_{US} = 525$, $n_{India} = 152$ for emotion categories, and $n_{US} = 927$ or 827, $n_{India} = 242$ or 205 for the different affective scales). F1, F2 and F3 represent the first, second and third formants. Q1 and Q3 are the first and third spectral quartiles. See Supplementary Note 11 for details.

Low-level auditory features are likely to support inferences made at early stages of processing. Thus, the finding that the low-level acoustic correlates of most emotion categories (such as Amusement or Fear) were much better preserved across cultures than those of Valence lends further support to the hypothesis that the recognition of emotion categories occurs at earlier stages of processing.

Discussion

Recent studies have documented that the voice is a rich medium of emotional communication, one with cross-cultural similarities and early developmental onset in terms of what emotions are conveyed in the voice. What is less well-understood is the taxonomy of emotions recognized from prosody—that is, how the emotions recognized from prosody are arranged within a semantic space—and how this taxonomy of emotion may be preserved across cultures.

Using mathematically based approaches[7,37], we examined the shared semantic space of the recognition of emotion from speech prosody in participants from the United States and India. Our focus was to test hypotheses related to three properties of this semantic space: its conceptualization—which focused explicitly on how emotion categories and scales of affect contribute to the recognition of emotion in prosody; its dimensionality—the number of distinct kinds of emotion conveyed in prosody; and its distribution of states, here focusing on the nature of the boundaries between emotion categories.

Guided by this conceptual framework, over 2,000 US and Indian participants judged 2,519 prosodically modulated speech samples produced by 100 actors from 5 distinct English-speaking cultures. They either judged each prosody sample in terms of which emotion category (from a list of 30) was expressed or rated the prosody sample on scales that capture 23 affective features theorized in appraisal and componential theories to account for emotion recognition. Applying large-scale statistical inference techniques, we compared the preservation of the recognition of 30 categories and 23 scales of affect across cultures, modelled the latent space that captured the shared variance in judgments between cultures and interrogated the

boundaries between the 12 categories that were found to underlie this latent space.

With respect to the dimensionality of the semantic space of emotion recognition, many studies of expressive signalling have focused on 6–8 categories of emotion and relied on either interrater agreement rates[12,13,21,42] or factor analysis[44,47,48] to characterize the recognition of emotion. Applying statistical modelling techniques to judgments of a vast array of stimuli, we uncovered 12 distinct emotions that were recognized in India and the United States. The 12 emotions—Adoration, Amusement, Anger, Awe, Confusion, Contempt, Desire, Disappointment, Distress, Fear, Interest and Sadness—that emerged from our analyses were highly correlated across India and the United States and most were also found to have distinct acoustic correlates that were preserved robustly across the two cultures (Fig. 6). It will be important to examine whether these emotions emerge in studies of other kinds of emotional vocalization—for example, vocal bursts or song—and whether they emerge in other cultures and, in particular, among non-English speakers. We note that this investigation yields evidence of the shared recognition not only of commonly studied emotional states, such as Anger, Fear, Sadness and Surprise, but also of less commonly studied emotional states, including traditionally understudied varieties of positive emotion, such as Adoration, Amusement, Awe, Desire and Interest (Fig. 4). Understanding the rich variety of emotions conveyed in prosody may be particularly useful for studies of the physiological and neural representations of distinct emotions (especially considering that conversation with subjects is already commonplace in most studies).

In addition, we wanted to examine one property of the distribution of emotional states within the semantic space of emotion recognition—the boundaries between emotion categories. In contrast to discrete theories of emotion[40,65–70], we find that emotional prosody is characterized not by discrete clusters of states, but by smooth gradients between emotion categories. Prosodic signals occupy gradients from Adoration to Amusement, Sadness to Distress, Interest to Confusion, and between many other categories (Figs. 4 and 5). These findings may help to explain past findings of interrater variability in the perception of emotional signals[12,72,88], suggesting that disagreement across raters in forced-choice rating tasks may reflect the intermediacy of states between categories signalled by expressive behaviour as opposed to just the indistinctness of the categories or individual differences in recognition[60]. Furthermore, these findings support a shift from the predominantly scientific focus of how discrete patterns of expression, physiology and neural activation distinguish discrete emotion states[89–93] toward an understanding of the continuous variability and blending together of emotion categories by continuously varying patterns of expression, physiology and neural activation[39,41,71].

Lastly, we examined one critical issue related to the conceptualization of emotion in prosody: whether the recognition of emotion in prosody is explained at a more basic, cross-cultural level by categorical labels or scales of affect. It has been posited that in the recognition of emotion, the signalling of Valence, usually along with Arousal, is a core, low-level interpretive process from which specific emotion categories are constructed[43,79]. In contrast to this claim, we found empirical evidence that the recognition from prosody of categories such as Amusement and Desire is better preserved across cultures than that of Valence (Fig. 1). Judgments along scales of affect, including Valence, are better predicted by category judgments from another culture than by the identical scale-of-affect judgments from another culture (Fig. 2 and Supplementary Fig. 4). Finally, the low-level acoustic correlates of the recognition of many emotion categories are extremely well-preserved across cultures, whereas those for Valence are not at all well-preserved (Fig. 8), suggesting that the recognition of emotion categories may occur at earlier stages of processing. Taken together, these results suggest that categories such as Amusement and Desire may be recognized more directly from prosody, and that judgments of broader affective features may subsequently be inferred in a more culture-specific manner from these categories.

It is important to note that the pattern of results observed in this investigation was potentially influenced by the kind of prosody that we studied, the emotion recognition response formats, and the cultures—both English-speaking—that we included. Given this, it will be important to extend the methods presented here to other kinds of prosody captured in contexts in which speakers are not directed to communicate specific emotions as in the present study[32,94]. It will be important to study more naturalistic, spontaneous forms of prosody and the range of emotions such forms of prosody communicate as well as the potentially broader semantic space of emotion that captures such signals[54,55,57].

We note that the present results pertain to similarities in the recognition of emotional prosody between two English-speaking cultures, the United States and India. By examining distinct English-speaking cultures (see Supplementary Discussion 7), we were able to interrogate the relative preservation of distinct emotional signals in prosodically modulated speech samples, while holding constant the effects of interpretations of the phonetic and semantic content of the sentences. However, cultures that adopt the English language may acquire certain prosodic conventions in addition to its lexicon, which may shape the prosodic communication of emotion, in part accounting for the high degrees of similarity across the United States and India in conveying distinct emotions through prosody. It will be important for future research to use similar methods to examine the structure of emotional prosody in other languages, such as French and Chinese, that have different prosodic conventions[32].

More generally, our findings fit with two general interpretations: that they are explained by the innate psychological basicness of varieties of emotional prosody in all humans, or by the acquired psychological basicness of varieties of emotional prosody in English speakers. Both interpretations point to the primacy of signals of emotion categories, such as Amusement, over signals of affective features, such as Valence. Nevertheless, it remains unclear whether the specific categories recognized from emotional prosody in the United States and India are universal to all human languages.

Recent studies of the recognition of nonverbal expressive signals in remote cultures do, however, point to broader universals in the recognition of vocal emotion categories. A previous study[23] assessed the recognition of nonverbal vocal bursts targeting 16 emotion categories in a remote culture in Bhutan; moderate to strong recognition (around 50% accuracy or greater, with chance = 25%) of more than half of the targeted categories was found in this study, including 7 of the categories found by the present study to be distinguished from prosody in the United States and India (Amusement, Anger, Awe, Desire, Interest, Fear and Sadness). Another study[16] assessed the recognition of 9 emotion categories from vocal bursts among Himba listeners from remote Namibian villages and in this study moderate to strong recognition (around 75% accuracy or greater, with chance = 50%) of more than half of the targeted categories was reported, including 4 found to be recognized from prosody in the present study (Amusement, Anger, Fear and Sadness). In the same study, English listeners were found to strongly recognize (>90% accuracy) 8 of the 9 targeted categories of emotion from recorded Himba vocalizations[72]. While the overlap in the mechanisms of recognizing vocal bursts and prosody is not well-understood, these recent findings offer early clues that the recognition of signals of many distinct categories of emotion from the human voice may be universal as opposed to unique to English-speaking cultures.

Finally, it is worth noting how the present findings dovetail with recent research on reported experiences of emotion. It was previously shown[7] that the subjective feelings that people report in response to viewing a wide range of evocative videos (2,185 in

total) reliably distinguish among 27 distinct varieties of emotion. As with the present findings regarding emotional prosody, categories emerged as more primary in determining the structure of experience, and these reported experiences were found to be organized along continuous gradients bridging categories of emotion, such as Interest and Awe. Together, these results converge on a taxonomy of emotion consisting of a rich array of distinct categories bridged by smooth gradients.

Debates over the structure of expressive signals are foundational to the science of emotion. They bear upon central theoretical claims about emotion and exert a profound influence on fields ranging from affective neuroscience to machine learning. Our method of interrogating how the varieties of emotional prosody are situated within a semantic space reveals a more complex taxonomy of expressive states than is typical in existing accounts of how emotions are organized. Prosodic signals reliably convey at least 12 distinct dimensions of emotion and are distributed along continuous gradients between them.

Methods

Speech samples. The 2,519 speech samples were drawn from the VENEC (vocal expressions of nineteen emotions across cultures) corpus, a large cross-cultural database[33]. Actors from Australia, India, Kenya, Singapore and the United States were provided with scenarios describing typical situations in which each of 18 emotions may be elicited and were instructed to enact finding themselves in similar situations. The emotion categories targeted by the VENEC corpus were Affection, Anger, Amusement, Contempt, Disgust, Distress, Fear, Guilt, Happiness, Interest, Lust, Negative surprise, Positive surprise, Pride, Relief, Sadness, Serenity, and Shame. See Supplementary Methods for further details.

Emotion judgments. Emotion judgments of the speech samples were obtained using Amazon Mechanical Turk. Three separate survey formats were used to obtain emotion judgments: one for the category judgments and two for the affective scale judgments. Each individual survey presented a subset of speech samples (30 for the category judgments, 12 for the affective scale judgments, assigned randomly) in an order randomized for each participant. A total of 2,345 English-speaking participants, including 1,969 US participants (1,095 females, mean age = 36 years) and 376 Indian participants (123 females, mean age = 30 years), took part in the study. For each judgments format, 10–15 judgments were collected of each speech sample from each culture. (No statistical methods were used to predetermine sample sizes, but our sample sizes are similar to those in a previous study in which similar methods captured over 90% of the systematic variability in judgments of emotional videos[7].) US participants rated 71.6 speech samples on average and Indian participants rated 249.2 speech samples on average. The experimental procedures were approved by the Institutional Review Board at the University of California, Berkeley. All participants gave their informed consent. See Supplementary Methods for further details.

Statistical analyses. Our statistical analyses are outlined briefly below. Data were analysed primarily using custom code in MATLAB version R2012b. Acoustic properties were computed using the BioSound Toolbox in Python (http://github.com/theunissenlab/BioSoundTutorials). Analyses were not performed blind to the conditions of the experiments. For a detailed description of each method, see Supplementary Methods.

Category judgment proportions. For each speech sample, we computed the proportion of participants who chose each category and the average judgments of each affective scale. To estimate the significance of the category judgment proportions of each speech sample we constructed a null distribution of category judgment proportions using a Monte Carlo simulation.

Signal correlations. To derive signal correlations between cultures for each judgment, we correlated the mean judgments from each culture across all speech samples and divided by the estimated explainable variance. Explainable variance was estimate by dividing the mean of the squared standard errors (estimated using bootstrapping) by the total variance and subtracting this quantity from 1. To calculate standard errors and P values for signal correlations, it was necessary to account for potential non-independence across ratings of different speech samples due to the fact that each rater rated multiple samples. To do so, we applied a non-parametric bootstrap approach, using stratified resampling across individual raters rather than individual ratings. We validate these methods by demonstrating that signal correlations accurately estimate the respective population-level correlations in Monte Carlo simulations (Supplementary Fig. 2).

Regression between category and affective scale judgments. We predicted affective scale judgments from category judgments using ordinary least-squares linear regression. Here, it may be worth acknowledging that methods specialized for sparse data could potentially have produced better prediction correlations. However, this only generates a more conservative interpretation of our findings that category judgments explain the preservation of affective scale judgments across cultures.

PPCA. We determined the number of dimensions necessary to explain the preservation of emotion category recognition across cultures by introducing a method called PPCA, which has two versions, correlational and covariational. Covariational PPCA maximizes the objective function $\mathrm{Cov}(X\alpha, Y\alpha)$ whereas correlational PPCA maximizes the objective function $\mathrm{Corr}(X[X^TX]^{-1/2}\alpha, Y[Y^TY]^{-1/2}\alpha)$. We tested the significance of each component by applying each version of PPCA in a leave-one-rater-out fashion, determining whether the held-out ratings projected onto each component were consistently positively correlated with component scores of ratings from the other country using a non-parametric Wilcoxon signed-rank test[83]. After determining the number of significant PPCs, we applied varimax rotation, generating more interpretable components. To compute P and q values for the scores of each individual speech sample on each component, we used a Monte Carlo simulation of the category ratings.

It is worth acknowledging that we do not establish here that PPCA is applicable to all distributions of data. However, we do show that PPCA generates accurate results on randomly simulated data that are distributed identically to those of the present study, but with varying numbers of underlying dimensions (Fig. 3).

Maps. To visualize the distribution of speech samples within the multidimensional space derived using PPCA, we applied a method called t-SNE. We then assigned a colour to each speech sample in the map corresponding to a weighted average of the unique colours of its top 2 scores on the 12 categorical judgment dimensions.

Continuous versus discrete category models. We compared how well continuous versus discrete category models predicted affective scale ratings using ordinary least-squares regression. For the discrete models, we used ordinary least-squares analyses to predict the affective scale judgments from the maximally loading PPC, which was converted to a dummy variable (1 for the maximally loading PPC, 0 otherwise) to form a fully discrete model, and to a continuous intensity (keep the maximally loading PPC, convert others to 0) to form a discrete model with intensity. For these analyses, we averaged across ratings from the United States and India. To test for a difference in variance explained between the continuous and discrete category models, we used across-rater bootstrap resampling.

Acoustic measures. To analyse the acoustic correlates of emotion recognition, we computed twelve acoustic measurements of each speech sample. (1) duration, (2) pause time, (3) F0, (4) maximum F0, (5) minimum F0, (6)–(8) F1–3, (9) Spectral Q1, (10) spectral centroid, (11) spectral Q3 and (12) pitch saliency.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Code availability

Custom MATLAB analysis code can be requested from https://goo.gl/forms/3q0y2Vvi1KinMft13.

Data availability

The 2,519 speech samples used in the present study and their ratings can be requested from https://goo.gl/forms/3q0y2Vvi1KinMft13. Publications incorporating the speech samples should reference the previous study[33].

Received: 19 October 2017; Accepted: 15 January 2019;
Published online: 11 March 2019

References

1. Keltner, D. & Haidt, J. Social functions of emotions at four levels of analysis. *Cogn. Emot.* **13**, 505–521 (1999).
2. Nesse, R. M. Evolutionary explanations of emotions. *Hum. Nat.* **1**, 261–289 (1990).
3. Campos, B., Shiota, M. N., Keltner, D., Gonzaga, G. C. & Goetz, J. L. What is shared, what is different? Core relational themes and expressive displays of eight positive emotions. *Cogn. Emot.* **27**, 37–52 (2013).
4. Oveis, C., Spectre, A., Smith, P. K., Liu, M. Y. & Keltner, D. Laughter conveys status. *J. Exp. Soc. Psychol.* **65**, 109–115 (2016).
5. Gonzaga, G. C., Keltner, D., Londahl, E. A. & Smith, M. D. Love and the commitment problem in romantic relations and friendship. *J. Pers. Soc. Psychol.* **81**, 247–262 (2001).
6. ten Brinke, L. & Adams, G. S. Saving face? When emotion displays during public apologies mitigate damage to organizational performance. *Organ. Behav. Hum. Decis. Process.* **130**, 1–12 (2015).

7. Cowen, A. S. & Keltner, D. Self-report captures 27 distinct categories of emotion bridged by continuous gradients. *Proc. Natl Acad. Sci. USA* **114**, E7900–E7909 (2017).
8. Schirmer, A. & Adolphs, R. Emotion perception from face, voice, and touch: comparisons and convergence. *Trends Cogn. Sci.* **21**, 216–228 (2017).
9. Singer, T. & Lamm, C. The social neuroscience of empathy. *Ann. NY Acad. Sci.* **1156**, 81–96 (2009).
10. Frühholz, S., Ceravolo, L. & Grandjean, D. Specific brain networks during explicit and implicit decoding of emotional prosody. *Cereb. Cortex* **22**, 1107–1117 (2012).
11. Bach, D. R. et al. The effect of appraisal level on processing of emotional prosody in meaningless speech. *Neuroimage* **42**, 919–927 (2008).
12. Cordaro, D. T. et al. Universals and cultural variations in 22 emotional expressions across five cultures. *Emotion* **18**, 75–93 (2018).
13. Elfenbein, H. A. & Ambady, N. On the universality and cultural specificity of emotion recognition: a meta-analysis. *Psychol. Bull.* **128**, 203–235 (2002).
14. Keltner, D. & Cordaro, D. T. in *Emotion Researcher* (Scarantino, A. ed.) Available at http://emotionresearcher.com/understanding-multimodal-emotional-expressions-recent-advances-in-basic-emotion-theory/ (2015).
15. Norenzayan, A. & Heine, S. J. Psychological universals: what are they and how can we know? *Psychol. Bull.* **131**, 763–784 (2005).
16. Sauter, D. A., Eisner, F., Ekman, P. & Scott, S. K. Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proc. Natl Acad. Sci. USA* **107**, 2408–2412 (2010).
17. Filippi, P. *et al.* Humans recognize emotional arousal in vocalizations across all classes of terrestrial vertebrates: evidence for acoustic universals. *Proc. R. Soc. B* **284**, 20170990 (2017).
18. Parr, L. A., Waller, B. M. & Vick, S. J. New developments in understanding emotional facial signals in chimpanzees. *Curr. Dir. Psychol. Sci.* **16**, 117–122 (2007).
19. Snowdon, C. T. in *Handbook of Affective Sciences* (eds Davidson, R. J. & Scherer, K. R.) 457-480 (Oxford Univ. Press, Oxford, 2002).
20. Filippi, P. Emotional and interactional prosody across animal communication systems: a comparative approach to the emergence of language. *Front. Psychol.* **7**, 1393 (2016).
21. Adolphs, R. Neural systems for recognizing emotion. *Curr. Opin. Neurobiol.* **12**, 169–177 (2002).
22. Russell, J. A. Is there universal recognition of emotion from facial expressions? A review of the cross-cultural studies. *Psychol. Bull.* **115**, 102–141 (1994).
23. Cordaro, D. T., Keltner, D., Tshering, S., Wangchuk, D. & Flynn, L. M. The voice conveys emotion in ten globalized cultures and one remote village in Bhutan. *Emotion* **16**, 117–128 (2016).
24. Gendron, M., Roberson, D., van der Vyver, J. M. & Barrett, L. F. Cultural relativity in perceiving emotion from vocalizations. *Psychol. Sci.* **25**, 911–920 (2014).
25. Hertenstein, M. J. & Campos, J. J. The retention effects of an adult's emotional displays on infant behavior. *Child Dev.* **75**, 595–613 (2004).
26. Juslin, P. N. & Laukka, P. Communication of emotions in vocal expression and music performance: different channels, same code? *Psychol. Bull.* **129**, 770–814 (2003).
27. Keltner, D. et al. in *Handbook of Emotions* 4th edn (eds Lewis M., Haviland-Jones, J. M. & Barrett, L. F.) 467–482 (Guilford, New York, 2016).
28. Wu, Y., Muentener, P. & Schulz, L. E. One- to four-year-olds connect diverse positive emotional vocalizations to their probable causes. *Proc. Natl Acad. Sci. USA* **114**, 11896–11901 (2017).
29. Titze, I. R. & Martin, D. W. Principles of voice production. *J. Acoust. Soc. Am.* **104**, 1148 (1998).
30. Scherer, K. R. & Bänziger, T. Emotional expression in prosody: a review and an agenda for future research. In *Proc. Speech Prosody 2004* 359–366 (2004).
31. Mitchell, R. L. C. & Ross, E. D. Attitudinal prosody: what we know and directions for future study. *Neurosci. Biobehav. Rev.* **37**, 471–479 (2013).
32. Hancil, S. *The Role of Prosody in Affective Speech* (Peter Lang, New York, 2009).
33. Laukka, P. et al. 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**, 686–705 (2016).
34. Nordström, H., Laukka, P., Thingujam, N. S., Schubert, E. & Elfenbein, H. A. Emotion appraisal dimensions inferred from vocal expressions are consistent across cultures: a comparison between Australia and India. *R. Soc. Open Sci.* **4**, 170912 (2017).
35. Paulmann, S. & Uskul, A. K. Cross-cultural emotional prosody recognition: evidence from Chinese and British listeners. *Cogn. Emot.* **28**, 230–244 (2014).
36. Scherer, K. R., Banse, R. & Wallbott, H. G. Emotion inferences from vocal expression correlate across languages and cultures. *J. Cross Cult. Psychol.* **32**, 76–92 (2001).
37. Cowen, A. S. & Keltner, D. Clarifying the conceptualization, dimensionality, and structure of emotion: response to Barrett and colleagues. *Trends Cogn. Sci.* **22**, 274–276 (2018).
38. Laukka, P. *et al.* Cross-cultural decoding of positive and negative non-linguistic emotion vocalizations. *Front. Psychol.* **4**, 353 (2013).
39. Parr, L. A., Cohen, M. & de Waal, F. Influence of social context on the use of blended and graded facial displays in chimpanzees. *Int. J. Primatol.* **26**, 73–103 (2005).
40. Ekman, P. in *The Nature of Emotion* (eds Ekman, P. & Davidson, R. J.) 15–19 (Oxford Univ. Press, Oxford, 1992).
41. Harris, R. J., Young, A. W. & Andrews, T. J. Morphing between expressions dissociates continuous from categorical representations of facial expression in the human brain. *Proc. Natl Acad. Sci. USA* **190**, 21164–21169 (2012).
42. Russell, J. A. Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychol. Bull.* **115**, 102–141 (1994).
43. Russell, J. A. Core affect and the psychological construction of emotion. *Psychol. Rev.* **110**, 145–172 (2003).
44. Smith, C. A. & Ellsworth, P. C. Patterns of cognitive appraisal in emotion. *J. Pers. Soc. Psychol.* **48**, 813–838 (1985).
45. Frijda, N. H., Kuipers, P. & ter Schure, E. Relations among emotion, appraisal, and emotional action readiness. *J. Pers. Soc. Psychol.* **57**, 212–228 (1989).
46. Scherer, K. R. The dynamic architecture of emotion: evidence for the component process model. *Cogn. Emot.* **23**, 1307–1351 (2009).
47. Watson, D. & Tellegen, A. Toward a consensual structure of mood. *Psychol. Bull.* **98**, 219–235 (1985).
48. Posner, J., Russell, J. A. & Peterson, B. S. The circumplex model of affect: an integrative approach to affective neuroscience, cognitive development, and psychopathology. *Dev. Psychopathol.* **17**, 715–734 (2005).
49. Russell, J. A circumplex model of affect. *J. Pers. Soc. Psychol.* **39**, 1161–1178 (1980).
50. Ang, J., Dhillon, R., Krupski, A., Shriberg, E. & Stolcke, A. Prosody-based automatic detection of annoyance and frustration in human-computer dialog. In *Proc. 7th International Conference on Spoken Language Processing* 2037–2040 (2002).
51. Laukka, P., Neiberg, D., Forsell, M., Karlsson, I. & Elenius, K. Expression of affect in spontaneous speech: acoustic correlates and automatic detection of irritation and resignation. *Comput. Speech Lang.* **25**, 84–104 (2011).
52. Provine, R. R. & Fischer, K. R. Laughing, smiling, and talking: relation to sleeping and social context in humans. *Ethology* **83**, 295–305 (1989).
53. Vidrascu, L. & Devillers, L. Real-life emotion representation and detection in call centers data. In *Proc. 3784th Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 739–746 (Springer, 2005).
54. Sauter, D. A. & Fischer, A. H. Can perceivers recognise emotions from spontaneous expressions?. *Cogn. Emot.* **32**, 504–515 (2018).
55. Anikin, A. & Lima, C. F. Perceptual and acoustic differences between authentic and acted nonverbal emotional vocalizations. *Q. J. Exp. Psychol.* **71**, 622–641 (2018).
56. Scherer, K. R. Vocal markers of emotion: comparing induction and acting elicitation. *Comput. Speech Lang.* **27**, 40–58 (2013).
57. Juslin, P. N., Laukka, P. & Bänziger, T. The mirror to our soul? Comparisons of spontaneous and posed vocal expression of emotion. *J. Nonverbal Behav.* **42**, 1–40 (2018).
58. Gupta, V., Hanges, P. J. & Dorfman, P. Cultural clusters: methodology and findings. *J. World Bus.* **37**, 11–15 (2002).
59. Jaju, A., Kwak, H. & Zinkhan, G. M. Learning styles of undergraduate business students: cross-cultural comparison between the US, India, and Korea. *Mark. Educ. Rev.* **12**, 49–60 (2002).
60. Barrett, L. F. Are emotions natural kinds? *Perspect. Psychol. Sci.* **1**, 28–58 (2006).
61. Ekman, P. What scientists who study emotion agree about. *Perspect. Psychol. Sci.* **11**, 31–34 (2016).
62. Ekman, P. & Cordaro, D. What is meant by calling emotions basic. *Emot. Rev.* **3**, 364–370 (2011).
63. Keltner, D. & Lerner, J. S. in *Handbook of Social Psychology* 5th edn (eds Fiske, S. T. et al., Wiley Online Library, Hoboken NJ, 2010).
64. Lazarus, R. S. Progress on a cognitive–motivational–relational theory of emotion. *Am. Psychol.* **46**, 819–834 (1991).
65. Roseman, I. J. Appraisal determinants of discrete emotions. *Cogn. Emot.* **5**, 161–200 (1991).
66. Etcoff, N. L. & Magee, J. J. Categorical perception of facial expressions. *Cognition* **44**, 227–240 (1992).
67. Harmon-Jones, C., Bastian, B. & Harmon-Jones, E. The discrete emotions questionnaire: a new tool for measuring state self-reported emotions. *PLoS ONE* **11**, e0159915 (2016).
68. Izard, C. E. Four systems for emotion activation: cognitive and noncognitive processes. *Psychol. Rev.* **100**, 68–90 (1993).
69. Johnson-Laird, P. N. & Oatley, K. The language of emotions: an analysis of a semantic field. *Cogn. Emot.* **3**, 81–123 (1989).
70. Shiota, M. N. et al. Beyond happiness: building a science of discrete positive emotions. *Am. Psychol.* **72**, 617–643 (2017).

71. Samson, A. C., Kreibig, S. D., Soderstrom, B., Wade, A. A. & Gross, J. J. Eliciting positive, negative and mixed emotional states: a film library for affective scientists. *Cogn. Emot.* **30**, 827–856 (2016).
72. Gendron, M., Roberson, D., van der Vyver, J. M. & Barrett, L. F. Perceptions of emotion from facial expressions are not culturally universal: evidence from a remote culture. *Emotion* **14**, 251–262 (2014).
73. Laukka, P., Neiberg, D. & Elfenbein, H. A. Evidence for cultural dialects in vocal emotion expression: acoustic classification within and across five nations. *Emotion* **14**, 445–449 (2014).
74. Mehrabian, A. & Russell, J. *An Approach to Environmental Psychology* (MIT Press, Cambridge MA, 1974).
75. Osgood, C. E. Dimensionality of the semantic space for communication via facial expressions. *Scand. J. Psychol.* **7**, 1–30 (1966).
76. Sauter, D. A. & Scott, S. K. More than one kind of happiness: can we recognize vocal expressions of different positive states? *Motiv. Emot.* **31**, 192–199 (2007).
77. Simon-Thomas, E. R., Keltner, D. J., Sauter, D., Sinicropi-Yao, L. & Abramson, A. The voice conveys specific emotions: evidence from vocal burst displays. *Emotion* **9**, 838–846 (2009).
78. Benjamini, Y. & Yu, B. The shuffle estimator for explainable variance in FMRI experiments. *Ann. Appl. Stat.* **7**, 2007–2033 (2013).
79. Barrett, L. F. Valence is a basic building block of emotional life. *J. Res. Pers.* **40**, 35–55 (2006).
80. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).
81. Barrett, L. F., Lindquist, K. A. & Gendron, M. Language as context for the perception of emotion. *Trends Cogn. Sci.* **11**, 327–332 (2007).
82. Abdi, H. & Williams, L. J. Partial least squares methods: partial least squares correlation and partial least square regression. *Comput. Toxicol.* **930**, 549–579 (2013).
83. Hardoon, D. R., Szedmak, S. & Shawe-Taylor, J. Canonical correlation analysis: an overview with application to learning methods. *Neural Comput.* **16**, 2639–2664 (2004).
84. Wilcoxon, F. Individual comparisons by ranking methods. *Biom. Bull.* **1**, 80–83 (1945).
85. Van Der Maaten, L. & Hinton, G. Visualizing data using *t*-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
86. Scherer, K. R. Vocal affect expression: a review and a model for future research. *Psychol. Bull.* **99**, 143–165 (1986).
87. Ringeval, F. et al. AV+EC 2015: The first affect recognition challenge bridging across audio, video, and physiological data. In *Proc. 5th International Workshop on Audio/Visual Emotion Challenge* 3–8 (ACM, 2015).
88. Haidt, J. & Keltner, D. Culture and facial expression: open-ended methods find more expressions and a gradient of recognition. *Cogn. Emot.* **13**, 225–266 (1999).
89. Kragel, P. A. & LaBar, K. S. Multivariate neural biomarkers of emotional states are categorically distinct. *Soc. Cogn. Affect. Neurosci.* **10**, 1437–1448 (2015).
90. Kreibig, S. D. Autonomic nervous system activity in emotion: a review. *Biol. Psychol.* **84**, 394–421 (2010).
91. Lench, H. C., Flores, S. A. & Bench, S. W. Discrete emotions predict changes in cognition, judgment, experience, behavior, and physiology: a meta-analysis of experimental emotion elicitations. *Psychol. Bull.* **137**, 834–855 (2011).
92. Vytal, K. & Hamann, S. Neuroimaging support for discrete neural correlates of basic emotions: a voxel-based meta-analysis. *J. Cogn. Neurosci.* **22**, 2864–2885 (2010).
93. Wager, T. D. et al. in *Handbook of Emotions* 3rd edn (eds Lewis, M. et al.) 249–271 (Guilford, New York, 2008).
94. Scherer, K. & Bänziger, T. in *Blueprint for Affective Computing: A Sourcebook* (eds Scherer, K. R., Banziger, T., & Roesch, E.) 166–176 (Oxford Univ. Press, Oxford, 2010).
95. G'Sell, M. G., Wager, S., Chouldechova, A. & Tibshirani, R. Sequential selection procedures and false discovery rate control. *J. R. Stat. Soc. B.* **78**, 423–444 (2016).

Acknowledgements

We thank R. Rosipal for devising a correlational version of PPCA and F. Theunissen for providing input regarding acoustic analyses. Research reported in this publication was supported by the US National Institute of Mental Health under award number T32-MH020006-16A1 and by the Thomas and Ruth Ann Hornaday Chair in Psychology at the University of California, Berkeley. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

P.L. and H.A.E. contributed all speech samples; A.S.C. and D.K. designed the research with input from P.L. and H.A.E.; A.S.C. performed research, contributed analytic tools and analysed data; and A.S.C. and D.K. wrote the paper with input from P.L., H.A.E. and R.L.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41562-019-0533-6.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to A.S.C.

nature research

Corresponding author(s): Alan S. Cowen

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	Surveys were coded in Javascript and HTML and launched on Amazon Mechanical Turk.
Data analysis	Data were analyzed primarily using custom code in Matlab (versions 2017a and 2018a). Acoustic properties were computed using the BioSound Toolbox in Python (http://github.com/theunissenlab/BioSoundTutorials).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The 2,519 speech samples used in the present study and their ratings can be requested here: https://goo.gl/forms/3q0y2Vvi1KinMft13.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	A quantitative examination of the emotions people recognize in speech prosody in both the US and India.
Research sample	A total of 2345 English-speaking participants recruited on Amazon Mechanical Turk, including 1969 US participants (1095 females, mean age = 36 y) and 376 Indian participants (123 females, mean age = 30 y), took part in the study. Mechanical Turk samples typically span a wide geographic and demographic range. The use of Mechanical Turk made it feasible to collect over one million human judgments.
Sampling strategy	Three separate survey formats were used to obtain emotion judgments: one for the category judgments and two for the affective scale judgments. Each individual survey presented a subset of speech samples (30 for the category judgments, 12 for the affective scale judgments, assigned randomly) in an order randomized for each participant. For each judgments format, 10-15 judgments were collected of each speech sample from each culture. No statistical methods were used to pre-determine sample sizes, but our sample sizes are similar to those in a previous study in which similar methods captured over 90% of the systematic variability in judgments of emotional videos (Cowen & Keltner, 2017).
Data collection	Data collection was performed online using Amazon Mechanical Turk.
Timing	Data were collected in May, 2017.
Data exclusions	No data were excluded from analyses.
Non-participation	No participants dropped out.
Randomization	Each individual survey presented a subset of speech samples (30 for the category judgments, 12 for the affective scale judgments, assigned randomly) in an order randomized for each participant.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	See above.
Recruitment	Participants opted to participate in surveys on Amazon Mechanical Turk entitled "Rate emotions expressed by the voice". It is unclear how self-selection bias (perhaps toward people interested in emotion or the human voice) would impact results.