Published in final edited form as:

Nat Hum Behav. 2019 April ; 3(4): 369–382. doi: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,*], **Dr. Petri Laukka**[2], **Dr. Hillary Anger Elfenbein**[3], **Runjing Liu**[4], **Dr. Dacher Keltner**[1]

[1]Department of Psychology, University of California, Berkeley, 2121 Berkeley Way, Berkeley, CA 94720

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

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

[4]Department of Statistics, University of California, 367 Evans Hall, Berkeley, CA 94720

Abstract

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 five 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 two cultures. Analyses of the semantic and acoustic structure of emotion recognition reveal that emotion categories drive emotion recognition 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 (https://s3-us-west-1.amazonaws.com/venec/map.html), reveal a complex, high-dimensional space of emotional states recognized cross-culturally in speech prosody.

Emotion recognition is fundamental to human social interaction. Brief emotional displays in the face and voice by nearby adults guide infants' and children's responses to their environment, and figure prominently in how 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 should not surprise that the recognition of facial

[*]To whom correspondence should be addressed. alan.cowen@berkeley.edu.

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

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

Code availability. Custom Matlab analysis code can be requested here: https://goo.gl/forms/3q0y2Vvi1KinMft13.

Competing interests. The authors declare no competing interests.

expression and emotion-related vocalization is known to be processed in specific brain regions [7–11], to be preserved to a considerable extent across many cultures [12–16], and to 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 emotion [13,21,22]. More recently, scientists have begun to document the varied ways in which humans communicate emotion with the voice. With short bursts of sound, known as vocal bursts, humans can communicate upwards of 15 emotions, a finding now replicated in over a dozen cultures, including two remote cultures with minimal contact with the West (e.g., [16,23]; but see [24] for findings of greater cultural relativity). With fleeting emotional vocalizations, parents communicate to infants what is worthy of approach or warranting avoidance [25], adults infer a person's rank within a social hierarchy [4], and singers convey specific emotions in song [26] (for review, see [27]). By the age of 2, children can readily identify at least five positive emotions from brief emotion-related vocalizations [28].

In the present investigation, we focus on emotional prosody—the non-lexical 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: Including Timbre in Prosody.) 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 upwards of 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. That is, what is the mapping from the variations in emotional prosody people hear to the complex network of words and phrases that people (including scientists) rely on to represent emotion? How many *distinct* emotions do people recognize in the complex array of variations in prosody they hear in their daily lives? What drives their recognition of emotion, emotion categories (e.g., "awe", "fear") or broader scales that capture core affect appraisals (valence, arousal)? What is the structure of the categories that people rely on to represent emotion inferred from sound: are they discrete or bridged by gradients of meaning?

In the present investigation we seek new answers to these questions. We do so by examining 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 in representing emotion[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 emotion (see Supplementary Discussion 2: Emotion Categories, Affective Features, Scales, and Dimensions 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 signaled by expressive behavior. 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 (e.g., are they discrete or not).

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, it is of note, 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 stands in contrast to a traditional focus in the literature on the recognition of prototypical expressions or visual morphs between them [12,40–43]. 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 behavior of interest – experience or expression, for example. In doing so, studies need to move beyond foundational work that suggested emotions may be organized within a space [44] defined by its two to three broadest dimensions to include the affective features of cognitive appraisal theory [34,45,46] and componential theorizing [47]. Such theories describe how affective features other than valence and arousal are needed to account for the wide array of emotions studied today [44,48]. 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: Multidimensional Reliability Analysis). This large-scale statistical inference approach contrasts with the reliance of typical emotion recognition research on either univariate recognition accuracy (reliability) [12,13,16,21,43] or factor analysis [45,49,50]. 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: Limitations of Traditional Methods for further detail regarding methodological limitations of past studies).

A recent examination of the semantic space of reported emotional experience [7] validated these methodological approaches, suggesting that they can be fruitfully extended to understand the semantic space of emotion recognition. In this previous study, participants reported on the emotional responses to over 2000 videos in terms of a wide array of emotion categories and in terms of 14 affective features derived from appraisal and componential theories of emotion [44–50]. These responses were analyzed to derive a semantic space of

reported emotional experience [7]. This 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 from US and Indian judgments of prosodically modulated, lexically identical speech samples produced by actors from five different cultures imagining themselves 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 [51–53] and as much as a quarter of speech in emotional contexts [54], differing modestly from naturalistic vocalizations in terms of their average perceptual and acoustic features [55–58]. By comparing how participants from India and the US interpret speech samples richly varying only in their prosodic features, we can ascertain how the meaning of emotional prosody may be preserved across two very distinct English-speaking cultures [59,60] within a shared semantic space, including the relative primacy of emotion categories and affective features.

At stake in the study of semantic space of emotion recognition are answers to questions of central theoretical import. First, how should emotional expressions be conceptualized: to what extent do they convey specific categories of emotion, such as awe and fear [40,61–64], and information about affective features, such as valence and arousal [44,45,47,65,66] (and can one of these manners of conceptualizing expression be accounted for by the other)? Second, how many varieties of emotion conveyed by emotional expressions are distinct, thus mapping to separate semantic dimensions? Third, do emotional expressions occupy discrete clusters—families of states such as awe, interest, and surprise [40,66–71], or do they lie along continuous gradients [7,39,41,44,48,72]? And finally, to what extent are the aforementioned properties of emotional expressions preserved across cultures [12,13,15,16,73,74]? To answer these questions and derive a cross-cultural semantic space of the recognition of emotion from prosody, we collected judgments from participants in the US and India of the most extensive stimulus set of emotional prosody in terms of number of emotions considered and cultures of origin of speakers, ideal for deriving a semantic space of emotion recognition for reasons we outlined above.

2345 participants from the US 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 2519 speech samples [33,75]. 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 (US, 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; italicized categories parallel those targeted with scenarios presented during the recording of the speech samples (see Supplementary Methods 1: Recording the Speech Samples).

A second group of participants rated each of the 30 different speech samples they judged in terms of 23 different affective features. These features were culled from dimensional and componential theoretical accounts of the appraisal processes proposed to underlie emotion recognition and experience [44,45,47,66,76,77], 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 Methods 2: Judgment Surveys 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 9-point Likert scales (1 = negative levels or none of the feature, 5 = neutral or moderate levels, 9 = extreme levels of the feature).

Based on 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 Methods 2: Judgment Surveys). The categories used for judgment included the emotions the actors were instructed to target [33] as well as emotions previously found to be conveyed by the voice [23,78,79]. The collection of a wide range of judgments of 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 is more potent in explaining variance in the other 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 signaled 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 derive from whether participants' judgments match experimenter expectations to ascertain the cross-cultural similarity in emotion recognition [12,13,16,23,24,38]. For reasons we outline in Supplementary Discussion 5: Interrater Agreement, we operationalized emotion recognition in terms of inter-rater agreement, a more data driven approach to observer consensus in emotion recognition. Guided by this conceptual approach, we first analyze the combined data from Indian and US participants to verify that we obtained reliable judgments of emotion, one way of determining how many emotions were recognized in the 2519 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 ($q < .05$, Monte Carlo simulation using empirical base rates; see Supplementary Fig. 1 for distribution of interrater agreement rates and Supplementary Methods 3: Testing Category Judgments Proportions). Fifty-six percent of the 2519 speech samples 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 selected by real participants). These levels of interrater agreement are comparable to those documented in past studies of emotion prosody [33]. While 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—one way to address whether the conceptualization of emotional prosody is driven more by categories or affective features. In past studies, cross-cultural similarity has most typically been ascertained by comparing the rates with which members of different cultures label expressive behaviors with the same emotion terms [12,16,23,24,38]. This approach does not capture how members of cultures also use emotion concepts (either emotion categories or affective features) to label non-target expressions in similar fashion, data critical to understanding cultural similarities in how individuals recognize emotion in expressive behavior. 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 2519 samples of emotional prosody. This analysis reveals the extent to which US and Indian participants use the emotion categories and affective features in similar fashion when labeling the meaning of the 2519 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[80] of these mean judgments (see Supplementary Methods 4: Explainable Variance and 5: Testing Cross-Cultural Signal Correlations 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 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 (e.g., "amusement") are psychologically constructed from more basic appraisals of "core affect" (valence, 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, presumably constructed out of more basic affective appraisals [44,61,81]. As evident 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 we considered, including valence and arousal. With cross-cultural signal correlations (r) exceeding .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 US than that of valence ($r = .67$; 90% confidence interval: $.55 < r < .78$). Further, cross-cultural signal correlations were *significantly* greater for anger ($r = .94$; 90% confidence interval: $.83 < r < 1$; $p = .001$, two-tailed bootstrap test; $q < .05$, Benjamini-Hochberg false discovery rate [FDR] correction[82]; see Supplementary Methods 5: Testing Cross-Cultural Signal Correlations) than for valence. Furthermore, many emotion categories were significantly better preserved across cultures than the recognition of arousal ($r = .39$; 90% confidence interval: $.20 < r < .58$), including amusement, anger, contentment, desire, fear, pain, sadness, and positive surprise ($r = .92, .94, .94, .87, .88, .77, .96$, and $.90$; 90% confidence intervals: $.72, .83, .63, .63, .71, .62, .75, 61 < r < 1, 1, 1, 1, 1, .90, 1, 1$; all $p < .01$, $q < .05$, two-tailed bootstrap test). The finding that the recognition of certain emotion categories is better preserved across cultures than that of valence—considered a "basic building block of emotional life" [81]—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 [44,61,81,83].

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 category judgments, rather than affective scale judgments, from one culture to be better predictors of affective scale judgments from another culture. Gathering separate judgments of a full complement of emotion categories and affective scales across 2519 samples of emotional prosody allowed for a rigorous test of this possibility.

To test these 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 Methods 6: Regression Analyses). These analyses ascertain whether emotion category ratings are stronger predictors of affective feature judgments across cultures, or vice-versa, and are represented in 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. It is first of note that emotion categories are better preserved than judgments of affective features across India and the US (C > D; correlation of .80 vs. .59, p = .041, two-tailed bootstrap test; C-D=.21, 90% confidence interval: .041<C-D<.39; see Supplementary Methods 6: Regression Analyses 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 (a) Indian *category* judgments predicted US *affective scale* judgments far better than Indian *affective scale* judgments predicted US *affective scale* judgments (D1 > D; correlation of .80 vs. .59, p=.012, two-tailed bootstrap test; D1-D=.21, 90% confidence interval: .06<D1-D<.35), and that (b) 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 .65 vs. .59, p=.12, two-tailed bootstrap test; D2-D=.050, 90% confidence interval: −.01<D2-D<.12). It is further of note that 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). Based on these results, it is more plausible that judgments of general affective features (valence, arousal, etc.) are psychologically constructed from the recognition of emotion categories (amusement, fear etc.) 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 documented 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. These findings set the stage for addressing our next question: How many *distinct* varieties of

emotion were preserved in the recognition of emotional prosody across the two cultures; that is, what is the dimensionality of the cross-cultural recognition of emotion from prosody? More specifically, we have not yet ruled out whether some of the categories were redundant (e.g., synonyms); for example, there may have been interrater reliability in labeling 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, what is the minimum number of dimensions necessary to account for commonalities in the recognition of emotion across cultures?

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 co-vary across two datasets measuring 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 $\mathrm{Cov}[\mathbf{X}\alpha_i, \mathbf{Y}\alpha_i]$. It shares features of partial least squares correlation analysis [PLSC][84], canonical correlation analysis [CCA][85], 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 ß, maximizing $\mathrm{Cov}[\mathbf{X}\alpha_i, \mathbf{Y}ß_i]$ or $\mathrm{Corr}[\mathbf{X}\alpha_i, \mathbf{Y}ß_i]$, PPCA derives just one, α. The goal here is to find dimensions of recognition common to both datasets X and Y. Our method reduces to PCA when the two datasets to which it is applied are identical, so that the objective becomes $\mathrm{Cov}[\mathbf{X}\alpha_i, \mathbf{X}\alpha_i] = \mathrm{Var}[X\alpha_i]$, but PCA and 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 $\mathrm{Corr}[\mathbf{X_{wh}}\alpha_i, \mathbf{Y_{wh}}\alpha_i]$ rather than the covariance. See Supplementary Methods 7: PPCA for further details and discussion.

Given that we previously found the categories explained the cross-cultural preservation of the affective scales, we applied PPCA to the US and Indian category judgments of the 30 emotions conveyed by the 2519 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 fashion 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[86]. 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 we analyze here. Hence, we established using simulations that they would produce accurate results given the distribution of responses we observe. See Fig. 3 for results of simulations.

As we show in Fig. 4, PPCA revealed that 12 distinct semantic dimensions, or kinds, of emotion, were recognized in prosody and significantly preserved across the US and India participant judgments (Fig. 4b; out-of-sample $r \geq .066$, $q < .001$ across all held-out raters, $q < .05$ across held-out raters from each country individually, ForwardStop sequential FDR corrected[86] one-tailed Wilcoxon signed-rank test[86] [one-tailed because we are only interested in positive cross-cultural correlations]). We thus 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 upper left plot reveals the proportion of variance explained by each dimension [PPC] uncovered by PPCA, in data from each culture; the bottom left plot reveals the proportion of preserved covariance for each dimension, as well as the corresponding correlation and its significance.

Of note, the application of canonical correlation analysis (CCA)[85] also resulted in 12 significant dimensions (out-of-sample $r \geq .049$, $q < .05$ across all held-out raters, $q < .05$ across held-out raters from each country individually, ForwardStop sequential FDR corrected[86] one-tailed Wilcoxon signed-rank test[86]; see Supplementary Methods 7: PPCA), as did the application of correlational PPCA (out-of-sample $r \geq .041$, $q < .001$ across all held-out raters, $q < .05$ across held-out raters from each country individually, ForwardStop sequential FDR corrected[86] one-tailed Wilcoxon signed-rank test[86]). 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 8th dimension emerge when using PCA methods 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 USA 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, 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 US. 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 past studies' conclusions that several emotions can be conveyed across cultures with prosody (anger, contempt, fear, interest, desire, relief, 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 Methods 1: Recording the Speech Samples], 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 are these categories of emotion recognized from prosody distributed within a semantic space? Do they lie within discrete clusters, as predicted by basic emotion theories [40,66–71], or along continuous gradients between emotion categories, as recently documented in our investigation of reported emotional experience [7]? As one can see in Fig. 6, we find that the emotional states conveyed by prosody lie along continuous gradients between categories rather than discrete clusters (as with emotional experience). These gradients between categories are evident when we visualize smooth variations in the categorical judgment profiles of the 2519 speech samples using a method called t-distributed stochastic neighbor embedding, or t-SNE [88]. 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 thus conducted t-SNE 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 Methods 8: Maps for further details. In Fig. 4, t-SNE is used to visualize the smooth gradients between emotion categories, represented in different colors, and the extent to which they are preserved across cultures. To allow further exploration of the categories of emotion signaled 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: $.82 < r^2 < .89$), a fully discrete model (with 12 indicator variables denoting the maximal dimension each speech sample loaded on; i.e., one variety of emotion at a time) predicting 68% of the variance in the affective scales (90% confidence interval: $.65 < r^2 < .71$), 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: $.72 < r^2 < .79$). Both discrete models performed significantly worse than the full 12-dimensional model ($p<.001$, two-tailed bootstrap test; see Supplementary Methods 9: Continuous vs. Discrete Models 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 (e.g., 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 $r = -.21$, $p < .001$, two-tailed bootstrap test, 90% confidence interval: $-.25 < r < -.18$; Spearman's $\rho = -.16$, $p < .001$, two-tailed bootstrap test, 90% confidence interval: $-.19 < \rho < -.12$). 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.

Fig. 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 reveals 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 [30,32,33,89]. Within this theorizing, 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 [10,11,47]. Past work has examined how broad, lower-level acoustic properties (e.g., fundamental frequency) are associated with emotion judgments [30,32,33,90]. To what extent are associations between acoustic features of prosody and emotion category and appraisal feature judgments 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 thinking about 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 center 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 [30,32,33,57]. To interrogate the emotional correlates of these acoustic properties in each culture, we computed them for the 2519 speech samples, correlated them with our 12 principal preserved components (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 Methods 11: Measuring Cross-Cultural Acoustic Correlates for details).

The top row of Fig. 8 presents the correlations between low-level acoustic features and emotion category or affective scale judgments. The bottom row of Fig. 8 presents the extent to which the associations between low-level acoustic features and emotion category or affective scale judgments are preserved across India and the US. As one can see, 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. Namely, the cross-cultural Spearman correlation in acoustic correlates exceeded .95 for 5 of the PPCs and exceeded .8 for all but distress and contempt. By contrast, the correlations between acoustic features and the recognition of valence—considered a "basic building block of emotional life"[81]—were considerably less well preserved across cultures ($\rho = .40$, 90% confidence interval: $.02 < \rho < .76$; significantly lower than ρ for 5 of the 12 PPCs [amusement, anger, awe, desire, and disappointment; $\rho = 1, .99, .99, .99, .95$; $p = .002, .004, .002, .004, .008$], $q < .05$, two-tailed bootstrap test; see Supplementary Methods 11: Measuring Cross-Cultural Acoustic Correlates). 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 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 US and India.

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 (amusement, fear etc.) 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 US and India. Our focus was to test hypotheses related to three properties of this semantic space: its conceptualization, focusing explicitly on how emotion categories and scales of affect contribute to the recognition of emotion in prosody; its dimensionality, or 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 2000 US and Indian participants judged 2,519 prosodically modulated speech samples produced by 100 actors from five distinct English-

speaking cultures. They either judged each prosody sample in terms of which emotion category, from a list of 30, was expressed, or they 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, modeled 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 signaling have focused on 6–8 categories of emotion and relied on either interrater agreement rates [12,13,21,43] or factor analysis [45,49,50] to characterize the recognition of emotion. Applying statistical modeling techniques to judgments of a vast array of stimuli, we uncovered 12 distinct emotions that were recognized in India and the US. Twelve emotions—adoration, amusement, anger, awe, confusion, contempt, desire, disappointment, distress, fear, interest, and sadness—emerged in our analyses, were highly correlated across India and the US, and most were found as well to have distinct acoustic correlates that were also 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—e.g., vocal bursts, 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).

Our second interest was 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 [40,66–71], 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 explain past findings of interrater variability in the perception of emotional signals [12,73,91], suggesting that disagreement across raters in forced-choice rating tasks may reflect the intermediacy of states between categories signaled by expressive behavior as opposed to just the indistinctness of the categories or individual differences in recognition [61]. Furthermore, these findings support a shift from the predominant scientific focus on how discrete patterns of expression, physiology, and neural activation distinguish discrete emotion states [92–96] toward an understanding of the continuous variability and blending together of emotion categories by continuously varying patterns of expression, physiology, and neural activation [39,41,72].

Our final interest was to examine 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 signaling of valence, usually along with arousal, is a core, low-level interpretive process from which specific emotion categories are constructed [44,81]. In contrast to this claim, we find empirically 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, 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 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 present study's methods to other kinds of prosody captured in contexts in which speakers are not directed to communicate specific emotions as in the present study [32,97]. It will be important to study more naturalistic, spontaneous forms of prosody, and the range of emotions such forms of prosody communicate and the potentially broader semantic space of emotion that captures such signals [55,56,58].

We note that the present results pertain to similarities in the recognition of emotional prosody between two English-speaking cultures, the US and India. By examining distinct English-speaking cultures (see Supplementary Discussion 7: Cultural Differences), 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 US 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 US 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 vocal emotion categories. Cordaro and colleagues [23] assessed the recognition of nonverbal vocal bursts targeting 16 emotion categories in a remote culture in Bhutan, finding moderate to strong recognition (around 50% accuracy or greater, with chance = 25%) of more than half of the targeted categories, including 7 of the categories found by the present study to be distinguished from prosody in the US and India (amusement, anger, awe, desire, interest, fear, and sadness). Sauter and colleagues [16] assessed the recognition of 9 emotion categories from vocal bursts among Himba listeners from remote Namibian villages, also reporting moderate to strong recognition (around 75% accuracy or greater, with chance = 50%) of more than half of the targeted categories, 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 (but see [73]). 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. Cowen and Keltner [7] found that the subjective feelings people report in response to viewing a wide range of evocative videos, 2185 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.

2519 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 USA 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 2345 English-speaking participants, 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. 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 [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 analyzed primarily using custom code in Matlab[98] . 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 detailed description of each method, see Supplementary Methods.

Category judgment proportions.

For each speech sample, we computed (1) the proportion of participants who chose each category and (2) 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 (OLS) linear regression. Here, it may be worth acknowledging that methods specialized for sparse data could potentially have produced better prediction correlations. However, this only provides for 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 principal preserved component analysis (PPCA), which has two versions, correlational and covariational. Covariational PPCA maximizes the objective function $\mathrm{Cov}(\mathbf{X}\alpha, \mathbf{Y}\alpha)$ whereas correlational PPCA maximizes the objective function $\mathrm{Corr}(\mathbf{X}[\mathbf{X}^{\mathrm{T}}\mathbf{X}]^{-1/2}\alpha, \mathbf{Y}[\mathbf{Y}^{\mathrm{T}}\mathbf{Y}]^{-1/2}\alpha)$. We tested the significance of each component by applying each version of PPCA in a leave-one-rater-out fashion, determining whether 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[85]. 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 establish that PPCA generates accurate results on randomly simulated data 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-distributed stochastic neighbor embedding (t-SNE). We then assigned a color to each speech sample in the map corresponding to a weighted average of the unique colors of its top two 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 (OLS) regression. For the discrete models, we used OLS 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 US 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 analyze 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements.

We thank Dr. Roman Rosipal for devising a correlational version of PPCA and Frederic Theunissen for providing input regarding acoustic analyses. Research reported in this publication was supported by the U.S. 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.

References

1. Keltner D & Haidt J Social functions of emotions at four levels of analysis. Cogn. Emot 13, 505–521 (1999).
2. Nesse RM Evolutionary explanations of emotions. Hum. Nat 1, 261–289 (1990). [PubMed: 24222085]
3. Campos B, Shiota MN, Keltner D, Gonzaga GC & Goetz JL What is shared, what is different? Core relational themes and expressive displays of eight positive emotions. Cogn Emot 27, 37–52 (2013). [PubMed: 22716231]
4. Oveis C, Spectre A, Smith PK, Liu MY & Keltner D Laughter conveys status. J. Exp. Soc. Psychol 65, 109–115 (2013).
5. Gonzaga GC, Keltner D, Londahl EA & Smith MD Love and the commitment problem in romantic relations and friendship. J. Pers. Soc. Psychol 81, 247–262 (2001). [PubMed: 11519930]
6. ten Brinke L & Adams GS Saving face? When emotion displays during public apologies mitigate damage to organizational performance. Organ. Behav. Hum. Decis. Process 130, 1–12 (2015).
7. Cowen AS & Keltner D Self-report captures 27 distinct categories of emotion bridged by continuous gradients. Proc. Natl. Acad. Sci 201702247 (2017). doi:10.1073/pnas.1702247114
8. Schirmer A & Adolphs R Emotion perception from face, voice, and touch: Comparisons and convergence. Trends in Cognitive Sciences 21, 216–228 (2017). [PubMed: 28173998]
9. Singer T & Lamm C The social neuroscience of empathy. Ann. N. Y. Acad. Sci (2009). doi: 10.1111/j.1749-6632.2009.04418.x
10. Frühholz S, Ceravolo L & Grandjean D Specific brain networks during explicit and implicit decoding of emotional prosody. Cereb. Cortex 22, 1107–1117 (2012). [PubMed: 21750247]
11. Bach DR et al. The effect of appraisal level on processing of emotional prosody in meaningless speech. Neuroimage 42, 919–927 (2008). [PubMed: 18586524]
12. Cordaro DT et al. Universals and cultural variations in 22 emotional expressions across five cultures. Emotion (2017). doi:10.1037/emo0000302
13. Elfenbein HA & Ambady N On the universality and cultural specificity of emotion recognition: A meta-analysis. Psychol. Bull (2002). doi:10.1037/0033-2909.128.2.203
14. Keltner D & Cordaro DT in Emotion Researcher (ed. Scarantino A) (2015).
15. Norenzayan A & Heine SJ Psychological universals: What are they and how can we know? Psychol. Bull (2005). doi:10.1037/0033-2909.131.5.763
16. Sauter DA, Eisner F, Ekman P & Scott SK Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. Proc. Natl. Acad. Sci (2010). doi:10.1073/pnas.0908239106

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 Biol. Sci (2017). doi:10.10/rspb.2017.0990
18. Parr LA, Waller BM & Vick SJ New developments in understanding emotional facial signals in chimpanzees. Curr. Dir. Psychol. Sci (2007). doi:10.1111/j.1467-8721.2007.00487.x
19. Snowdon CT in Handbook of Affective Sciences (2002).
20. Filippi P Emotional and interactional prosody across animal communication systems: A comparative approach to the emergence of language. Frontiers in Psychology 7, (2016).
21. Adolphs R Neural systems for recognizing emotion. Current Opinion in Neurobiology (2002). doi:10.1016/S0959-4388(02)00301-X
22. Russell JA Is there universal recognition of emotion from facial expressions? A review of the cross-cultural studies. Psychol Bull 115, 1 (1994).
23. Cordaro DT, Keltner D, Tshering S, Wangchuk D & Flynn LM The voice conveys emotion in ten globalized cultures and one remote village in Bhutan. Emotion (2016). doi:10.1037/emo0000100
24. Gendron M, Roberson D, van der Vyver JM & Barrett LF Cultural relativity in perceiving emotion from vocalizations. Psychol. Sci 25, 911–920 (2014). [PubMed: 24501109]
25. Hertenstein MJ & Campos JJ The retention effects of an adult's emotional displays on infant behavior. Child Dev 75, 595–613 (2004). [PubMed: 15056208]
26. Juslin PN & Laukka P Communication of Emotions in Vocal Expression and Music Performance: Different Channels, Same Code? Psychological Bulletin 129, 770–814 (2003). [PubMed: 12956543]
27. Keltner D, Tracy J, Sauter DA, Cordaro DC & McNeil G in Handbook of Emotions (ed. Barrett LF, Lewis M, & J. M. H. J.) 467–482 (Guilford Press, 2016).
28. Wu Y, Muentener P & Schulz LE One- to four-year-olds connect diverse positive emotional vocalizations to their probable causes. Proc. Natl. Acad. Sci 201707715 (2017). doi:10.1073/pnas.1707715114
29. Titze IR & Martin DW Principles of Voice Production. J. Acoust. Soc. Am 104, 1148 (1998).
30. Scherer KR & Bänziger T Emotional expression in prosody: a review and an agenda for future research. Proc. Speech Prosody (2004).
31. Mitchell RLC & Ross ED Attitudinal prosody: What we know and directions for future study. Neuroscience and Biobehavioral Reviews (2013). doi:10.1016/j.neubiorev.2013.01.027
32. Hancil S The role of prosody in affective speech (Peter Lang, 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. Personal. Soc. Psychol. Interpers. Relations Gr. Process (2016). doi:10.1037/pspi0000066
34. Nordström H, Laukka P, Thingujam NS, Schubert E & Elfenbein HA Emotion appraisal dimensions inferred from vocal expressions are consistent across cultures: A comparison between Australia and India (2017).
35. Paulmann S & Uskul AK Cross-cultural emotional prosody recognition: Evidence from Chinese and British listeners. Cogn. Emot (2014). doi:10.1080/02699931.2013.812033
36. Scherer KR, Banse R & Wallbott HG Emotion Inferences from Vocal Expression Correlate Across Languages and Cultures. J. Cross. Cult. Psychol (2001). doi:10.1177/0022022101032001009
37. Cowen AS & Keltner D Clarifying the conceptualization, dimensionality, and structure of emotion: Response to Barrett and colleagues. Trends in Cognitive Sciences 22, 274–276 (2018). [PubMed: 29477775]
38. Laukka P et al. Cross-cultural decoding of positive and negative non-linguistic emotion vocalizations. Front. Psychol (2013). doi:10.3389/fpsyg.2013.00353
39. Parr LA, Cohen M & De Waal F Influence of social context on the use of blended and graded facial displays in chimpanzees. Int. J. Primatol (2005). doi:10.1007/s10764-005-0724-z
40. Ekman P in The Nature of Emotion (eds. Ekman& Davidson) 15–19 (Oxford University Press, 1992).

41. Harris RJ, Young AW & Andrews TJ Morphing between expressions dissociates continuous from categorical representations of facial expression in the human brain. Proc. Natl. Acad. Sci (2012). doi:10.1073/pnas.1212207110
42. Keltner D in What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS) (2012). doi:10.1093/acprof:oso/9780195179644.003.0007
43. Russell JA Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. Psychol. Bull (1994). doi:10.1037/0033-2909.115.1.102
44. Russell JA Core affect and the psychological construction of emotion. Psychol Rev 110, 145 (2003). [PubMed: 12529060]
45. Smith CA & Ellsworth PC Patterns of cognitive appraisal in emotion. J Pers Soc Psychol 48, 813 (1985). [PubMed: 3886875]
46. Frijda NH, Kuipers P & ter Schure E Relations among emotion, appraisal, and emotional action readiness. J. Pers. Soc. Psychol (1989). doi:10.1037/0022-3514.57.2.212
47. Scherer KR The dynamic architecture of emotion: Evidence for the component process model. Cogn Emot 23, 1307–1351 (2009).
48. Posner J, Russell JA & Peterson BS The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. Dev Psychopathol 17, 715–734 (2005). [PubMed: 16262989]
49. Russell J A circumplex of affect. J Pers Soc Psychol 36, 1152–1168 (1980).
50. Watson D & Tellegen A Toward a consensual structure of mood. Psychol. Bull (1985). doi: 10.1037/0033-2909.98.2.219
51. Ang J, Dhillon R, Krupski A, Shriberg E & Stolcke A Prosody-based automatic detection of annoyance and frustration in human-computer dialog in ICSLP 2002 - {I}nterspeech 2002. Proceedings of the 7th International Conference on Spoken Language Processing 2037–2040 (2002).
52. 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).
53. Provine RR & Fischer KR Laughing, Smiling, and Talking: Relation to Sleeping and Social Context in Humans. Ethology 83, 295–305 (1989).
54. Vidrascu L & Devillers L Real-life emotion representation and detection in call centers data in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 3784 LNCS, 739–746 (2005).
55. Sauter DA & Fischer AH Can perceivers recognise emotions from spontaneous expressions? Cognition and Emotion 1–12 (2017). doi:10.1080/02699931.2017.1320978 [PubMed: 27788634]
56. Anikin A & Lima CF Perceptual and acoustic differences between authentic and acted nonverbal emotional vocalizations. Quarterly Journal of Experimental Psychology 1–21 (2017). doi: 10.1080/17470218.2016.1270976
57. Scherer KR Vocal markers of emotion: Comparing induction and acting elicitation. Comput. Speech Lang 27, 40–58 (2013).
58. Juslin PN, Laukka P & Bänziger T The Mirror to Our Soul? Comparisons of Spontaneous and Posed Vocal Expression of Emotion. J. Nonverbal Behav 42, (2018).
59. Gupta V, Hanges PJ & Dorfman P Cultural clusters: Methodology and findings. J. World Bus (2002). doi:10.1016/S1090-9516(01)00070-0
60. Jaju A, Kwak H & Zinkhan GM Learning Styles of Undergraduate Business Students: Cross-Cultural Comparison between the US, India, and Korea. Mark. Educ. Rev (2002). doi: 10.1080/10528008.2002.11488787
61. Barrett LF Are emotions natural kinds? Persp Psychol Sci 1, 28–58 (2006).
62. Ekman P What scientists who study emotion agree about. Persp Psychol Sci 11, 31–34 (2016).
63. Ekman P & Cordaro D What is meant by calling emotions basic. Emot Rev 3, 364–370 (2011).
64. Keltner D & Lerner JS in Handbook of Social Psychology (eds. Fiske ST., Gilbert DT. & Lindzey G.) (Wiley Online Library, 2010).

65. Lazarus RS Progress on a cognitive-motivational-relational theory of emotion. Am Psychol 46, 819 (1991). [PubMed: 1928936]
66. Roseman IJ Appraisal determinants of discrete emotions. Cogn Emot 5, 161–200 (1991).
67. Etcoff NL & Magee JJ Categorical perception of facial expressions. Cognition 44, 227–240 (1992). [PubMed: 1424493]
68. Harmon-Jones C, Bastian B & Harmon-Jones E The Discrete Emotions Questionnaire: A new tool for measuring state self-reported emotions. PLoS One 11, 83–111 (2016).
69. Izard CE Four systems for emotion activation: Cognitive and noncognitive processes. Psychol. Rev (1993). doi:10.1037/0033-295X.100.1.68
70. Johnson-Laird PN & Oatley K The Language of Emotions: An Analysis of a Semantic Field. Cogn. Emot (1989). doi:10.1080/02699938908408075
71. Shiota MN et al. Beyond happiness: Toward a science of discrete positive emotions. Am Psychol
72. Samson AC, Kreibig SD, Soderstrom B, Wade AA & Gross JJ Eliciting positive, negative and mixed emotional states: A film library for affective scientists. Cogn. Emot (2016). doi: 10.1080/02699931.2015.1031089
73. Gendron M, Roberson D, van der Vyver JM & Barrett LF Perceptions of emotion from facial expressions are not culturally universal: Evidence from a remote culture. Emotion 14, 251–262 (2014). [PubMed: 24708506]
74. Keltner Dacher, Cordaro DT. Understanding Multimodal Emotional Expression: Recent Advances in Basic Emotion Theory. Emot. Res (2015).
75. Laukka P, Neiberg D & Elfenbein HA Evidence for cultural dialects in vocal emotion expression: Acoustic classification within and across five nations. Emotion (2014). doi:10.1037/a0036048
76. Mehrabian A & Russell J An approach to environmental psychology (T. Press, 1974).
77. Osgood CE Dimensionality of the semantic space for communication via facial expressions. Scand J Psychol 7, 1–30 (1966). [PubMed: 5908205]
78. Sauter DA & Scott SK More than one kind of happiness: Can we recognize vocal expressions of different positive states? Motiv. Emot (2007). doi:10.1007/s11031-007-9065-x
79. Simon-Thomas ER, Keltner DJ, Sauter D, Sinicropi-Yao L & Abramson A The voice conveys specific emotions: Evidence from vocal burst displays. Emotion (2009). doi:10.1037/a0017810
80. Benjamini Y & Yu B The shuffle estimator for explainable variance in FMRI experiments. Ann. Appl. Stat 7, 2007–2033 (2013).
81. Barrett LF Valence is a basic building block of emotional life. Journal of Research in Personality (2006). doi:10.1016/j.jrp.2005.08.006
82. Benjamini Y & Hochberg Y Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B 57, 289–300 (1995).
83. Barrett LF, Lindquist KA & Gendron M Language as context for the perception of emotion. Trends Cogn Sci 11, 327–332 (2007). [PubMed: 17625952]
84. Abdi H, & Williams LJ Partial least squares methods: partial least squares correlation and partial least square regression In Computational Toxicology (pp. 549–579). Humana Press, Totowa, NJ (2013).
85. Hardoon DR, Szedmak S, & Shawe-Taylor J Canonical correlation analysis: An overview with application to learning methods. Neural Computation 16, 12, 2639–2664 (2004). [PubMed: 15516276]
86. Wilcoxon F Individual Comparisons by Ranking Methods. Biometrics Bull 1, 80 (1945).
87. G'Sell MG, Wager S, Chouldechova A & Tibshirani R Sequential selection procedures and false discovery rate control. J. R. Stat. Soc. Ser. B Stat. Methodol 78, 423–444 (2016).
88. Maaten LVD & Hinton G Visualizing data using t-SNE. J Mach Learn Res 9, 2579–2605 (2008).
89. Scherer KR Vocal Affect Expression. A Review and Model for Future Research. Psychological Bulletin 99, 143–165 (1986). [PubMed: 3515381]
90. Ringeval F et al. AV+EC 2015: The First Affect Recognition Challenge Bridging Across Audio, Video, and Physiological Data. Proc. AVEC 2015, Satell. Work. ACM-Multimedia 2015 3–8 (2015). doi:10.1145/2808196.2811642

91. Haidt J & Keltner D Culture and Facial Expression: Open-ended Methods Find More Expressions and a Gradient of Recognition. Cogn. Emot (1999). doi:10.1080/026999399379267

92. Kragel PA & LaBar KS Multivariate neural bioandmarkers of emotional states are categorically distinct. Soc Cogn Affect Neurosci 10, 1437–1448 (2015). [PubMed: 25813790]

93. Kreibig SD Autonomic nervous system activity in emotion: A review. Biol. Psychol 84, 394–421 (2010). [PubMed: 20371374]

94. Lench HC, Flores SA & Bench SW Discrete emotions predict changes in cognition, judgment, experience, behavior, and physiology: a meta-analysis of experimental emotion elicitations. Psychol Bull 137, 834–855 (2011). [PubMed: 21766999]

95. 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). [PubMed: 19929758]

96. Wager TD et al. in Handbook of Emotions (eds. Lewis M., Haviland-Jones JM. & Barrett LF.) 3, 249–271 (Guilford Press New York, NY, 2008).

97. Scherer K & Bänziger T On the use of actor portrayals in research on emotional expression. Bluepr. Affect. Comput. A Sourceb 166–176 (2010).

98. MATLAB. MATLAB *MATLAB* R2012b (2012). doi:10.1201/9781420034950

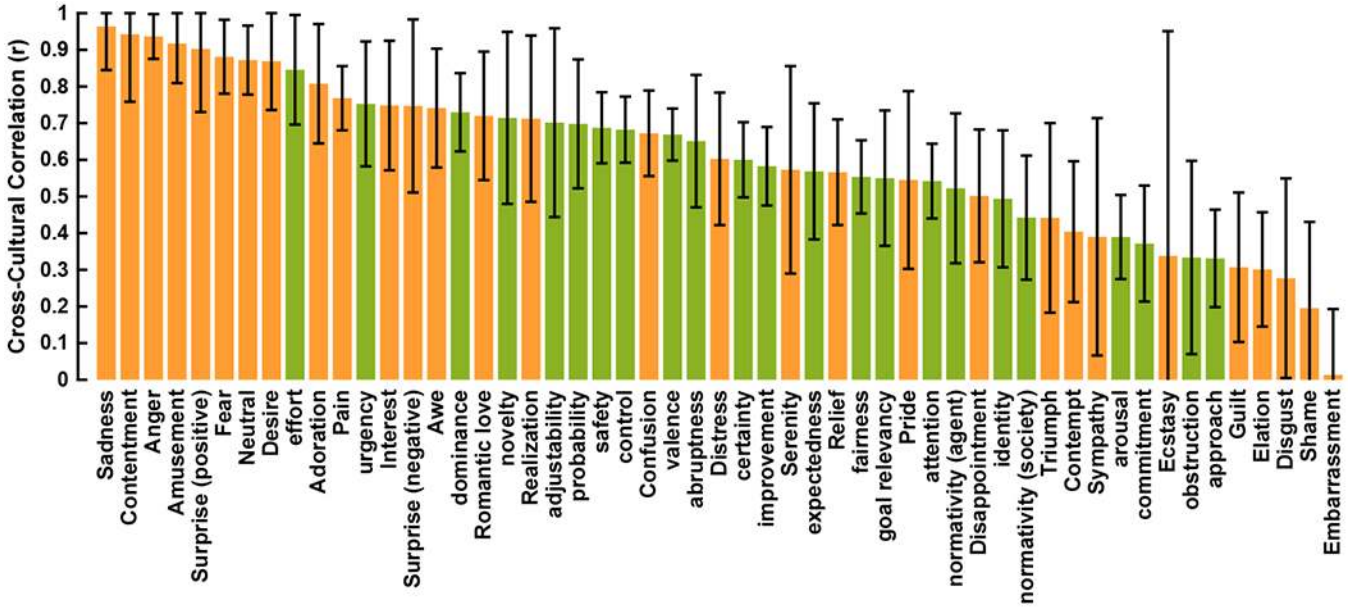

Fig. 1. Correlations in the meaning of emotional prosody across cultures.

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 US across all 2519 speech samples. It is found by correlating the mean responses by Indian participants with the mean responses by US participants across the 2519 speech samples, then dividing by the explainable variance[80] in responses from each culture. Error bars represent standard error estimated by bootstrapping across raters. (For category surveys, participant sample size n_{USA} = 525, n_{India} = 152, and for the two affective scale surveys, n_{USA}=927 and 827 and n_{India}=242 and 205. See Supplementary Methods 4: Explainable Variance and 5: Supplementary Methods 5: Testing Cross-Cultural Signal Correlations for details regarding explainable variance and standard error estimation; 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.)

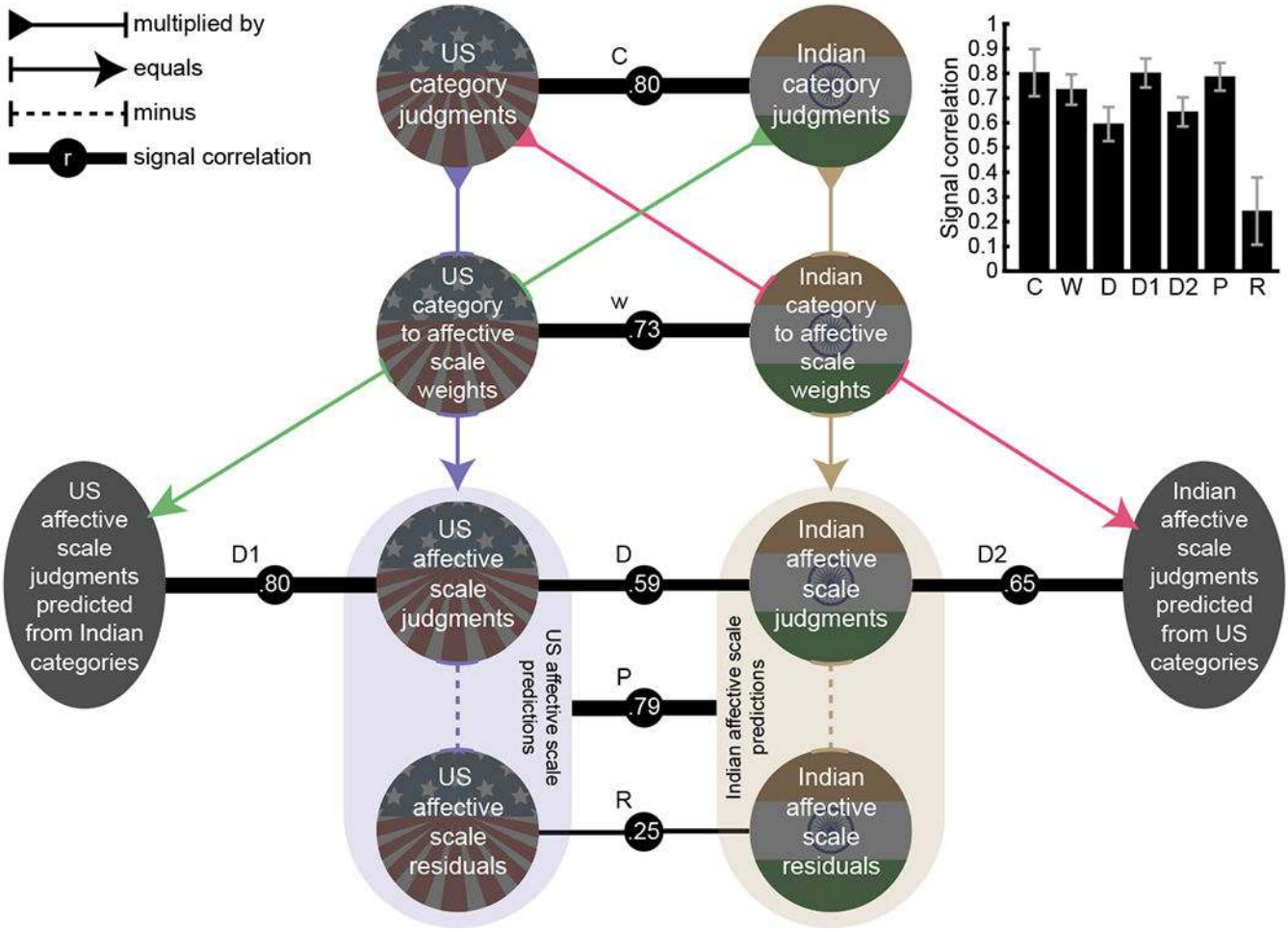

Fig. 2. The preserved recognition of emotion categories accounts for the preservation of affective feature judgments across cultures.

Each circle (or ellipse) represents a matrix of estimators relevant to the recognition of emotion attributes—categories or affective features—in the US (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 at the top left. 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 on the top right. Category judgments are significantly better preserved than affective scale judgments (C>D, p=.041, two-tailed bootstrap test; C-D=.21, 90% confidence interval: .041<C-D<.39). Moreover, category judgments are better than affective scale judgments at predicting affective scale judgments from another culture (D1>D, p=.012, two-tailed bootstrap test; D2>D, p=.12; D1-D=.21, 90% confidence interval: .06<D1-D<.35; D2-D=.050, 90% confidence interval: −.01<D2-D<.12). The variance in the affective scale judgments left over after removing the predictions of the categories is less correlated across cultures (R=.25 between residuals, 90% confidence interval: .024<R<.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 explainable variance[80] for each matrix [see Supplementary Methods 6: Regression Analyses]. Error bars in the top right plot represent standard error estimated by bootstrapping across raters (for category surveys, n_{USA}=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 likely even better explained by the preservation of categorical judgments than indicated here.)

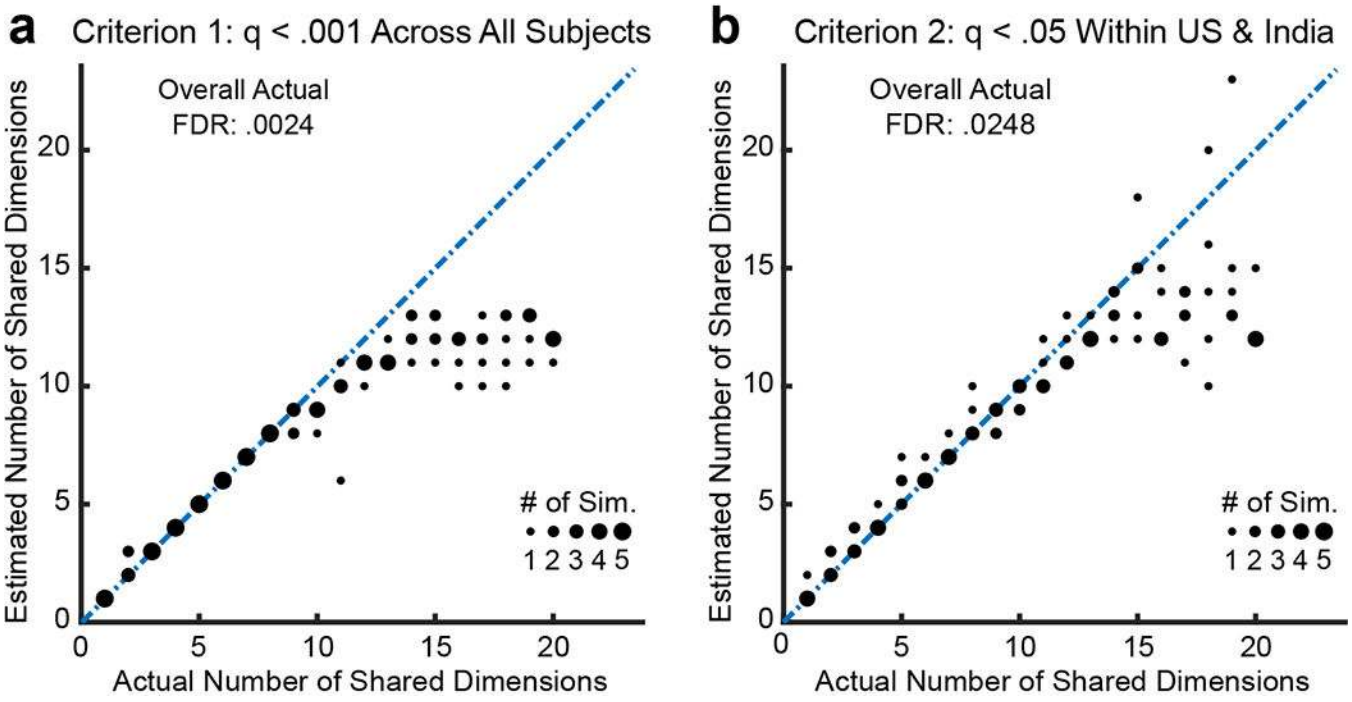

Fig. 3. Verifying that PPCA accurately estimates the number of shared dimensions.

To test whether 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 varying 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 2519×30 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, $\mathbf{X_{USA}}$ and $\mathbf{X_{IND}}$, projecting $\mathbf{X_{USA}}$ 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 probability given by the percentage of ratings each rater actually contributed. Finally, we applied leave-one-rater-out PPCA to determine the p-values for extracted dimensions (Supplementary Methods 7: PPCA). Plotted here are 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:** $q < .001$ across all held-out raters; **b:** $q < .05$ across held-out raters from each country individually; ForwardStop sequential FDR corrected[86] one-tailed Wilcoxon signed-rank test[85]). 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.)

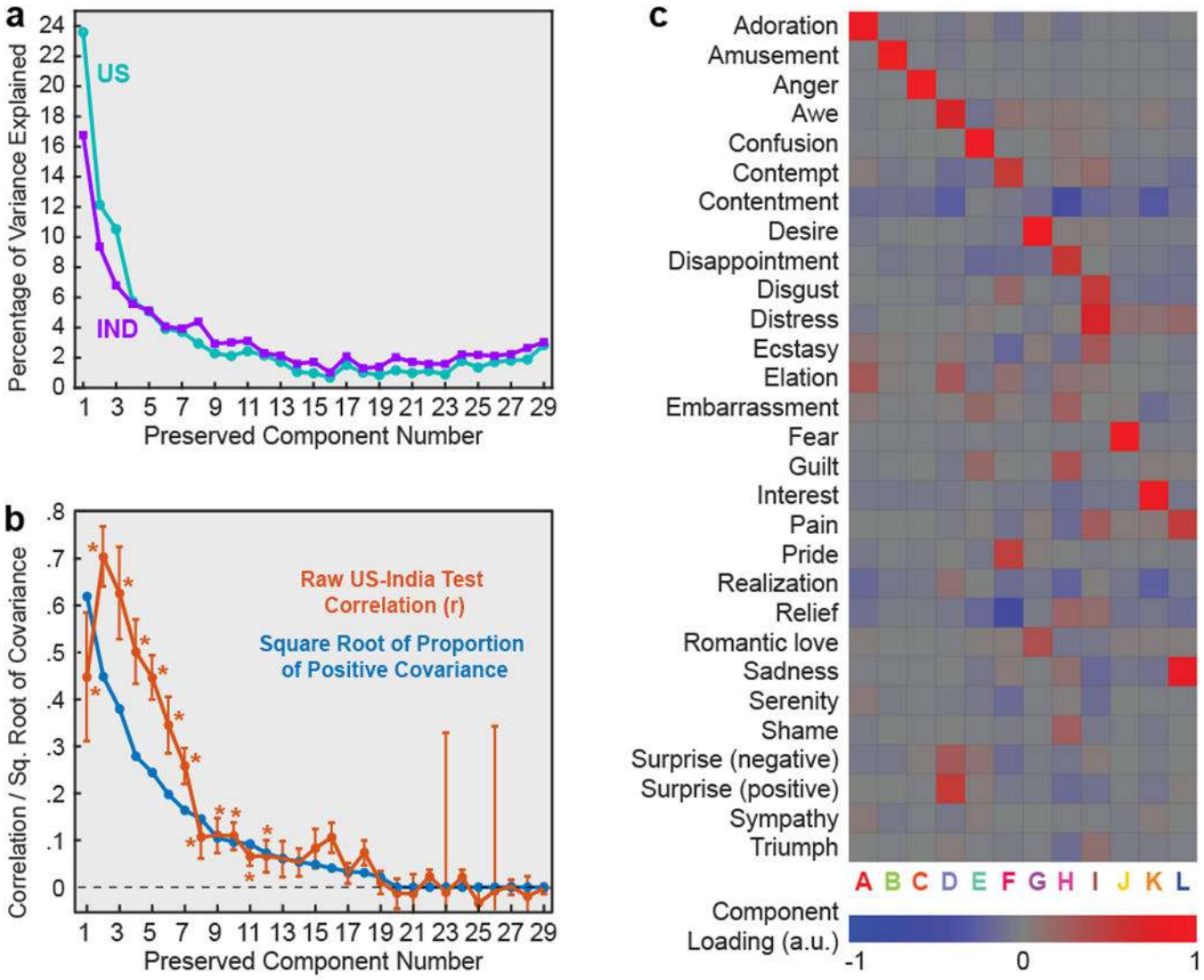

Fig. 4. 12 distinct varieties of emotional prosody are preserved across cultures via category recognition.

(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 Methods 7: PPCA for details). The test correlation was significant for 12 PPCs (out-of-sample $r \geq .066$, $q < .001$ across all held-out raters, $q < .05$ across held-out raters from each country individually, ForwardStop sequential FDR corrected[87] one-tailed Wilcoxon signed-rank test[86]). Error bars represent 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 refer to each component as a distinct category (not to be confused with the raw categorical judgments).

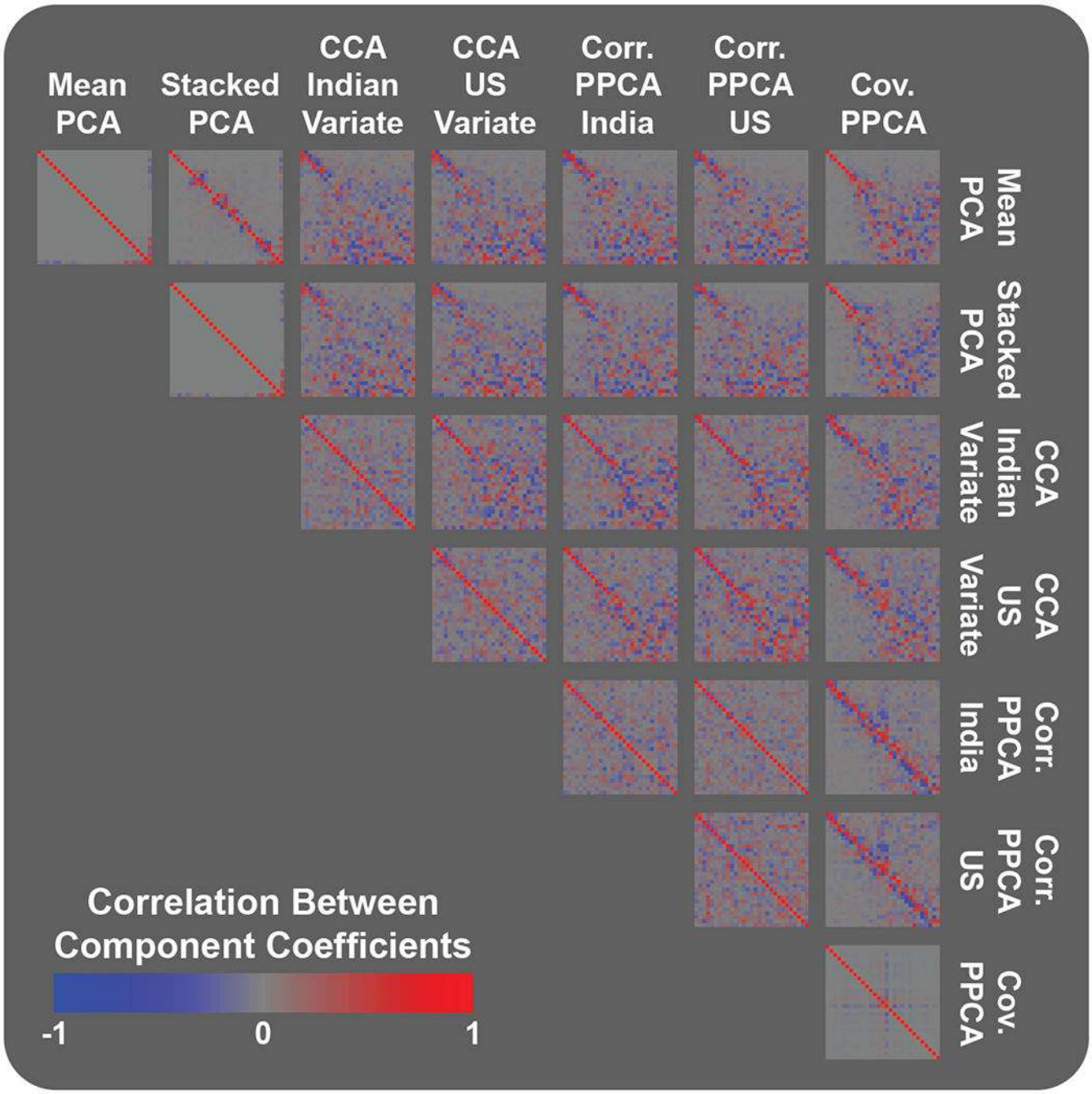

Fig. 5. Correlations between coefficients of components extracted from US and Indian category judgments using different methods.

Each method was used to extract 29 components after excluding the "Neutral" category. The actual component coefficients on 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/Mean PCA, in terms of descending canonical correlation for CCA, and in terms of descending correlation/covariance for correlational/covariational PPCA. Note that CCA extracts an entirely separate

latent space for each culture, and correlational PPCA extracts a slightly modified latent space for each culture ($[\mathbf{X}^{\mathrm{T}}\mathbf{X}]^{-1/2}\alpha_1$ and $[\mathbf{Y}^{\mathrm{T}}\mathbf{Y}]^{-1/2}\alpha_1$). In general, early components (the first seven to ten) are similar across the different methods. The solution derived using covariational PPCA, our primary focus, shares similarities with those derived both by 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.

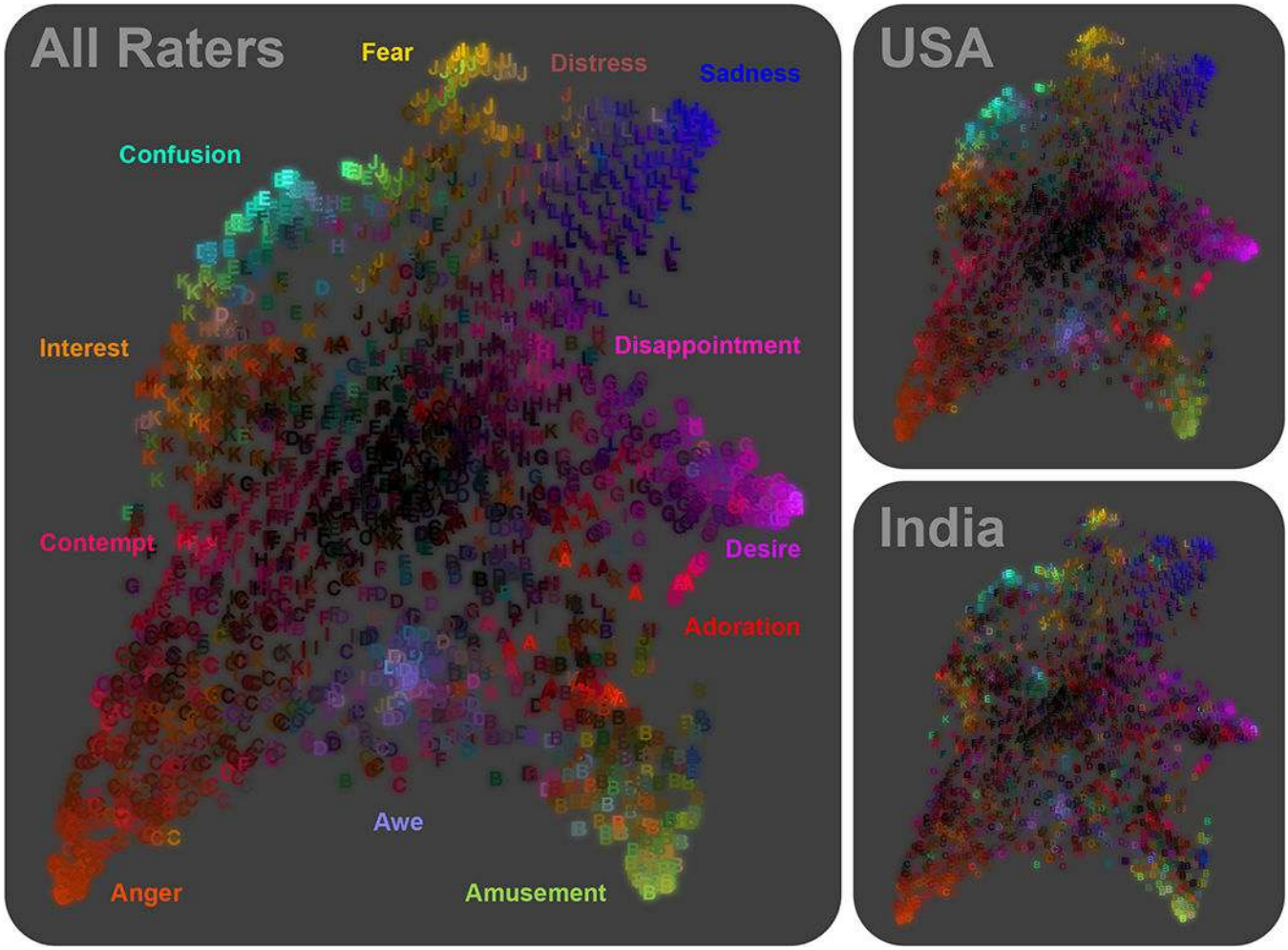

Fig. 6. Visualizing the 12-dimensional structure of emotion conveyed by prosody.

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-distributed stochastic neighbor embedding (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 the data is in fact two-dimensional; see Supplementary Discussion 6: Visualization Along Two Dimensions). The individual speech samples are plotted along these axes as letters that correspond to their highest-loading categorical judgment dimension (with ties broken alphabetically) and are colored using a weighted average of colors corresponding to their scores on the 12 categorical judgment dimensions (see Supplementary Methods 8: Maps for details). The resulting map reveals gradients from amusement to adoration, anger to contempt, and more. For an interactive version of this map, see https://s3-us-west-1.amazonaws.com/venec/map.html. The smaller maps, colored 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 ($q < .05$, Monte Carlo simulation using rates of each category judgment, Benjamini-Hochberg FDR corrected[82]; see Supplementary Methods 10: Testing PPC Scores for details). Categories are often blended together, combining adoration with amusement, anger with contempt, awe with interest, sadness with fear, and more.

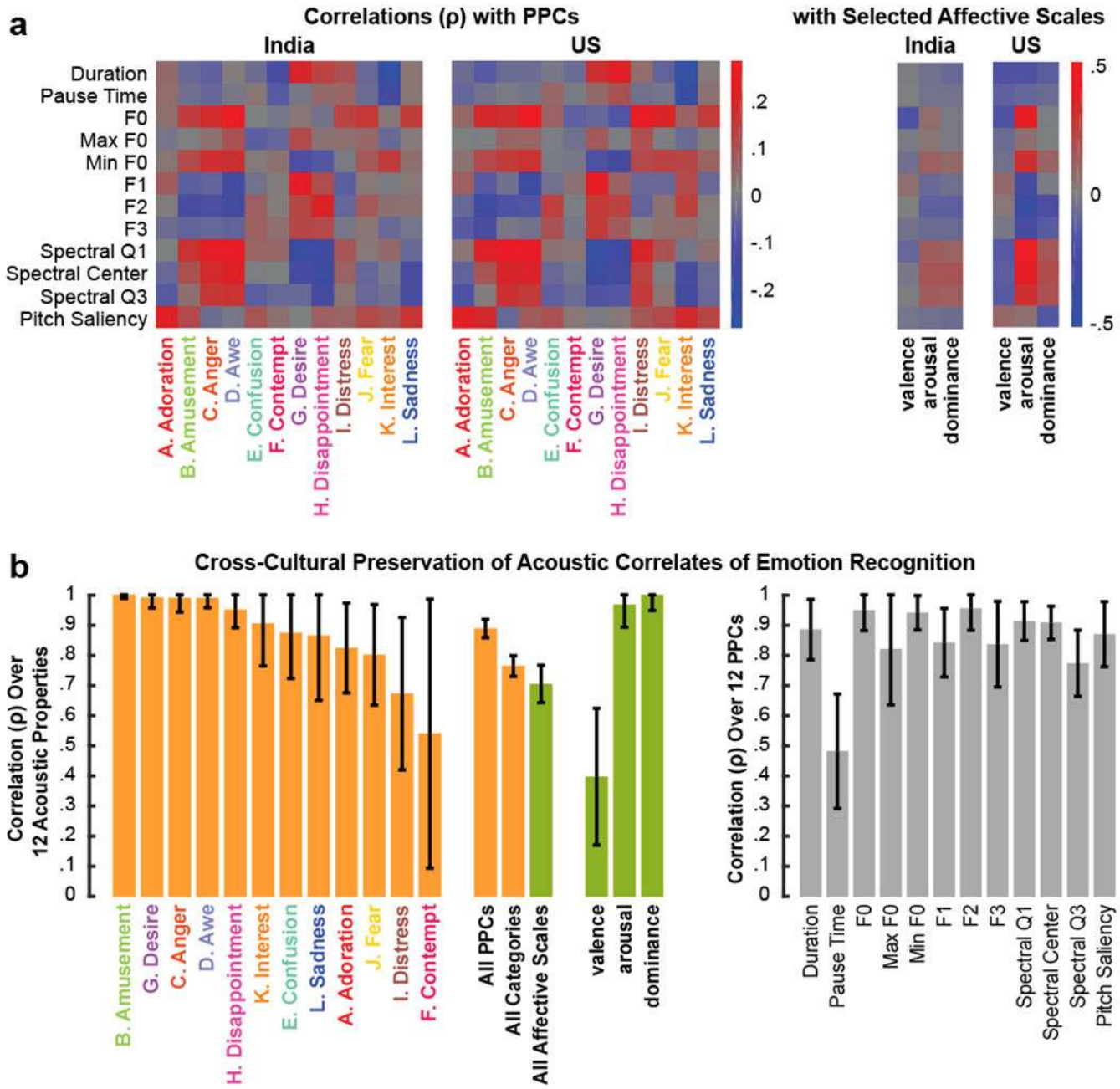

Fig. 8. Low-level acoustic correlates of emotion recognition and their preservation across cultures.

(a) Correlations (ρ) between each acoustic property and judgments from eachculture across the 2519 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 Fig. 8 shows results for every category and dimension. **(b) Cross-cultural signal correlations in acoustic correlates of emotion category and affective feature recognition.** Each colored bar represents the Spearman correlation between a given column of the above acoustic correlation matrices

across cultures (individual dimensions A-F, valence/arousal/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 extremely well preserved across cultures, whereas those of valence are considerably less well preserved. Error bars represent standard error (participant sample size n_{USA} = 525, n_{India} = 152 for emotion categories, and n_{USA}=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 Methods 11: Measuring Cross-Cultural Acoustic Correlates for details.