BRILL

Journal of Cognition and Culture 8 (2008) 135–148

www.brill.nl/jocc

Vocal Emotion Recognition Across Disparate Cultures

Gregory A. Bryant [a, b, *]
H. Clark Barrett [a, c]
[a] FPR-UCLA Center for Culture, Brain, and Development, UCLA,
341 Haines Hall, Box 951553, Los Angeles, CA, 90095-1553, USA
[b] Department of Communication Studies, UCLA,
2303 Rolfe Hall, Los Angeles, CA 90095, USA
[c] Department of Anthropology, Center for Behavior, Evolution, and Culture,
UCLA, 341 Haines Hall, Box 951553, Los Angeles, CA, 90095-1553, USA
*Corresponding author, e-mail: gabryant@ucla.edu

Abstract

There exists substantial cultural variation in how emotions are expressed, but there is also considerable evidence for universal properties in facial and vocal affective expressions. This is the first empirical effort examining the perception of vocal emotional expressions across cultures with little common exposure to sources of emotion stimuli, such as mass media. Shuar hunter-horticulturalists from Amazonian Ecuador were able to reliably identify happy, angry, fearful and sad vocalizations produced by American native English speakers by matching emotional spoken utterances to emotional expressions portrayed in pictured faces. The Shuar performed similarly to English speakers who heard the same utterances in a content-filtered condition. These data support the hypothesis that vocal emotional expressions of basic affective categories manifest themselves in similar ways across quite disparate cultures.

Keywords

vocal emotion, speech, evolutionary psychology, cross-cultural comparisons, universals

Emotional expressions are strategically deployed physical displays with communicative function, and can include bodily gestures, facial movements and vocalizations. Beginning with Darwin, evolutionary biologists have proposed a variety of reasons why emotional expressions might exist in animals, and why they might exhibit features that are universal within a given species. For example, emotional expressions involved in the communication of information about potentially dangerous environmental events, such as the approach of a predator, can be understood in the light of evolutionary theories that explain why animals provide useful information to members of their own species, from inclusive fitness theory to reciprocal altruism to costly signaling

DOI: 10.1163/156770908X289242

theory (Hauser, 1996; Maynard Smith and Harper, 2003). Among the reasons that such displays might have universal properties is that ambiguity could have high fitness costs, resulting in selection for universals in both emotion production and perception.

In the case of humans, it is clear that there is substantial cultural variation in emotional expression (Elfenbein and Ambady, 2002). However, universals of emotional expression likely exist, just as they do in other species. There is considerable evidence of universal facial expressions for basic emotions such as happiness, sadness, anger, fear, surprise, and disgust (Ekman, 1972; Elfenbein and Ambady, 2002). Moreover, research suggests that autonomic nervous system activity differentially patterns across basic emotions (Ekman, 1999a). Despite such evidence, the idea of "basic" emotions has elicited much controversy (e.g., Russell, 1994; Turner and Ortony, 1992). Because every communicative act is a unique and complex event, it can be difficult to decide on criteria for detecting universals among the many dimensions of variation. It can also be problematic to impose discrete category boundaries on continua of expression. However, these problems of theory and measurement do not preclude the co-existence of both universal features and cultural variation in many types of expressions. Ekman (1972) described cultural display rules that underlie the variability between cultures in affective communication. Variations in emotional expressions generally involve rules regarding the types of contexts in which particular expressions are produced, and what kinds of patterned behaviors are associated with what emotions.

Although there is a large literature on universals in emotional expression, this is the first empirical effort examining perception of vocal emotional expressions in a culture with relatively little exposure to common sources of emotion stimuli, such as mass media. In the current set of studies, we demonstrate that vocal emotions can be reliably identified across two quite disparate cultures (American college students and the Shuar, a South American indigenous population), and that the identification can be attributed to global (i.e., suprasegmental) acoustic properties of the utterances. Despite significant differences in the ways and contexts individuals express themselves emotionally in these two cultures, Shuar participants were able to identify basic emotional vocalizations in English, and performed similarly to English speakers who heard the same utterances in a content-filtered condition.

Basic emotions can be recognized in the voice independent of verbal information (see Scherer, 1986), as can more abstract categories of language communication such as sarcasm (Bryant and Fox Tree, 2005). Prosodic features in speech (pitch, loudness, duration, and spectral properties) often form stereotyped configurations that relate in systematic ways to emotional categories (e.g., Cosmides, 1983). For instance, vocalizations conveying happiness tend to be high in average pitch, loud, high in pitch variability, and fast. Sad vocalizations tend to be low average pitch, soft, low in pitch variability and slow (Scherer, 1986). Many studies have described these distinctions through perceptual coding, acoustic analyses, and rating experiments. A recent study showed that even when a set of prototypical emotional vocal expressions were manipulated to create a continuum, listeners' judgments of these stimuli were categorical, suggesting that when inferring emotions from voice, listeners assume discrete categories (Laukka, 2005). Universals in human vocal production have been proposed as well, including universals in infant directed (ID) speech, which serve in part to convey affective information, such as approval or disapproval. Similar acoustic patterns in ID speech have been identified in all cultures studied to date, and are likely universal (Fernald, 1992; Bryant and Barrett, 2007).

Cross-cultural research with adult monolinguals has demonstrated that vocal emotions can be reliably identified across language groups (see Scherer *et al.*, 2001 and Elfenbein and Ambady, 2002 for reviews). Most of this work involves participants listening to voice actors portraying different emotions, matching them with emotional terms, and/or providing judgments on various rating scales. Critics of this research generally complain that the use of forced-choice methodology artificially narrows what would otherwise be quite variable percepts. Nevertheless, participants do reliably match vocal expressions with basic emotion categories, revealing at least some degree of universality across languages in affective vocalization patterns. For example, Scherer *et al.* (2001) found that judges in nine countries reliably recognized four different emotions better than chance from content-free speech (i.e., selected combinations of meaningless syllables from six Indo-European languages). Hit rates and error patterns across emotions were similar across all cultures, though accuracy did decrease as a function of language dissimilarity. Scherer *et al.* (2001) presented these data as evidence for universal inference rules underlying emotion decoding.

Another criticism of this research concerns the overall similarities between experimental subjects across cultures, even across different language groups. Participants are often college-aged students in university settings, and therefore share many traits that diminish the degree to which they should be considered culturally distinct. Similar exposure to television and other western media make findings of universality suspect. Our studies represent the first empirical demonstration of cross-cultural vocal emotion identification in a traditional, subsistence-based, non-Western society (the Shuar of Amazonia; Harner, 1972; German and Barrett, 2005). Exposure to western media is relatively low in this indigenous group. We also avoided the methodological pitfalls associated with using emotional labels by having participants match vocalizations to emotional faces that were specifically imitated in the vocal stimuli. This method has the advantage of presenting a task with a correct answer, as opposed to other judgment studies that ask for ratings of voices that do not necessarily reflect any emotional categories presented. It also has the advantage of allowing subjects to simultaneously assess both face options while listening to a particular voice sample, which would not be possible if we asked them to choose between two voice samples while looking at a particular face.

We predicted that Shuar participants would be able to reliably identify basic emotions in spoken English, and that native English speakers would perform similarly to the Shuar participants on the same task when the vocal stimuli were content-filtered (i.e., no discriminable words). In Experiment 1, native English speakers judged emotional vocalizations produced by English speakers, demonstrating how well native speakers could be expected to perform on this task. We expected performance to be quite high given its relative ease. Experiment 2 repeated Experiment 1, but with the Shuar, hunter-horticulturalists of Amazonian Ecuador. We expected Shuar participants to perform above chance, reflecting the presence of universal emotion cues in speech, but not nearly at the same level as English speakers, who were hearing utterances spoken in their own language. In Experiment 3, native English speakers judged the same utterances as in Experiments 1 and 2, but the items were low-pass filtered so no words were discernable, though global prosodic information (i.e., pitch, loudness, duration and spectral information) was retained. Participants were expected to perform worse than participants in Experiment 1 (due to the degraded stimuli), and perform similarly to Shuar participants in Experiment 2, because both Shuar and English listeners would have access to only global prosodic properties of the stimuli that are known to convey affective information (for full descriptions, see Frick, 1985 and Scherer, 1986). If universals in emotional expression are communicated through such acoustic information, in the absence of semantic content, Shuar and English listeners should both be above chance, and have similar levels of accuracy in judging these utterances.

Method

Participants

In Experiment 1, 28 (12 male and16 female) undergraduates at the University of California, Santa Cruz participated for course credit. All participants were native English speakers. In Experiment 2, the participants were 23 adults (15 male and 8 female; age range 17–63; $M = 32.5$) from two Shuar villages in Morona Santiago and Pastaza provinces, Ecuador. All Shuar participants were taught Spanish as part of a Shuar/Spanish bilingual education program, but Shuar is their first and primary language. In Experiment 3, participants were 30 (15 male and 15 female) University of California, Los Angeles undergraduates who received $5 for participation.

Materials

Twenty simple propositional sentences were created (e.g., "The dog is in the house", "She ate the fish") that were readily translatable to Shuar (for a future study) and relatively emotion-neutral. Two native English speakers (1 male/ 1 female) each produced 5 emotional versions of 10 different sentences. The utterances were produced while trying to imitate emotional expressions portrayed in five different emotional faces (described below). Thus, a total of 100 utterances were created (10 sentences×5 emotions×2 speakers).

The voices were recorded digitally using Cool Edit Pro, Version 2.1 (Syntrillium Software, 2000) on a desktop PC, using an AKG C-535EB condenser microphone and a Mackie 1202-VLZ mixing board. The utterances were transferred to audio cassette in 5 separate lists of 20 items each. The items were counterbalanced so that all participants heard all twenty utterances, an equal number of male and female voices, and an equal number of items from each emotional category. Vocal stimuli in Experiments 1 and 2 were presented on a portable cassette recorder with a built-in speaker. This device was used in the field for practical purposes, and so was used in Experiment 1 for methodological consistency. In Experiment 3, a portable CD player with built-in speakers was used (see below).

For Experiment 3, all utterances were content-filtered in order to remove the words while retaining global prosodic information. The 20 utterances were low-pass filtered at 450 Hz (all frequencies above 450 Hz were reduced with a 60 dB/octave roll-off) using the scientific filter function in Cool Edit Pro. In problem areas where words could be reliably identified by naïve listeners, lower cut off values (all > 250 Hz) were used. Pitch, loudness, and duration properties remained, but acoustic information (i.e., high frequency information) crucial for understanding words was eliminated. The utterances were normed with a group of participants from the same pool as the main experimental participants to ensure they could not be understood. The stimuli were transferred to CD and presented to participants on a portable CD player with speakers (as opposed to cassette) in order to accommodate the degraded audio quality.

Face stimuli were obtained from NimStim face stimulus set developed by The Research Network on Early Experience and Brain Development. The stimulus set has undergone preliminary validation indicating high agreement across children and adult raters (Tottenham *et al.*, 2002). Five emotional faces (angry, happy, sad, fearful and disgust) were used, all produced by the same individual. The faces were printed in black and white, approximately 4' × 5' in size, and laminated.

Procedure

Participants were told they were taking part in a study examining emotions and the voice. It was explained that they would be listening to voices that were produced while the speaker was looking at one of five emotional faces. During the instructions, all faces were presented at once. Participants were then described the task of listening to each sentence one time, and choosing one of two presented faces. They were reminded that one of the presented faces was the face the speaker was attempting to imitate when he/she recorded the utterance. Each participant received two practice trials, and then twenty experimental trials. Face presentation was counterbalanced for orientation (right/left) and pairing (equal number of pairings between all categories). All Shuar participants were instructed in Spanish. In Experiment 3, participants were asked to report any words they understood in the utterances, and no correct identifications were made. After all trials, participants were asked to identify the emotion portrayed in each face, and were then debriefed.

Results

Performance was expected to be highest in Experiment 1 (English speakers judging unfiltered English utterances), and lowest in Experiment 2 (Shuar speakers judging unfiltered English utterances. In Experiment 3 (English speakers judging content-filtered English utterances), performance was expected to be worse than participants in Experiment 1, and comparable to participants in Experiment 2 because both groups were reliant exclusively on global prosodic information.

For all three experiments we conducted logistic regression analyses to test whether overall hit rates (number of correct judgments divided by total number of judgments for each subject) and hit rates for each emotion category across subjects within each experiment differed significantly from chance. Because our dependent variable (hit rate) is binary, we used logistic regression as an alternative to a linear statistical model. All chi square tests are Wald chi square statistics.

We ran a logistic regression predicting response from condition, testing for differences in performance between experiments. The overall model was significant, χ^2 (2, n = 81) = 154.86, P<0.001. As predicted, participants in all experiments were able to reliably match the emotional voices to their corresponding faces better than chance (see Fig. 1). In Experiment 1, participants identified the correct face 80% of the time (mean hit rate = 0.80, SD = 0.079), which was significant, χ^2 (1, *n* = 81) = 220.82, *P*<0.001. In Experiment 2, participants identified the correct face 60% of the time (mean hit rate = 0.60, SD = 0.079), which was also significantly better than chance, χ^2 (1, *n* = 81) = 100.0, *P*<0.001. In Experiment 3, participants identified the correct face 58% of the time (mean hit rate = 0.58, SD = 0.11) and this result was also significant, χ^2 (1, *n* = 81) = 113.42, *P*<0.001. Planned comparisons revealed a significant difference in performance between Experiment 1 and Experiment 2, χ^2 (1, *n* = 81) = 100.06, *P*<0.001, between Experiment 1 and Experiment 3, χ^2 (1, *n* = 81) = 113.45, *P*<0.001; but not between Experiment 2 and Experiment 3, χ^2 (1, *n* = 81) = 0.41, *P* = ns.

Following these analyses of global hit rates for each experiment, we examined performance for each emotion category within each experiment. Tables 1 through 3 give confusion matrices for each experiment. Confusion matrices show how subject judgments were distributed between correct and incorrect categories for every combination of paired categories. In addition, hit rates (number of times an emotion was correctly selected divided by the number of paired comparisons in which it was available as a correct choice), false alarm rates (number of times an emotion was incorrectly selected divided by the number of paired comparisons in which it was available as an incorrect choice), and d-prime (*Z*-transformed hit rates minus *Z*-transformed false alarms rates) are shown at the bottom of each table.

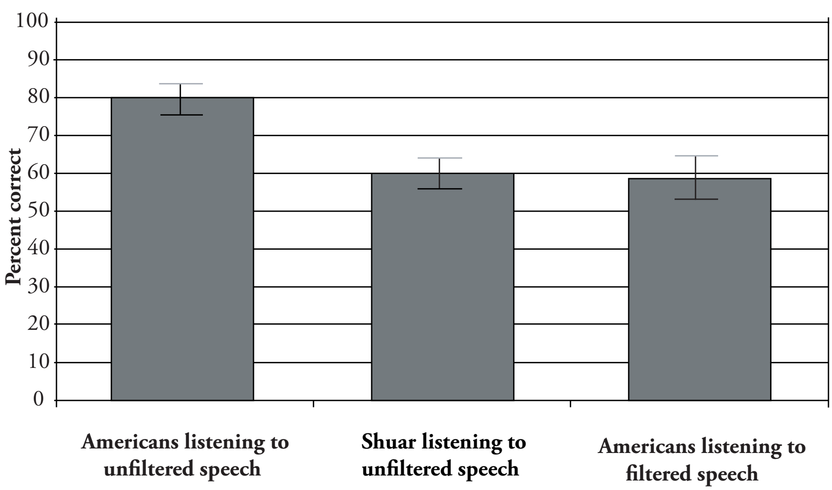

Figure 1. Mean percentages and standard error of correct responses identifying vocal emotion across three experiments.

Table 1
Confusion matrix, hit rates and false alarm rates for Experiment 1 (Americans listening to unfiltered speech)

True category	Selected category					
	Angry	Fear	Happiness	Sadness	Disgust	Total
Angry	**104**	0	7	0	1	112
Fear	1	**95**	6	1	9	112
Happiness	11	10	**86**	2	3	112
Sadness	2	3	0	**107**	0	112
Disgust	18	11	3	22	**58**	112
Total	136	119	102	132	71	560
Hit rate (%)	93*	85*	77*	96*	52*	80*
False alarm rate (%)	29	21	14	22	12	20
d-prime	2.03	1.82	1.80	2.46	1.24	1.70

Tables 1–3 show the number of times that a given emotion was selected (across the top) for a trajectory that had been generated with a particular original true emotion (down the left side). The main diagonal (in bold) thus presents correct categorizations. Responses are pooled across participants. The significance tests are all Wald chi squares. *$P<0.01$.

Table 2
Confusion matrix, hit rates and false alarm rates for Experiment 2 (Shuar listening to unfiltered speech)

True category	Selected category					
	Angry	Fear	Happiness	Sadness	Disgust	Total
Angry	**60**	10	6	10	6	92
Fear	7	**58**	8	6	13	92
Happyiness	2	15	**65**	8	2	92
Sadness	6	10	12	**57**	7	92
Disgust	15	9	13	18	**37**	92
Total	90	102	104	99	65	460
Hit rate (%)	65*	63**	71*	62**	40	60**
False alarm rate (%)	33	48	42	46	30	40
d-prime	0.84	0.39	0.74	0.41	0.82	0.52

* $P<0.01$, ** $P<0.05$.

Table 3
Confusion matrix, hit rates and false alarm rates for Experiment 3 (Americans listening to filtered speech)

True category	Selected category					
	Angry	Fear	Happiness	Sadness	Disgust	Total
Angry	**61**	16	8	21	14	120
Fear	8	**70**	18	11	13	120
Happiness	12	16	**64**	12	16	120
Sadness	9	6	9	**83**	13	120
Disgust	9	9	12	17	**73**	120
Total	99	117	111	144	129	600
Hit rate (%)	51	58†	53	69*	61**	59*
False alarm rate (%)	32	39	39	51	47	41
d-prime	0.50	0.49	0.36	0.48	0.36	0.43

* $P<0.01$, ** $P<0.05$, † $P<0.10$.

Hit Rate Analyses

For each experiment, we analyzed hit rates for each emotion category (see Tables 1–3). We generated coefficient (β) values for hit rates in each of the emotion categories in all three experiments, and used Wald chi square tests to determine whether these β values differed significantly differed from zero ($\beta = 0$ is equivalent to $p(\text{hit}) = 0.5$). In Experiment 1 (American unfiltered), all categories were recognized better than chance except disgust: Anger: χ^2 $(1, n = 81) = 48.86$, $P<0.001$; Fear: χ^2 $(1, n = 81) = 42.64$, $P<0.001$; Happiness: χ^2 $(1, n = 81) = 28.62$, $P<0.001$; Sadness: χ^2 $(1, n = 81) = 44.89$, $P<0.001$; Disgust: χ^2 $(1, n = 81) = 0.14$, P = ns. Experiment 2 (Shuar unfiltered) had a similar pattern, with all categories recognized significantly better than chance, although disgust was marginally significant: Anger: χ^2 $(1, n = 81) = 12.98$, $P<0.01$; Fear: χ^2 $(1, n = 81) = 6.1$, $P<0.05$; Happiness: χ^2 $(1, n = 81) = 14.75$, $P<0.001$; Sadness: χ^2 $(1, n = 81) = 5.15$, $P<0.05$; Disgust: χ^2 $(1, n = 81) = 3.46$, $P = 0.06$. In Experiment 3 (American filtered), sadness and disgust were recognized significantly better than chance, and fear was marginally significant: Anger: χ^2 $(1, n = 81) = 0.03$, P = ns.; Fear: χ^2 $(1, n = 81) = 3.31$, $P = 0.07$; Happiness: χ^2 $(1, n = 81) = 0.53$, P = ns.; Sadness: χ^2 $(1, n = 81) = 16.73$, $P<0.001$; Disgust: χ^2 $(1, n = 81) = 5.52$, $P<0.05$.

To test for differences between emotion categories within each experiment, we used each of the five emotion categories as the reference group in five separate logistic regression analyses to generate coefficient (β) values for every comparison within each experiment. Wald chi squares were used to check if these β values differed significantly from zero (i.e., $p(\text{hit})_{\text{Emotion x}} = p(\text{hit})_{\text{Emotion y}}$) (see Table 4).

Table 4
Differences between emotion categories across three experiments

Emotion pairs	Experiment 1			Experiment 2			Experiment 3		
	β	χ^2	*P*	β	χ^2	*P*	β	χ^2	*P*
Anger-Fear	−0.844	5.20	0.02	−0.095	0.12	0.72	0.303	1.33	0.25
Anger-Happy	−1.37	12.0	0.001	0.250	1.0	0.32	0.10	0.18	0.66
Anger-Sad	0.498	0.64	0.42	−0.141	0.28	0.59	0.775	10.1	0.001
Anger-Disgust	−2.49	24.1	0.000	−1.03	15.5	0.000	0.407	2.22	0.14
Fear-Happy	−0.524	2.79	0.10	0.344	1.21	0.27	−0.203	0.325	0.57
Fear-Sad	1.34	4.63	0.03	−0.046	0.32	0.86	0.471	2.16	0.14
Fear-Disgust	−1.65	25.4	0.000	−0.93	12.5	0.000	0.104	0.10	0.75
Happy-Sad	1.87	12.11	0.000	−0.391	1.85	0.18	0.674	6.86	0.009
Happy-Disgust	−1.12	10.82	0.001	−1.28	20.61	0.000	0.307	1.08	0.30
Sad-Disgust	−2.99	27.46	0.000	−0.884	10.43	0.000	−0.368	1.46	0.23

Because we had no a priori directional predictions for these comparisons, we provide these results as an exploratory analysis, and make no particular claims about why the data pattern as they do. In Experiment 1 (American unfiltered), Anger and Sadness were recognized better than the other categories, but did not significantly differ from one another, and Disgust was recognized significantly worse than all other categories. Fear and Happiness were not reliably different from one another. In Experiment 2 (Shuar unfiltered), Disgust was also recognized at a significantly lower rate than all other categories. But the other four categories were not recognized significantly different from one another. In Experiment 3 (American filtered), Sadness was recognized significantly better than Happiness and Anger, but no other differences between categories were reliable.

Discussion

Emotional expressions are designed to transmit unambiguous affective information between conspecifics and, thus, in humans, these expressions should exhibit some universal characteristics identifiable across languages and cultures. Much of the evidence for universals in emotional expression comes from research on facial expressions and, to a lesser degree, voices. Most of the cross-cultural work on vocal emotions has been done using adult populations from groups with exposure to common sources of emotional stimuli, such as mass media. The exception to this is cross-linguistic research examining universals in the production and perception of ID speech in mothers and infants (reviewed in Fernald, 1992) and the perception of ID speech in Shuar adults (Bryant and Barrett, 2007). A more rigorous test of the prediction that basic emotions are expressed, to some degree, in a universal manner in adult-directed speech, involves the examination of people in traditional societies that have relatively less exposure to western media and people. This research is the first demonstration, to our knowledge, of cross-cultural vocal emotion recognition in a traditional society.[1] Shuar individuals from Amazonian Ecuador were able to reliably identify happy, angry, fearful, and sad vocalizations produced by American native English speakers by matching emotional spoken utterances to emotional expressions portrayed in pictured faces. These data lend solid support to the hypothesis that vocal emotional expressions of basic affective categories manifest themselves in similar ways across quite disparate cultures.

Against this general finding of ability to recognize emotion in the voice cross-culturally, we found generally poor performance in recognizing disgust, even by native English speakers hearing unfiltered English speech (Experiment 1). Previous studies have found that subjects can have difficulty distinguishing disgust expressions from other emotions, such as surprise, anger, and contempt (Haidt and Keltner, 1999). In this case we suspect that the problem might be that the disgust face stimulus was poor. None of the Shuar participants assigned the label disgust to the disgust face, and only 25% of the native English speakers did so. Moreover, many participants across all three studies thought the portrayal was "crying" or "upset" rather than disgust. These results suggest that our disgust face was not a good representative of the stereotyped configuration documented for disgust (Ekman, 1999b). Additionally, many of the pairings of disgust as a true category paired with other emotions elicited high false alarm rates suggesting an aversion to picking the disgust face. For these reasons, conclusions about disgust recognition on the basis of this study should remain tentative. However, this does not impact the results for the other four emotion categories, for which we found no problems with the face stimuli.

Aside from poor recognition of disgust, the analyses of performance for individual emotion categories revealed other similarities and differences between

[1] Recent preliminary findings exploring the perception of happiness in the voice with the Himba in Southern Africa are promising (Sauter *et al.*, 2006).

the experiments. Curiously, Americans listening to the same utterances content-filtered were better able to identify disgust than participants in Experiment 1 who heard them unfiltered. By removing words, the filtering process can make certain acoustic distinctions easier to notice that are otherwise obscured by lexical information, especially in decontextualized utterances (Bryant and Fox Tree, 2005). For example, the distinction between angry and happy vocalizations was easier for American participants in the filtered condition (relative to other distinctions from anger), but more difficult for American participants in the unfiltered condition. The words might have obscured the subtle acoustic distinctions between these types of vocal expressions.

One interesting culture-specific phenomenon observable in these data was that Shuar participants often chose the fearful face when listening to a happy utterance. But the error did not occur in reverse, that is, they did not tend to associate happy faces with fear utterances. Overall for the Shuar, happiness was the best identified emotion but this pattern disappeared when paired with fear. This makes sense in that happiness and fear do share many acoustic features, such as high pitch, fast tempo, and high pitch variability (Scherer, 1986). But display rules governing the expression of happiness in Shuar culture might have also contributed to confusions between these categories of vocal emotion (in traditional Shuar culture, for example, verbal exchanges between strangers are often serious and formal in tone, sometimes implicitly or explicitly aggressive, which might influence the way emotions like fear and happiness are expressed; see Juncosa, 1999). Moreover, Shuar participants had relatively low hit rates and high false alarm rates for fear. Further research might explore the degree to which this confusion is the result of culturally-specific expression patterns and/or acoustic similarity effects.

These results also shed light on what acoustic features contribute to people's judgments of emotional speech. Discriminability measures indicated that native English speakers listening to utterances that were content-filtered performed similarly to Shuar participants that heard the utterances unfiltered, although differences were apparent across the different emotional categories. When specific linguistic information is removed, but prosodic information is retained, native speakers might use a fair amount of the same information for making emotional judgments that non-speakers use. The information remaining after low-pass filtering is global. Low-pass filtering retains prosodic information that acts across whole utterances, as opposed to local prosodic cues that act on smaller segments for making linguistic distinctions. There is substantial evidence that these types of prosody are processed in different brain regions (see Baum and Pell, 2000), and are produced by functionally distinct vocal physiology systems (McRoberts *et al.*, 1995).

We stress, again, that because emotional expression is a complex and multi-faceted phenomenon, cultural universals and cultural variation in emotion expression and perception coexist. However, despite the cultural variation that we see, and all of the forces serving to maintain differences between cultures, it is possible for individuals to communicate across fairly wide cultural boundaries, such as that separating American college students from Shuar hunter-horticulturalists in the Ecuadorian Amazon. There may be many other such universals of human communication waiting to be discovered.

Acknowledgements

This project was made possible by a research grant from the UCLA-FPR Center for Culture, Brain, and Development. Thanks to Valerie Edmon, Ray Gibbs, Kevin Haley, Michael Mitchell, Christina Moya, John Schumann, Muneera Shariff, Azadeh Ramazani, and the UCLA Experimental Biological Anthropology Lab. We especially thank the Shuar communities of Chapints, Chinimpi and Shauk for their cooperation, and special thanks to Nicole Wilson for her invaluable assistance in the field.

References

Baum, S. R. and Pell, M. D. (1999). The neural bases of prosody: Insights from lesion studies and neuroimaging. *Aphasiology* 13, 581-608.

Bryant, G. A. and Barrett, H. C. (2007). Recognizing intentions in infant-directed speech: Evidence for universals. *Psychological Science* 18, 746-751.

Bryant, G. A. and Fox Tree, J. E. (2005). Is there an ironic tone of voice? *Language and Speech* 48, 257-277.

Cosmides, L. (1983). Invariances in the acoustic expression of emotion during speech. *Journal of Experimental Psychology: Human Perception and Performance* 9, 864-881.

Elfenbein, H. A. and Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin* 128, 208-235.

Ekman, P. (1972). Universals and cultural differences in facial expressions of emotion, in J. Cole (Ed.), *Nebraska Symposium on Motivation, 1971*, volume 19, pp. 207-282. University of Nebraska Press, Lincoln, NE.

—— (1999a). Basic emotions, in T. Dalgleish and M. Power (Eds.), *Handbook of Cognition and Emotion.* Wiley, Chichester.

—— (1999b). Facial expressions, in T. Dalgleish and M. Power (Eds.), *Handbook of Cognition and Emotion.* Wiley, Chichester.

Fernald, A. (1992). Human maternal vocalizations as biologically relevant signals: An evolutionary perspective, in J. H. Barkow, L. Cosmides and J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture*, pp. 391-428. Oxford University Press, New York, NY.

German, T. P. and Barrett, H. C. (2005). Functional fixedness in a technologically sparse culture. *Psychological Science* 16, 1-5.

Haidt, J. and Keltner, D. (1999). Culture and facial expression: Open-ended methods find more expressions and a gradient of recognition. *Cognition and Emotion* 13, 225-266.

Harner, M. J. (1972). *The Jivaro: People of the sacred waterfalls*. University of California Press, Berkeley, CA.

Hauser, M. D. (1996). *The evolution of communication*. MIT Press, Cambridge, MA.

Juncosa, J. E. (1999). *Etnografía de la comunicación verbal Shuar.* Ediciónes Abya-Yala, Quito.

Laukka, P. (2005). Categorical perception of vocal emotion expressions. *Emotion* 5, 277-295.

Maynard Smith, J. and Harper, D. (2003) *Animal signals*. Oxford University Press, Oxford.

McRoberts, G. W., Studdert-Kennedy, M. and Shankweiler, D. P. (1995). The role of fundamental frequency in signaling linguistic stress and affect: Evidence for a dissociation. *Perception and Psychophysics* 57, 159-174.

Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological Bulletin* 115, 102-141.

Sauter D., Wiland J., Warren J., Eisner F., Calder A. and Scott S. K. (2005). Sounds of joy: An investigation of vocal expressions of positive emotions. *Journal of Cognitive Neuroscience* (Suppl.), S 61.

Scherer. K. R. (1986). Vocal affect expression: A review and model for future research. *Psychological Bulletin* 99, 143-165.

Scherer, K. R., Banse, R. and Wallbott, H. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology* 32, 76-92.

Tottenham, N., Borscheid, A., Ellertsen, K., Marcus, D. J. and Nelson, C. A. (2002). *Categorization of facial expressions in children and adults: Establishing a larger stimulus set.* Poster presented at the Cognitive Neuroscience Society annual meeting, San Francisco, CA.

Turner T. J. and Ortony, A. (1992). Basic emotions: can conflicting criteria converge? *Psychological Review* 99, 566-571.