J Nonverbal Behav (2009) 33:107–120
DOI 10.1007/s10919-008-0065-7

ORIGINAL PAPER

Recognizing Emotions in a Foreign Language

Marc D. Pell · Laura Monetta · Silke Paulmann · Sonja A. Kotz

Published online: 21 January 2009

Abstract Expressions of basic emotions (joy, sadness, anger, fear, disgust) can be recognized pan-culturally from the face and it is assumed that these emotions can be recognized from a speaker's voice, regardless of an individual's culture or linguistic ability. Here, we compared how monolingual speakers of Argentine Spanish recognize basic emotions from pseudo-utterances ("nonsense speech") produced in their native language and in three foreign languages (English, German, Arabic). Results indicated that vocal expressions of basic emotions could be decoded in each language condition at accuracy levels exceeding chance, although Spanish listeners performed significantly better overall in their native language ("in-group advantage"). Our findings argue that the ability to understand vocally-expressed emotions in speech is partly independent of linguistic ability and involves universal principles, although this ability is also shaped by linguistic and cultural variables.

Keywords Emotional speech processing · Affective prosody · Vocal expression · Cultural factors · Cross-linguistic group study

Introduction

Human expressive behaviors which communicate joy, anger, disgust, sadness, and fear are thought to possess certain invariant properties which allow them to be recognized independent of culture and learning (Ekman and Friesen 1971; Izard 1977). As highlighted by a recent meta-analysis (Elfenbein and Ambady 2002), the major source of evidence for 'universality' in the recognition of emotional displays has been cross-cultural studies of the

M. D. Pell (✉) · L. Monetta · S. Paulmann
School of Communication Sciences and Disorders, McGill University,
1266, ave. des Pins Ouest, Montréal, QC H3G 1A8, Canada
e-mail: marc.pell@mcgill.ca
URL: www.mcgill.ca/pell_lab

S. A. Kotz
Research Group on Neurocognition of Rhythm in Communication,
Max Planck Institute of Human Cognitive and Brain Sciences, Leipzig, Germany

face; this work demonstrates that individuals from different cultural backgrounds (e.g., Western/non-Western; literate/pre-literate) recognize basic emotions at levels well exceeding chance when presented in a forced-choice response format (see Russell 1994 for a detailed discussion of this literature). For many researchers, these data argue that processes for recognizing emotional displays involve universal principles which are innately shared across human cultures, linked to the shared neurophysiological consequences of experiencing basic emotions (Ekman 1992; Izard 1994).

While less abundant, supporting evidence of cross-cultural agreement in how basic emotions are recognized from a speaker's *vocal* expressions has been reported (Albas et al. 1976; Scherer et al. 2001; Thompson and Balkwill 2006; Van Bezooijen et al. 1983). During speech communication, listeners attend to changes in pitch, loudness, rhythm, and voice quality (*emotional prosody*) to form an impression about the speaker's emotion state in conjunction with linguistic decoding (Wilson and Wharton 2006). In one important study, Scherer et al. (2001) presented 30 emotionally-inflected but semantically-anomalous "pseudo-utterances" produced by four German actors to native speakers of nine different languages (eight European and Malay/Bahasa Indonesian, $n = 32$–70 participants per language). The authors found that all listener groups recognized *fear*, *joy*, *sadness*, *anger* and "neutral" utterances strictly from prosody at above chance accuracy levels (66% accuracy overall in a five-choice task). They attributed this capacity to universal principles or "inference rules" that listeners apply during the processing of emotions in a foreign language. The results further emphasized that German (i.e., native) listeners performed significantly better on this task than the other language groups, and interestingly, that language similarity appeared to influence vocal emotion recognition; listeners whose native language was more linguistically similar to German (e.g., Dutch) tended to be more accurate than those from a highly dissimilar language (e.g., Malay). These data imply that language and culture[1] also play an important role in how vocal emotions are recognized (see Elfenbein and Ambady 2002 for an overview).

In fact, studies of both the face and voice commonly report differences in the overall *degree* of cross-cultural agreement witnessed across study groups and/or for specific emotion types (Beaupré and Hess 2005; Ekman et al. 1987; Matsumoto 1993; Thompson and Balkwill 2006). Typically, these data reveal an "in-group advantage" for identifying emotions posed by members of the same rather than of a different culture (Elfenbein and Ambady 2002), underscoring that social factors play a critical role at stages of regulating and displaying emotions (i.e., nonverbal display rules, Ekman et al. 1987; Mesquita and Frijda 1992). Although ample comparative data on the voice are still lacking, it is possible that socio-cultural influences on emotion recognition are especially pronounced in the vocal channel due to the unique interplay of emotion and language in speech; in this context, "natural" cues to emotion in the voice are tightly intertwined with the acoustic-phonetic properties of individual speech sounds, suprasegmental patterns for signalling stress and prominence, and other cues which mark meaningful linguistic contrasts of a language (e.g., Pell 2001). It is possible that for vocal expressions, acquired principles of linguistic communication and/or segmental properties of a language constrain how emotions are detected in the cross-cultural setting (Mesquita and Frijda 1992; Sauter and Scott 2007; Scherer et al. 2001; Pell and Skorup 2008).

[1] As in much of the broader literature, we do not attempt here to differentiate what may be specific cultural or properly linguistic influences on vocal emotion expressions in the construction of our stimuli in each language condition. The relationship between culture and language is obviously complex and cannot be explained by the current approach.

The fact that vocal expressions of emotion are inherently dynamic and integrated with linguistic properties of an utterance has led to certain practical adjustments in how these phenomena are studied in the cross-cultural setting. Contrary to research on the face, researchers interested in the voice cannot present a valid "snapshot" which represents the vocal attributes of an emotion; moreover, research on vocally-conveyed emotions must control not only for possible cultural preferences in how emotions are naturally conveyed in this channel (i.e., display rules), but also how these cues might interact with language content for each group under study. Given the additional challenge of constructing stimuli which "isolate" the effects of emotional prosody in each language of interest (Pell 2006), it is perhaps unsurprising that cross-cultural research on the voice is not as rich as that on the face. These methodological considerations likely explain why many researchers in the vocal literature have adopted "unbalanced" experimental designs which involve repetition of items produced by a single cultural group to a number of different listener groups (Scherer et al. 2001), or presentation of vocal expressions produced by speakers of several different languages to listeners from a single language/culture (Thompson and Balkwill 2006). While these research designs may not allow for the precise separation of effects due to the culture of the speaker versus the listener, they are highly valuable in elucidating how different languages and linguistic ability influence vocal emotion recognition in a manner which ensures rigorous control of both emotional and linguistic characteristics of the stimuli.

In a recent undertaking which adopted such an approach, Thompson and Balkwill (2006) studied 25 English-speaking listeners who identified the emotion conveyed by "semantically-neutral" sentences produced by speakers of English, German, Chinese, Japanese and Tagalog. In each of the five language conditions, listeners were presented four tokens representing each of four basic emotions (*joy*, *sadness*, *anger*, *fear*). The results of this study demonstrated above-chance identification of all emotions in all languages, although English listeners performed significantly better in their native language (in-group advantage). The notion that vocal emotion recognition is influenced by linguistic similarity (Scherer et al. 2001) was not strongly indicated by Thompson and Balkwill's (2006) data as English listeners demonstrated comparable accuracy to identify emotions in a related language such as German (67.5% correct) and in an unrelated language such as Tagalog (72.2% correct). However, potential shortcomings of this study were that very few tokens were presented to listeners in each language condition ($n = 16$) and the vocal exemplars were not extensively piloted to establish their emotional validity to a group of native listeners of each language prior to cross-cultural presentation (stimuli were included based on the impressions of two native speakers of each language). Thus, while the data imply that vocal emotion recognition is governed by both language-independent (universal) and language-specific processes, these patterns merit independent verification to evaluate the effects of different languages on vocal emotion recognition in a new context and using a well-defined stimulus set.

To achieve these objectives, here we tested a group of monolingual speakers of Argentine Spanish to evaluate how well they recognize vocal emotions when expressed in a culturally-appropriate manner in their native language, Spanish, and in three foreign languages which vary in their similarity to Spanish (English, German, and Arabic). In contrast to previous studies which omitted "disgust" (Scherer et al. 2001; Thompson and Balkwill 2006), our stimuli included vocal expressions of Ekman's (1992) five basic emotions in each of the four language conditions; these expressions were always encoded in emotionally-inflected pseudo-utterances which contained no linguistic cues for identifying emotions. Based on the literature reviewed, we hypothesized that our participants would be able to accurately decode expressions of basic emotions at above chance levels in all four language conditions, providing additional evidence that vocal emotions can be

understood through 'universal principles'. We further predicted an overall in-group advantage for recognizing emotions in our participants' native language, Spanish, with the possibility that recognition scores might be lowest for the language that was most distantly related to Spanish (i.e., Arabic).

Method

Participants

Participants were 61 monolingual, Spanish-speaking adults (32 female, 29 male) attending university in San Juan or Cordoba, Argentina (mean age = 27.0 years, $SD = 4.0$). Selected participants had little or no direct experience with speakers of English, German, or Arabic as established by a language questionnaire prior to testing. This procedure also affirmed that each participant could converse only in Spanish and documented which individuals had been exposed to other languages and in what context. Over half of participants ($n = 35$) claimed some exposure to English at least "once in a while", primarily through movies and written texts. Additional participants claimed some knowledge of Italian ($n = 4$), French ($n = 4$), Portuguese ($n = 1$), and Hebrew ($n = 1$) but were not fluent in these languages.

Materials

The stimuli were recordings of emotional "pseudo-utterances" produced by native speakers of four different languages: (Argentine) Spanish, (Canadian) English, (Standard high) German, and (Jordanian/Syrian) Arabic. All materials for English, German, and Arabic were taken from a perceptually-validated stimulus inventory (Pell et al. 2005, 2009). Stimuli for Spanish were constructed in a comparable manner for the purpose of this investigation. The Spanish, English, and Arabic stimuli were prepared in Montréal, Canada and the German stimuli were prepared in Leipzig, Germany.

Emotion Elicitation Procedure

For each language condition, a similar but independent set of procedures was carried out to elicit and perceptually validate utterances which expressed vocal emotions in each language (see Pell et al. 2009 for complete details). A native speaker of each language first constructed a list of pseudo-utterances which retained natural phonological and morpho-syntactic properties of the target language while replacing meaningful content words (e.g., nouns, verbs) with plausible pseudo-words (e.g., for English: *The fector egzullin the boshent*). In the elicitation study, this list of items was then produced by two men and two women who were native speakers of the language to convey each of the five basic emotions (anger, disgust, fear, sadness, joy), "pleasant surprise", and "neutral" affect by modulating features of their voice. Speakers of each language were recruited based on having amateur experience in acting or public speaking and each speaker was tested entirely in their native language. To facilitate production of pseudo-utterances which were vocally inflected in a relatively natural manner by the speaker, a list of "lexicalized" utterances with an obvious verbal-semantic context for expressing each emotion (for fear: *The convict is holding a knife!*) was also constructed for each language. During the recording session, the speaker first produced the list of lexicalized utterances to express a given target emotion, followed

by the list of pseudo-utterances to express the same emotion using only their voice. Although only pseudo-utterances were employed in the present study, our elicitation procedure helped to ensure that speakers expressed emotions in pseudo-utterances in a way that was natural and not exaggerated. The order for expressing particular emotions was blocked during the elicitation study and this order varied across speakers. All recordings were captured onto digital media using a high-quality fixed microphone. The recorded sentences ranged between one and 3 s in duration (approximately 8–14 syllables) across languages when spoken naturally to express the different target emotions.

Emotion Validation Procedure

For each language separately, digital recordings of all emotional utterances were transferred to a computer, edited to isolate the onset and offset of each sentence, and then entered into a perceptual validation study to determine whether the intended emotion of the speaker was successfully encoded according to individuals from the *same* linguistic background. Each stimulus validation study recruited a separate group of young native listeners (Spanish: $n = 21$, mean age = 25.2 years; English: $n = 24$, mean age = 24.9 - years; German: $n = 24$, mean age = 24.2 years; Arabic: $n = 19$, mean age = 23.9 years). Each native listener judged the emotional meaning of all the pseudo-utterances produced in the same language in a forced-choice response format involving seven response alternatives (anger, disgust, fear, sadness, happiness, pleasant surprise, neutral). Seven emotional response categories were included in our stimulus validation studies because the goal of this work was to construct a stimulus inventory that could be used in a broad range of investigations about emotion and affective communication in healthy and brain-damaged individuals (e.g., Dara et al. 2008; Pell and Skorup 2008). Stimuli representing the seven emotions were fully randomized for presentation to native listeners in each language condition. The percentage of native listeners who accurately categorized the target emotion of each pseudo-utterance was then computed for each item, speaker, and language, where chance accuracy performance was always 14.3% in the validation study. These perceptual data were consulted to select representative exemplars of each emotion for each language condition as specified below.

Stimulus Selection Procedure

For the purpose of this study, we did not use all of the emotional categories available in our stimulus inventory, but rather, we restricted our purview to five emotions—anger, disgust, fear, sadness, and joy. According to Ekman (1992), these emotions have yielded high agreement in their recognition from facial expressions and should be considered 'basic', whereas the status of other expressions as basic emotions is less certain (it is for this reason that we omitted the category of surprise). Only those items which were strongly indicative of the five basic emotions as dictated by the native listener group in our validation study were considered here. Since not all of our speakers in the elicitation/validation study were equally adept at conveying all target emotions to native listeners in a reliable manner (Pell et al. 2009), we selected stimuli produced by one of the male and one of the female speakers recorded for each language condition who had obtained the highest and most consistent emotion target ratings across emotions. A minimal criterion of 57% correct emotional target recognition, or approximately 4× chance accuracy in the validation study, was adopted. The experiment also included non-emotional or "neutral" utterances produced by the same speakers which acted as filler items.

In total, 40 emotional utterances (5 emotions × 8 exemplars) and eight neutral utterances which obtained the highest consensus about the meaning expressed when judged by the respective native listener group were selected per language, with the provision that an equal number of sentences produced by the male and female speaker of each language were always selected per emotion. Only five of the 160 emotional items (3 Spanish disgust, 2 Arabic angry) did not reach our selection criterion in order to balance the number of tokens produced by male and female speakers in each emotion condition. For the selected items in each language, recognition of the five emotional expressions by native listeners was high overall: Spanish = 80%; English = 84%; German = 87%; Arabic = 71%. However, native recognition varied predictably for certain emotions and somewhat across languages (e.g., recognition of disgust was relatively poor for Spanish and Arabic). Table 1 provides a summary of the perceptual characteristics of the selected stimuli as a function of language and emotion.

Experimental Tasks/Procedure

The stimuli chosen for each language condition were separately entered into four identical emotion recognition tasks to test how Spanish listeners identify vocal emotions as a factor of language. All 48 items (40 emotional + 8 neutral utterances) within each language task were randomized within two presentation blocks of 24 trials each. Each language task was also preceded by a block of six practice items (which did not appear in the experimental blocks) to familiarize participants with the nature of the sentences in each case and individual characteristics of the speakers' voices. The 61 participants were tested individually in a quiet room during a single one-hour session. Auditory stimuli were played at a comfortable listening level over headphones by a portable computer and each item was presented only once to participants in each language condition.

Participants were instructed to listen closely to each sentence and then judge how the speaker feels based on their voice by choosing one of six verbal labels displayed (in Spanish) on the computer screen: anger (enojo), disgust (repugnancia), fear (miedo), sadness (tristeza), joy (alegría), and neutral (neutralidad). The computer recorded the accuracy of a push-button response following each trial. Participants were informed at the onset of testing that the sentences were not supposed to make sense and might sound "foreign" and that they should always make their decision by attending closely to characteristics of the speaker's voice. Participants were never given any details about the

Table 1 Perceptual characteristics of pseudo-utterances selected for cross-cultural presentation by language and emotion

Emotion	Language			
	Spanish	English	German	Arabic
Anger	93	85	96	68
Disgust	51	80	90	61
Fear	88	81	86	76
Sadness	74	94	85	86
Joy	95	82	78	66
Neutral	62	77	98	72

Note: The data refer to the percentage of native listeners in each language condition who correctly identified the target emotion in a stimulus validation study (in a 7-choice response paradigm)

country of origin of the speakers, or what specific languages they would hear, until after the experiment. The order of the four language conditions was systematically varied within the participant group so that approximately 15 participants performed each of the language tasks in each possible presentation order.

After the four language tasks were completed, each participant completed a short questionnaire to briefly probe their attitudes about the experiment. At this stage, they were informed what languages they had heard. They were asked: (1) Which of the language tasks did you find hardest for recognizing emotions overall?; (2) Which emotion did you find hardest to recognize overall (irrespective of language)? At the end of the questionnaire participants were compensated a small amount for their involvement.

Results

Mean recognition of the five emotional expressions and neutral stimuli by the 61 Spanish listeners (in % correct) is presented in Table 2 for each language condition, along with associated error patterns for each expression type. In descriptive terms, overall emotion recognition scores ranged from a high of 64% in Spanish to a low of 56% in German (English = 58%, Arabic = 59%). The data exemplify that each emotion was recognized accurately from vocal cues in all four language conditions, that is to say, the most frequent response assigned to each expression type by the Spanish listeners was always the target emotion. Error confusion patterns implied certain tendencies to confuse sadness or anger with "neutral" expressions, and to categorize fear as sadness, but these patterns were not uniform across language conditions. Accuracy was especially low in two specific conditions: to recognize disgust from German (28% correct) and to recognize joy from English (32%).

Impact of Language and Emotion on Vocal Emotion Recognition Scores

Since our hypotheses did not allow exact predictions about how emotion recognition would vary among the four languages, the hit rates (proportion correct) observed for each of the five basic emotions in each language condition were entered into a 4 × 5 repeated measures Analysis of Variance (ANOVA). The factors of language (Spanish, English, German, Arabic) and emotion (joy, anger, disgust, fear, sadness) served as repeated measures in this analysis and all significant effects yielded by the ANOVA were investigated using Tukey's HSD post hoc comparisons ($p < .01$), whenever relevant. The 4 × 5 ANOVA yielded a main effect of language on vocal emotion recognition scores, $F(3, 180) = 7.49$, $p < .0001$. Post hoc elaboration of the language main effect indicated that Argentine participants performed significantly better when judging vocal emotions in their native language, Spanish, than in each of the three foreign languages. Overall, there were no significant differences in emotion recognition when the accuracy of the Spanish participants was compared for utterances spoken in English, German, and Arabic (see Fig. 1).

In addition, the ANOVA yielded a significant main effect for emotion, $F(4, 240) = 63.59$, $p < .0001$, and a significant interaction of Language × Emotion, $F(12, 720) = 39.65$, $p < .0001$. Post hoc Tukey's comparisons performed on the interaction ($p < .01$) focused on how the recognition of each emotion varied as a function of language. These comparisons demonstrated that vocal attributes of *joy* were identified significantly more accurately in Spanish (89%) than in Arabic (59%) and German (57%), which in turn exceeded hit rates for joy in English (32%). Expressions of *anger* were identified more accurately in both Spanish (81%) and German (77%) when compared to

Table 2 Recognition of basic emotions by 61 Spanish listeners (% target recognition) when listening to pseudo-utterances produced by speakers of Spanish, English, German, and Arabic

Expression type	Response					
	Anger	Disgust	Fear	Sadness	Joy	Neutral
(a) Spanish						
Anger	**81**	4	1	1	6	7
Disgust	11	**43**	4	4	23	15
Fear	4	2	**57**	19	3	15
Sadness	2	5	6	**51**	1	35
Joy	3	1	2	1	**89**	4
Neutral	1	4	5	34	1	**55**
(b) English						
Anger	**67**	8	1	0	5	19
Disgust	7	**52**	12	20	2	7
Fear	4	2	**61**	21	8	4
Sadness	1	3	12	**74**	1	9
Joy	9	9	7	18	**32**	25
Neutral	4	4	3	8	4	**77**
(c) German						
Anger	**77**	9	3	0	8	3
Disgust	16	**28**	15	24	10	7
Fear	0	3	**51**	36	1	9
Sadness	0	2	19	**65**	4	10
Joy	31	3	2	1	**57**	6
Neutral	1	3	2	4	1	**89**
(d) Arabic						
Anger	**66**	9	2	2	3	18
Disgust	8	**45**	5	10	16	16
Fear	9	5	**53**	5	16	12
Sadness	1	1	5	**77**	1	15
Joy	4	3	5	5	**59**	24
Neutral	6	5	2	21	5	**61**

Note: Bold values refer to correct recognition of the intended expression type

English and Arabic (67% and 66%, respectively). Expressions of *disgust,* which were recognized relatively poorly when compared to other emotions, demonstrated significantly higher recognition rates in English (52%), Arabic (45%) and Spanish (43%) when compared to German (28%). *Sadness* was recognized with the *least* accuracy in Spanish (51%) which differed significantly from Arabic (77%), English (74%), and German (65%). Interestingly, *fear* showed no significant differences in recognition accuracy across the four languages (English = 61%; Spanish = 57%; Arabic = 53%; German = 51%).[2]

[2] To explore whether the sex of participants influenced our results, the 4 × 5 ANOVA was re-run with the additional between-groups factor of sex (male, female). There was no evidence that participant sex influenced the reported patterns of emotion recognition in the form of a significant main or interactive effect involving Language or Emotion (all F's < 1.32, p's $> .26$).

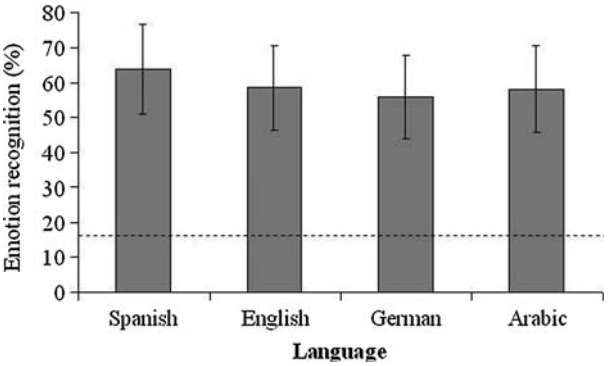

Fig. 1 Effects of language on the recognition of vocal emotions by 61 monolingual speakers of Argentine Spanish (in percent correct ± SD). Participants judged the emotion expressed by speakers from pseudo-utterances without meaningful language content; chance performance was approximately 17%

As described earlier, neutral stimuli were included as filler items in the experiment and were not entered in the main analysis to focus our discussion on effects due to basic emotions (Ekman 1992). However, to exemplify how language influenced recognition of this stimulus category, a one-way ANOVA was performed on the neutral data which yielded a significant effect for language, $F(3, 180) = 54.07$, $p < .0001$. Post hoc (Tukey's HSD) tests indicated that neutral utterances were recognized significantly better in German than in each of the other languages, and accuracy in English also exceeded that for Spanish.

Finally, it has been argued that raw hit rates are not always the best measure of accuracy in experiments which use forced-choice tasks because these data may not adequately control for bias in both stimulus frequency and *response frequency* (Wagner 1993). Although we controlled carefully for stimulus frequency in each of our language/emotion conditions, it is possible that our data reflect certain bias in how the Spanish participants used particular emotional response categories, which may also have differed across languages. To investigate this possibility briefly, unbiased hit rates ("H_u scores") were computed for the classification data for the Spanish group as a whole according to Wagner. These measures denote the unbiased proportion of correct responses observed in each language and emotion condition, where a score of zero reflects chance performance and a score of one reflects perfect performance. As demonstrated in Table 3, the overall accuracy of Spanish listeners was 0.37 for English, 0.38 for Arabic and German, and 0.42 for Spanish; this pattern strongly resembles the one described above for our main analysis as illustrated in Fig. 1. When viewed across languages, the recognition of basic vocal emotion expressions ranged from a low of 0.28 for disgust to a high of 0.52 for anger, a pattern which is also consistent with data presented in Table 2.

Subjective Influences of Language and Emotion on Vocal Emotion Recognition

To document some of the qualitative impressions of the Spanish participants about our independent variables, responses to the post-session questionnaire were analyzed to clarify: (1) which language condition was perceived as most difficult for recognizing emotions; and (2) which emotion was most difficult to recognize, irrespective of language. The vast majority (56/61 or 92%) of participants reported that the Arabic task was most difficult, whereas the remainder (5/61) found the Spanish task most difficult. A number of participants (16/61) claimed that in the Spanish condition, they had attempted to figure out what

Table 3 Unbiased hit rates (H_u scores) for the 61 Spanish listeners for each emotion by language condition (0 = chance, 1 = perfect performance)

Emotion	Language				Total
	Spanish	English	German	Arabic	
Anger	0.64	0.49	0.48	0.47	0.52
Disgust	0.31	0.35	0.16	0.30	0.28
Fear	0.42	0.39	0.28	0.39	0.37
Sadness	0.24	0.39	0.33	0.50	0.37
Joy	0.65	0.20	0.40	0.35	0.40
Neutral	0.23	0.42	0.64	0.26	0.39
Total	0.42	0.37	0.38	0.38	

the speaker was saying as they listened to pseudo-utterances. Irrespective of language, disgust was chosen as the emotion that was hardest to identify (50/61 or 82% of participants), followed by joy (10%) and fear (8%).

Discussion

Our results indicate that when adults listen to a foreign language they can successfully infer the speaker's emotional state strictly from their vocal inflections while speaking, consistent with previous findings (Albas et al. 1976; Scherer et al. 2001; Thompson and Balkwill 2006; Van Bezooijen et al. 1983). Specifically, we found that monolingual speakers of Argentine Spanish could categorize five basic emotions (Ekman 1992) from vocal attributes of their native language, Spanish, and three foreign languages (English, German, Arabic). Accuracy levels ranged between three and four times what would be expected by chance in our task (see Scherer et al. 2001 for similar findings). The ability to accurately categorize the emotions was necessarily tied to processes for decoding vocal, rather than linguistic, features of speech due to the presentation of pseudo-utterances which resembled each language but furnished no meaningful emotion-related linguistic cues. In overall terms, these findings support the hypothesis that vocal expressions of the emotions investigated, like corresponding facial expressions (Ekman and Friesen 1971), contain invariant or "modal" elements which are universally exploited by speakers and can be decoded across languages irrespective of the linguistic ability and experience of the listener (Scherer et al. 2001; Thompson and Balkwill 2006).

Our data simultaneously highlight an in-group advantage for recognizing vocally-expressed emotions presented in a listener's native language (Albas et al. 1976; Scherer et al. 2001; Thompson and Balkwill 2006; Van Bezooijen et al. 1983). While the impact of language varied for specific emotions (see below), there was a small but robust benefit of processing vocal cues to emotion when the participants listened to native Argentine speakers than to speakers of three foreign languages. One explanation for this finding is that the in-group advantage for recognizing emotions in the native language reflects learned, cultural differences in how emotions are expressed and understood through nonverbal behavior (Elfenbein et al. 2007). Another, potentially complementary explanation for our data is that the in-group advantage for recognizing vocal emotions is due to *interference* of language-specific features when listening to a foreign language (Mesquita and Frijda 1992; Scherer et al. 2001; Pell and Skorup 2008). In the latter case,

linguistically-assigned differences in the segmental inventory, intonational features, or in the accent or rhythmic structure of a foreign language could interact negatively with basic aspects of auditory speech processing and/or with processes for extracting salient emotional features from prosody. Pending further data, this account places the source of the in-group advantage at the stage of phonological and suprasegmental encoding which is likely to be more effortful when listening to a foreign language, with consequences on vocal emotion recognition (Beier and Zautra 1972; Pell and Skorup 2008).

Contrary to expectation, our comparisons supply no evidence that language *similarity* was an overall predictor of how Spanish participants recognized vocal emotions in the cross-cultural setting. Scherer and colleagues (2001), who presented a single set of pseudo-utterances produced by German actors to speakers of eight different European and one non-European language, observed that vocal emotion recognition was markedly poorer in the one language that was most dissimilar from German, Malay. Our investigation, which presented emotional stimuli in four distinct languages to a single cultural group, Argentine Spanish, included two foreign languages from a different branch of the same family (Indo-European: English, German) and one language from an independent family (Semitic: Arabic). Overall, there were no discernable differences in how Spanish listeners identified vocal emotions from English, German, and Arabic as indicated by the raw as well as unbiased hit rates, despite the fact that Arabic is most linguistically dissimilar from Spanish. Interestingly, these findings contrast with those of our questionnaire which showed that the vast majority of our participants (92%) *perceived* the Arabic task as most difficult for categorizing vocal emotions, although this was not reflected in their actual performance.

This outcome suggests that linguistic similarity is not a consistent factor which predicts the accuracy of vocal emotion recognition across languages. This conclusion fits with data reported by Thompson and Balkwill (2006), which show that English listeners make a comparable number of errors when identifying vocal emotions from German, Mandarin, Japanese, and Tagalog, although they perform significantly better for English. More research involving an even more diverse array of languages is needed to fully understand the effects of language distance or linguistic similarity on vocal emotion recognition. It is possible that, if language similarity were defined according to specific intonational or timing properties of a language, rather than by language typology or families as is common in the literature, future investigations would be better positioned to evaluate the importance of this variable and its relationship to the in-group processing advantage during vocal emotion processing.

Another contribution of this report is to qualify that the in-group advantage for recognizing emotions in the voice is a relatively general performance feature which does not apply evenly to all emotion types (Elfenbein and Ambady 2002). Independent of language, we found that vocal emotion recognition tended to be highest for anger (73%) and sadness (66%) and lowest for disgust (42%), consistent with the background literature (Banse and Scherer 1996; Pell et al. 2009; Scherer et al. 2001; Thompson and Balkwill 2006). When individual languages were inspected, there was little evidence in our data that listening to Spanish promoted systematically better recognition of specific emotional expressions in the voice when compared to the three foreign languages. Rather, the impact of language on emotion recognition rates seemed to vary in a unique manner for each emotion type, as well as for neutral utterances (this was also true for three of the five languages studied by Thompson and Balkwill 2006). Only one emotion, joy, demonstrated a clear processing advantage when it was produced by native Argentine speakers than speakers of all other languages; this emotion was further distinct in its range of recognition across languages,

varying from 89% in Spanish to 32% in English (or 0.65–0.20 in H_u scores, respectively). The pattern for joy is potentially meaningful because research implies that expressions of this emotion in the vocal channel are highly amenable to language differences and possible cultural stereotypes (Johnson et al. 1986; Juslin and Laukka 2003), as may be inferred from our data. This contrasts with *facial* expressions of joy which are associated with the smile and typically recognized with minimal difficulty or variation across cultures (Elfenbein and Ambady 2002).

However, there are also well known differences in how well individuals, including actors, portray specific emotions through nonverbal channels (Wallbott and Scherer 1986). As a result, due to the small number of speakers who portrayed emotions in each of our language conditions (one male, one female), the emotion-specific patterns we observed should be interpreted cautiously until further data are collected. Although we made strong attempts to choose stimuli which were perceptually valid to native listeners of the same language, some of the emotion-specific patterns observed could reflect individual differences which naturally occur at the stage of *encoding* discrete emotions (Banse and Scherer 1996; Scherer et al. 1991). That is, it remains possible that our Spanish listeners recognized certain emotions more reliably in specific language conditions because individual speakers (of whatever background) were relatively successful at encoding the prototypical acoustic-perceptual features of these emotions which could be accurately decoded by Spanish listeners via universal principles. We are investigating this possibility in an ongoing study which compares acoustic attributes of stimuli produced in each language in relation to the current perceptual findings.

In closing, our study builds on a growing literature which emphasizes the role of both universal principles and socio-cultural factors for recognizing emotion in the voice. Like facial expressions (Ekman et al. 1969, 1987), vocal expressions appear to contain discrete, pan-cultural elements as one of the physiological outcomes of experiencing a basic emotion which yield predictable effects on the human vocal apparatus and acoustic-perceptual properties of the voice (Scherer 1986). These modal tendencies are likely codified as natural signals which can be recognized in speech (Wilson and Wharton 2006) even when listeners are completely unfamiliar with corresponding linguistic features in the signal. However, language-specific properties of a foreign language hold the potential of reducing the *efficiency* of this processing in many contexts (Pell and Skorup 2008).

Although the scope of our current study was necessarily limited to Spanish listeners, extending our methods to include comparable groups of English, German, and Arabic listeners who have no knowledge of the other languages of interest would allow even further insights about the role of culture and linguistic ability on vocal emotion recognition. Also, note that our study focused on the recognition of relatively *unambiguous* portrayals of vocal emotion in each language; we did not include expressions which contain emotional blends which are commonly encountered in natural discourse, nor those which reflect instances where a speaker attempts to disguise or modify the intensity of their emotional expressions for social purposes (Hess et al. 2005). Since the emotional expressions in this study were posed, our data also do not reflect potential cultural differences due to learned social conventions which govern the antecedent events for experiencing emotions (Scherer 1997) which can play a role in how emotions are displayed. When these additional variables are considered fully in the context of our findings, one can speculate that socio-cultural dimensions of an interaction exert an even greater impact on how vocal emotions are recognized when encountered spontaneously in the cross-cultural setting (Kitayama and Ishii 2002).

Acknowledgements This work was supported by a Discovery grant from the Natural Sciences and Engineering Research Council of Canada (to M.D. Pell) and by the German Research Foundation (DFG FOR 499 to S.A. Kotz).

References

Albas, D., McCluskey, K., & Albas, C. (1976). Perception of the emotional content of speech: A comparison of two Canadian groups. *Journal of Cross-Cultural Psychology, 7*(4), 481–489.

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology, 70*(3), 614–636.

Beaupré, M., & Hess, U. (2005). Cross-cultural emotion recognition among Canadian ethnic groups. *Journal of Cross-Cultural Psychology, 36*(3), 355–370.

Beier, E., & Zautra, A. (1972). Identification of vocal communication of emotions across cultures. *Journal of Consulting and Clinical Psychology, 39*(1), 166.

Dara, C., Monetta, L., & Pell, M. D. (2008). Vocal emotion processing in Parkinson's disease: Reduced sensitivity to negative emotions. *Brain Research, 1188*, 100–111.

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion, 6*, 169–200.

Ekman, P., & Friesen, W. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology, 17*(2), 124–129.

Ekman, P., Friesen, W., O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., et al. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of Personality and Social Psychology, 53*(4), 712–717.

Ekman, P., Sorenson, E. R., & Friesen, W. V. (1969). Pan-cultural elements in facial displays of emotion. *Science, 164*, 86–88.

Elfenbein, H., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin, 128*(2), 203–235.

Elfenbein, H., Beaupré, M., Lévesque, M., & Hess, U. (2007). Toward a dialect theory: Cultural differences in the expression and recognition of posed facial expressions. *Emotion, 7*, 131–146.

Hess, U., Adams, R., & Kleck, R. (2005). Who may frown and who should smile? Dominance, affliction, and the display of happiness and anger. *Cognition and Emotion, 19*(4), 515–536.

Izard, C. E. (1977). *Human emotions*. New York: Plenum Press.

Izard, C. E. (1994). Innate and universal facial expressions: Evidence from developmental and cross-cultural research. *Psychological Bulletin, 115*(2), 288–299.

Johnson, W. F., Emde, R. N., Scherer, K. R., & Klinnert, M. D. (1986). Recognition of emotion from vocal cues. *Archives of General Psychiatry, 43*, 280–283.

Juslin, P., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin, 129*(5), 770–814.

Kitayama, S., & Ishii, K. (2002). Word and voice: Spontaneous attention to emotional utterances in two languages. *Cognition and Emotion, 16*(1), 29–59.

Matsumoto, D. (1993). Ethnic differences in affect intensity, emotion judgements, display rule attitudes, and self-reported emotional expression in an American sample. *Motivation and Emotion, 17*, 107–123.

Mesquita, B., & Frijda, N. (1992). Cultural variations in emotions: A review. *Psychological Bulletin, 112*(2), 179–204.

Pell, M. D. (2001). Influence of emotion and focus location on prosody in matched statements and questions. *Journal of the Acoustical Society of America, 109*, 1668–1680.

Pell, M. D. (2006). Judging emotion and attitudes from prosody following brain damage. In S. Anders, G. Ende, M. Junghofer, J. Kissler, & D. Wildgruber (Eds.), *Progress in Brain Research*, *156*, 303–317.

Pell, M. D., Kotz, S. A., Paulmann, S., & Alasseri, A. (2005). Recognition of basic emotions from speech prosody as a function of language and sex. *Abstracts of the Psychonomic Society 46th Annual Meeting* (Vol. 10, pp. 97–98).

Pell, M. D., Paulmann, S., Dara, C., Alasseri, A., & Kotz, S. A. (2009). *Factors in the recognition of vocally expressed emotions: A comparison of four languages*. Manuscript under review.

Pell, M. D., & Skorup, V. (2008). Implicit processing of emotional prosody in a foreign versus native language. *Speech Communication, 50*, 519–530.

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. A., & Scott, S. (2007). More than one kind of happiness: Can we recognize vocal expressions of different positive states? *Motivation and Emotion, 31*, 192–199.
Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin, 99*(2), 143–165.
Scherer, K. R. (1997). The role of culture in emotion-antecedent appraisal. *Journal of Personality and Social Psychology, 73*(5), 902–922.
Scherer, K. R., Banse, R., & Wallbott, H. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology, 32*, 76–92.
Scherer, K. R., Banse, R., Wallbott, H. G., & Goldbeck, T. (1991). Vocal cues in emotion encoding and decoding. *Motivation and Emotion, 15*(2), 123–148.
Thompson, W., & Balkwill, L.-L. (2006). Decoding speech prosody in five languages. *Semiotica, 158*(1/4), 407–424.
Van Bezooijen, R., Otto, S., & Heenan, T. (1983). Recognition of vocal expressions of emotion: A three-nation study to identify universal characteristics. *Journal of Cross-Cultural Psychology, 14*(4), 387–406.
Wagner, H. L. (1993). On measuring performance in category judgment studies of nonverbal behavior. *Journal of Nonverbal Behavior, 17*, 3–28.
Wallbott, H. G., & Scherer, K. R. (1986). Cues and channels in emotion recognition. *Journal of Personality and Social Psychology, 51*(4), 690–699.
Wilson, D., & Wharton, T. (2006). Relevance and prosody. *Journal of Pragmatics, 38*, 1559–1579.