Available online at www.sciencedirect.com

ScienceDirect

Speech Communication 50 (2008) 519–530

SPEECH COMMUNICATION

www.elsevier.com/locate/specom

Implicit processing of emotional prosody in a foreign versus native language

Marc D. Pell [*], Vera Skorup

McGill University, School of Communication Sciences and Disorders, 1266 Avenue des Pins ouest, Montréal, Québec, Canada H3G 1A8

Received 1 February 2008; received in revised form 13 March 2008; accepted 17 March 2008

Abstract

To test ideas about the universality and time course of vocal emotion processing, 50 English listeners performed an emotional priming task to determine whether they implicitly recognize emotional meanings of prosody when exposed to a foreign language. Arabic pseudo-utterances produced in a *happy, sad*, or *neutral* prosody acted as primes for a *happy, sad,* or '*false'* (i.e., non-emotional) face target and participants judged whether the facial expression represents an emotion. The prosody-face relationship (congruent, incongruent) and the prosody duration (600 or 1000 ms) were independently manipulated in the same experiment. Results indicated that English listeners automatically detect the emotional significance of prosody when expressed in a foreign language, although activation of emotional meanings in a foreign language may require increased exposure to prosodic information than when listening to the native language.

Keywords: Speech processing; Vocal expression; Cross-linguistic; Cultural factors; Semantic priming

1. Introduction

While there is an abundance of studies of how facial expressions of emotion are communicated and recognized in a cross-cultural setting (Ekman and Friesen, 1971; Ekman et al., 1987; Ekman et al., 1969; Izard, 1977), surprisingly little data have been gathered on how *vocal* expressions of emotion are recognized by individuals from different linguistic and cultural backgrounds. In the vocal channel, affect "bursts" (e.g., laughter) or emblems (e.g., "yuck") frequently refer to discrete emotion states and have been studied to some extent (see Schroder, 2003). Of greater interest here, vocal indicators of emotion are frequently embedded in spoken language and thus detectable in the suprasegmental content of speech, or *speech prosody*. The goal of the present investigation was to supply initial data on how emotions expressed through prosody are implicitly processed in a *foreign* language when compared to a listener's native language.

Dynamic changes in timing, pitch, loudness, and voice quality while speaking serve as the "observable" prosodic elements which combine to signal emotions in the vocal channel (Bachorowski, 1999; Frick, 1985; Juslin and Laukka, 2003; Scherer, 1986). For example, expressions of discrete emotions such as joy, sadness, and anger tend to correspond with distinct modulation patterns involving multiple prosodic elements over the course of an utterance (Banse and Scherer, 1996; Pell, 2001; Williams and Stevens, 1972). These patterns may be perceived categorically in speech to understand the speaker's emotion state (Laukka, 2005). There is ample evidence that vocal expressions of basic emotions can be perceptually identified at above-chance levels by native listeners of the same language (Pell, 2002; Pell et al., 2005; Scherer et al., 1991) and that these emotions are processed implicitly by listeners when exposed to their native language (Pell, 2005a,b).

Research on how listeners recognize vocal expressions of emotion produced in a foreign language is represented by

[*] Corresponding author. Tel.: +514 398 4133; fax: +514 398 8123.
E-mail address: marc.pell@mcgill.ca (M.D. Pell).
URL: http://www.mcgill.ca/pell_lab (M.D. Pell).

0167-6393/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.specom.2008.03.006

only a handful of studies to date (Albas et al., 1976; Beier and Zautra, 1972; Breitenstein et al., 2001; Kramer, 1964; McCluskey et al., 1975; Scherer et al., 2001; Thompson and Balkwill, 2006; Van Bezooijen et al., 1983). In all cases, these investigations required listeners to consciously identify or name the emotion conveyed by emotional prosody in an "off-line" manner. Beier and Zautra (1972) compared how American, Japanese, and Polish listeners recognize six vocal expressions of emotion (*happiness, fear, sadness, anger*, *indifference*, and *flirt*) from English utterances varying in length (*hello*, *good morning, how are you*, and a full sentence). Results of that study indicated that the American (i.e., native) listeners displayed an "in-group advantage" over the other two groups for recognizing vocal emotions when the stimuli were relatively short; however, as the duration of the sentences increased, cross-cultural recognition rates increased to the point that there were no significant differences among the three language groups when listening to full sentences. These data imply that the recognition of vocal emotion expressions in speech, unlike conjoined linguistic cues, is governed to some extent by 'universal' or culture-independent processes which are applied to these cues during speech processing; this argument fits with similar claims about why emotional facial cues can be recognized successfully across cultures (see Elfenbein and Ambady, 2002 for discussion). Interestingly, Beier and Zautra's (1972) findings further imply that the *amount of exposure* to vocal-prosodic cues in a foreign language is a significant factor in how these expressions are processed and understood, although no additional data are available to assess this particular claim.

Consistent with the idea that culture and language experience[1] play a role in vocal emotion recognition, several other researchers have observed an "in-group advantage" for identifying vocal emotions expressed by speakers of the same language when compared to speakers of a foreign language (Albas et al., 1976; Pell et al., submitted for publication-a; Scherer et al., 2001; Thompson and Balkwill, 2006; Van Bezooijen et al., 1983). Van Bezooijen et al. (1983) presented the sentence "two months pregnant", produced in Dutch to convey *disgust, surprise, shame, joy, fear, contempt, sadness, anger* or a *neutral* emotion, to separate groups of Dutch, Japanese, and Taiwanese listeners. Although Japanese and Taiwanese participants were able to recognize the majority of emotions with above-chance accuracy, the native Dutch listeners performed significantly better than the two foreign language groups overall. Similarly, Thompson and Balkwill (2006) noted that English-speaking listeners could identify emotions conveyed by "semantically-neutral" sentences produced in four different foreign languages (German, Chinese, Japanese, and Tagalog), although the listeners performed better in their native language (English). The same pattern of findings emerged from our recent investigation of 60 Spanish-speaking listeners who could accurately identify six emotions from pseudo-utterances spoken in Spanish, English, German, and Arabic, although the participants demonstrated a significant advantage for processing emotions in the native language, Spanish (Pell et al., submitted for publication-a). Collectively, these studies support the argument that both universal and language/culture-specific processes affect how emotions are processed in speech.

In the largest study to date, Scherer et al. (2001) recruited listeners from nine countries in Europe, Asia, and North America who were presented pseudo-sentences posed by German actors to express *fear, joy, sadness, anger,* or a *neutral* emotion. Again, comparison of the emotional judgements rendered by each language group highlighted that native German listeners performed most reliably overall, although the data also suggested that performance tended to be better for listeners whose native language was more linguistically related to German (e.g., Dutch, English). Nonetheless, all listener groups recognized the emotions at above chance performance levels and there was a high degree of similarity in emotion confusion patterns across language groups. Consistent with the literature reported, these findings were interpreted as evidence for universal "inference rules" which are applied by listeners to infer the emotional meaning of vocal expressions in a foreign language (Scherer et al., 2001). However, when exposed to languages which are increasingly dissimilar from the native language, the authors proposed that language-specific prosodic features which are not central to identifying emotions may progressively interfere in procedures for vocal emotion decoding (also Mesquita and Frijda, 1992).

Taken as a whole, these studies converge on the hypothesis that vocal expressions of emotion possess certain universally recognizable characteristics, similar to what has been argued for emotional face recognition (see Wilson and Wharton, 2006 for a recent view on how emotions are communicated through speech). By the same token, there is clear evidence that recognizing vocal cues to emotion in a foreign language is seldom as reliable as in-group judgements and language-related factors, such as stimulus duration or linguistic similarity, likely impact on decoding skills in the vocal channel. These issues merit closer examination given the general paucity of data on emotional speech processing and to address methodological limitations of the current literature which is poorly suited in many ways for understanding how vocal cues or other *dynamic* representations of emotion are processed in the context of language (Elfenbein and Ambady, 2002; Pell, 2005a).

1.1. Implicit processing of vocal emotions in speech

As mentioned, nearly all cross-cultural studies have required subjects to explicitly categorize vocal expressions

[1] In this experiment, as in much of the broader literature, we do not attempt to distinguish what may be specific cultural versus properly linguistic influences on vocal emotion expression and its processing. The relationship between culture and language is complex and cannot be explained by the current approach.

by matching the stimulus with an emotion label after the stimulus has been presented, using the forced-choice format. This approach has been the topic of some debate in the literature (Ekman, 1994; Izard, 1994; Russell, 1994) and from a speech processing viewpoint, this method does not afford a sensitive view of whether listeners exposed to a foreign language activate emotional meanings in real time ("on-line") as dynamic changes in speech prosody are encountered. One way to circumvent procedures which require listeners to name the meaning of emotional prosody is to implicitly gauge whether emotional features of prosody systematically 'prime' decisions about a related versus unrelated emotional face as indexed successfully by the Facial Affect Decision Task (Pell, 2005a,b), a nonverbal adaptation of the cross-modal Lexical Decision Task (Onifer and Swinney, 1981; Swinney, 1979).

A significant body of research has documented semantic priming effects, including emotion congruency effects, when processing various features of communicative stimuli (Innes-Ker and Niedenthal, 2002; Niedenthal et al., 1994; Rossell and Nobre, 2004). In keeping with this research, initial experiments using the Facial Affect Decision Task (FADT) have furnished evidence that the emotional features of prosody in speech prime subsequent decisions about a facial expression which shares the same emotion category membership as the prosodic stimulus, at least when these vocal expressions are produced by native speakers of the same language. In the FADT, participants render a YES/NO decision to indicate whether a static facial expression represents a 'true' emotion, where the face target either conforms to a prototypical display of emotion (e.g., happy, angry) or presents a facial configuration which does not represent a discrete emotion. In two published reports which presented different numbers and categories of emotions, listeners were passively exposed to emotionally-inflected pseudo-utterances (e.g., *Someone migged the pazing*) which preceded each face target and the emotional relationship of the two events was systematically manipulated across trials.

In both studies, results indicated that the accuracy and/or speed of facial affect decisions was systematically enhanced when the preceding vocal expressions were congruent rather than incongruent in emotion to the face; these systematic effects are hypothetically linked to emotion-based semantic knowledge shared by vocal and facial expressions of discrete emotions which yield priming in this processing environment (Pell, 2005a,b). Importantly, the results emphasize that listeners implicitly activate the emotional meaning of prosody when processing their native language, even when these features are not the focus of attention, and in the absence of any requirement to retrieve explicit emotional details about the prosody or the face stimulus (see related work by de Gelder and Vroomen (2000); Massaro and Egan (1996); Vroomen et al. (2001)).

An additional contribution of Pell, 2005b was to show that emotional priming effects due to prosody vary according to the *duration* of vocal emotion cues presented to the listener. Whereas vocal primes presented in the initial study were all relatively long in duration (i.e., greater than 1.5 s), *happy, sad,* and *neutral* pseudo-utterances presented by Pell (2005b) were cut into fragments of 300, 600, and 1000 ms from sentence onset to test whether activation of emotional meanings from prosody depends on a certain level of exposure to prosodic features of an utterance. In that study, prosodic primes were followed immediately by a related or unrelated face target which was presented at the offset of the prosody stimulus in each duration condition. Results indicated that emotion congruency effects were evident only in the 600 and 1000 ms prosody duration conditions (affecting decision latencies), with maximal priming of facial affect decisions when an emotionally-related prosody was presented for 600 ms. These patterns suggest that when listeners process their native language, the emotional meaning of prosody is implicitly activated according to a specific time course following initial exposure to relevant vocal cues.[2] The question of whether listeners implicitly process vocal cues to emotion in a completely unknown or foreign language has not been tested to date.

1.2. The current study

To advance knowledge of whether vocal cues to emotion are *implicitly* processed in a foreign language–and to test specific ideas about whether exposure duration to prosody has a critical influence on these processes (Beier and Zautra, 1972)—the methods of Pell, 2005b were closely replicated here but using a cross-linguistic variant of the FADT. In our previous experiments, English listeners were always presented "English-like" pseudo-utterances which had been produced to convey different emotions according to cultural conventions appropriate for (Canadian) speakers of English. In the present study, we modified our procedures in one critical way: English listeners were presented pseudo-utterances spoken in a foreign and linguistically distant language, Arabic, which were constructed in a similar manner and then produced to convey emotions according to conventions appropriate to speakers of (Palestinian) Arabic. The duration of prosodic stimuli was again manipulated to address the time course issue (Beier and Zautra, 1972; Pell, 2005b) by cutting Arabic primes to 600 or 1000 ms in duration.

The initial goal of the study was to evaluate whether English-speaking listeners are sensitive to emotions encoded by prosody in Arabic, as inferred by the presence of emotion congruency (priming) effects in the FADT, thus serving as a unique "case study" of cross-cultural processing of vocal emotions for an unknown language. A second goal was to compare our new data on processing emotions in a foreign language with highly comparable, published

[2] The data also suggested that priming effects related to "sad" and "happy" vocal expressions might each have a distinct time course, with "sad" expressions showing greater priming in the longest (1000 ms) stimulus duration condition (Pell, 2005b).

data on the processing of vocal emotions in the listeners' native language, English (Pell, 2005b). Based on the published literature, we predicted that English listeners would implicitly detect the emotional meanings of prosody encoded by Arabic speakers, yielding reliable priming effects in the FADT data, owing to 'universal' procedures or inference rules for deriving this information from a foreign language (Scherer et al., 2001). However, we also expected that language/cultural differences would influence how English listeners responded to emotional prosody in Arabic, for example, by delaying priming effects when processing emotions in a foreign language rather than the native language (Beier and Zautra, 1972).

2. Methods

2.1. Participants

Fifty native speakers of Canadian English (25 male, 25 female) averaging 23.6 (±8.9) years in age and 15.1 (±1.9) years in formal education took part in the study. Participants responded to an advertisement posted on the McGill campus and were paid a nominal fee after completing the study (CAD$10). Each participant reported normal hearing and normal or corrected-to-normal vision when questioned at study onset. A detailed language questionnaire administered prior to testing ensured that none of the participants was familiar with the Arabic language in any manner.

2.2. Materials

Emotionally-inflected pseudo-utterances ("primes") were matched with static faces representing a 'true' or a 'false' expression of emotion ("targets") over a series of trials (see details below; also Pell, 2005a,b). As noted earlier, the methods of Pell, 2005b were closely replicated with the main exception that sentences presented as the prosodic primes were constructed to resemble Arabic and were emotionally-inflected in a way that was natural for speakers of Arabic rather than English. A description of how both the English and Arabic pseudo-utterances were elicited and perceptual data pertaining to these inventories was gathered is provided in full elsewhere (Pell et al., 2005; Pell et al., submitted for publication-b).

2.2.1. Prosodic stimuli

Experimental primes were semantically anomalous pseudo-utterances which conveyed a *happy, sad*, or *neutral* emotion through prosody when produced by two male and two female speakers of Arabic. The prosodic materials were created in the following manner. Prior to recording, a list of 20 Arabic pseudo-utterances, each approximately 10 syllables in length, was constructed by replacing the content words of real Arabic utterances with sound strings that were phonologically licensed by Arabic but semantically meaningless to native listeners (e.g., أغْلاضْ الأخْوام صَبيرَة). Because pseudo-utterances retained appropriate phonological and some grammatical properties of Arabic (e.g., gender or other inflectional markers), they were highly "language-like" to native listeners and could be produced with relative ease by Arabic speakers to naturally express specific emotions through prosody.

Four Arabic speakers were recruited in Montréal (Canada) to produce the same list of pseudo-utterances in each of seven distinct emotions (anger, disgust, sadness, fear, happiness, pleasant surprise, neutral). All speakers were native to the Middle-East (Syrian/Jordanian dialect), were young adults studying at McGill University (mean age = 24.8 years), and had been in Montréal for less than three years in duration. Speakers were selected for having lay experience in some aspects of public speaking (e.g., radio, member of public speaking groups). All communication with the speakers and during the recording session was conducted entirely in Arabic. Each speaker was recorded separately in a sound-attentuated room; the order in which each speaker expressed each of the seven emotions was varied across encoders. During the elicitation study, speakers were encouraged to produce the list of pseudo-utterances to express each target emotion in a natural manner as if talking to the examiner, in a way that avoided exaggeration. All utterances were recorded onto digital audiotape using a high-quality head-mounted microphone and then digitally transferred to a computer and saved as individual sound files.

To perceptually validate the recordings elicited for each target emotion, an emotion recognition task which included all exemplars of the seven emotion categories (560 pseudo-utterances in total) was presented to 19 native Arabic listeners (Pell et al., 2005). For each utterance, the listener categorized which of the seven emotional categories was being expressed by the speaker's voice by selecting the corresponding emotion term from a printed list on a computer screen. As described by Pell et al. (submitted for publication-b), Arabic listeners showed relatively poor consensus at times about the emotion conveyed by pseudo-utterances, especially for pleasant surprise and disgust, highlighting that our speakers probably had difficulty simulating emotions for many of the items originally produced.[3] For the current study, we selected 36 highly robust items produced by the four speakers—12 distinct exemplars representing *happy, sad*, and *neutral* prosody—which had been identified at a minimum 70% consensus level where chance performance in our validation task was 14.3% (see Table 1). Utterances conveying a *happy* or *sad* prosody were of primary theoretical interest in the experiment and neutral utterances were used largely as filler items to control for the proportion of related trials in the exper-

[3] As reported by Pell et al. (submitted for publication-b), the following consensus rates were observed for the Arabic listener group when judging emotional pseudo-utterances (where chance performance = 14%): anger = 62.4%; disgust = 53.5%; fear = 59.6%; sadness = 72.5%; happiness = 56.3%; pleasant surprise = 46.4%; neutral = 60.4%.

Table 1
Characteristics of the prosody and face stimuli presented in the experiment

Event type	Parameter	Happiness	Sadness	Neutral
Prosodic primes	Perceived target recognition[a] (%)	75.0	88.2	78.9
	Fundamental frequency (*mean*[b])	1.39	−1.57	−0.90
	Fundamental freq. (*variation*[b])	0.79	0.31	0.66
	Speaking rate (syllables/second)	5.08	5.21	5.99
	Intensity variation ([b])	0.59	0.51	0.59
Face targets	Perceived target recognition[c] (%)	99.8	93.1	–
	Perceived target intensity[d] (0–5)	3.9	3.4	–
	Perceived valence[e] (−5 to 5)	+3.1	−2.8	–

[a] Based on responses of 19 Arabic listeners in a 7-choice forced identification task, where chance performance = 14.3%.

[b] Utterance measures were normalized in reference to each speaker and then averaged across speakers who produced exemplars of each emotion. For fundamental frequency mean, all utterances produced by a given speaker was transformed as follows: $F_{0\mathrm{norm}} = (F_{0\mathrm{i}} - F_{0\mathrm{Mean}})/\mathrm{sd}$, where $F_{0\mathrm{i}}$ is the observed utterance F_0 of the item in Hertz, $F_{0\mathrm{Mean}}$ is the speaker's typical mean F_0 averaged for a broad array of their utterances (Pell, 2002), and sd is the standard deviation. For fundamental frequency and intensity variation, the F_0 or intensity range for each utterance (maximum–minimum) was divided by the corresponding mean for that item and then averaged across speakers, by emotion.

[c] Based on responses of 26 English participants in an 8-choice forced identification task including an open response category, where chance performance = 12.5%.

[d] Represents the perceived intensity of the intended target emotion when correctly identified by the 26 participants in the forced-choice face identification task.

[e] As rated by a separate group of 14 English participants using an 11-point positive–negative scale.

iment as a whole. Following Pell (2005b), each experimental prime utterance was normalized to a peak intensity of 70 dB to control for unavoidable differences in the sound level of the source recordings across speakers. Each prime was then cut from the onset of the stimulus to form utterance "fragments" of two distinct durations: 600 ms and 1000 ms. Prior to editing, the average duration of selected utterances was 1462 ms for *happy*, 1408 ms for *sad*, and 1264 ms for the set of *neutral* stimuli. All auditory stimuli were edited using Praat software.

2.2.2. Facial stimuli

Experimental targets were selected from an inventory of digital bitmaps portraying emotional facial expressions posed by three male and three female actors (Pell, 2002). The vast majority of these tokens were also employed in our previous study of native emotion processing (Pell, 2005b). Half of the targets were 'true' facial expressions of emotion which reliably depicted a *happy* ($n = 12$) or *sad* ($n = 12$) facial expression, as determined by a consensus rate of 80% correct target identification in our validation study involving 32 healthy adults (Pell, 2002). The other half of the targets were 'false' emotional expressions ($n = 24$) which were produced by the same actors through movements of the brow, mouth, jaw, and lips but which could not be identified as a recognizable emotion by at least 60% of the validation group. Examples of 'true' and 'false' face targets portrayed by one female actor are presented in Fig. 1. In addition, an overview of major perceptual features associated with both the prosodic primes and facial targets selected for the experiment is supplied by Table 1.

2.3. Experimental design

The experiment was composed of 288 trials in which a prosodic prime was systematically paired with a face target. Following each trial, the participant always judged whether the facial expression represents a true expression of emotion (YES/NO response). Half of the trials ended with a 'true' facial expression of emotion as the target (YES trials) and half ended with a 'false' facial expression as the target (NO trials). Prime and target stimuli were posed by an equal number of male and female actors and only stimuli

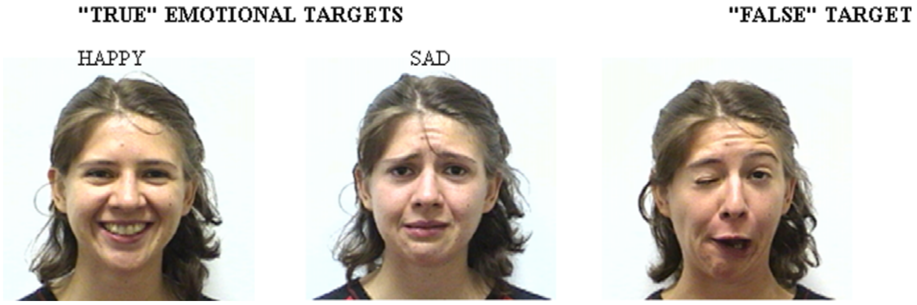

Fig. 1. Examples of 'true' and 'false' facial expressions posed by one female encoder.

produced by members of the same sex were combined within trials. In addition, the pairing of prosody-face events within trials was conducted separately for the two prosody duration conditions (600 ms, 1000 ms) so that the two prosodic fragments prepared from a given utterance were always paired with the same face tokens according to experimental conditions described below.

For YES trials which terminated in a *happy* or *sad* facial expression ($n = 144$), prosody-face events were combined to form a set of "congruent", "incongruent", and "neutral" trials. For *congruent* trials, the 12 happy and the 12 sad prosodic primes were combined with 'true' facial targets depicting the same target emotion (i.e., *sad–sad* and *happy–happy* trials), separately for the 600 ms and 1000 ms prime condition. *Incongruent* trials employed the same prime-target events but combined these to form an emotional mismatch between the two channels (*sad–happy* and *happy–sad*). *Neutral* trials consisted of the 12 *neutral* prosody primes matched with each of the 12 *happy* and 12 *sad* facial targets (*neutral–sad* and *neutral–happy*). These methods resulted in a total of 48 trials in each of the congruent, incongruent, and neutral conditions (2 emotions × 2 prosody durations × 12 items). A comparable number of NO trials was prepared by combining each of the 12 *happy, sad,* and *neutral* prosodic primes with two distinct 'false' face targets (from a total of 24 'false' targets selected for the experiment), again separately for the two prosody duration conditions. All YES and NO trials were intermixed and pseudo-randomly inserted into nine separate blocks consisting of 32 trials each. Since each prime stimulus repeated four times and each face stimulus repeated six times within the experiment, each block was inspected to avoid exact stimulus repetitions. Some of the blocks were also adjusted to ensure that trials represented a relatively equal proportion of male to female actors, true and false face targets, 600 and 1000 ms prosodic primes, and congruent, incongruent and neutral trials.

2.4. Procedure

Subjects were tested individually in a quiet, well-lit room during a single 45 min session. SuperLab 2.0 presentation software (Cedrus Corporation) was used to administer the experiment and to record accuracy and latency responses. Auditory stimuli were presented through high-quality adjustable headphones connected to a laptop computer, whereas face targets were displayed on a 37.5 cm wide computer monitor. At the onset of each trial, a visual fixation marker was presented on the computer screen for 350 ms, followed by a 500 ms silent interval, the prosodic fragment was played through the headphones, and the face target was presented on the computer screen immediately at the *offset* of the prosodic stimulus. The facial target remained visible on the computer screen until the participant pressed one of two buttons labelled YES or NO on a Cedrus 7-button response box.

Following previous FADT administrations, participants were instructed to attend closely to the nature of the facial expression presented on the computer screen and to decide whether or not the facial expression represents an emotion by pressing the appropriate button in each case. Participants were informed that they would hear sentences which would sound "foreign" but that their goal was to concentrate closely on the face and to judge the emotional status of the facial expression as accurately and as quickly as possible. Response speed (ms) and accuracy were recorded. To familiarize subjects with the stimuli and experimental procedure, two practice blocks were first run in which the participant learned to execute facial affect decisions to isolated faces (practice block 1) and to faces preceded by prosodic fragments resembling those in the experiment (practice block 2). During the practice blocks only, the computer provided written feedback about the subjects' accuracy ("Incorrect") and/or speed of their facial affect decisions ("Please try to respond faster"). After practice, the nine experimental blocks were presented in a pre-established random order which was varied across subjects, separated by short rest breaks between blocks and in the middle of the session.

3. Results

The overall error rate in the experiment was 6.96% (±6.58) of all trials, with an error rate of 7.08% (±8.71) for YES trials and 6.83% (±8.75) for NO trials for the 50 participants. Following Pell (2005b), data for one male participant whose experimental error rate exceeded 33.3% overall were excluded from further analysis owing to apparent difficulties to understand the nature of the task. Latency data for the 49 remaining subjects were normalized by removing values greater than 2000 ms or shorter than 300 ms from the data set (<0.1% of total values) and by replacing individual subject latencies greater than 2 standard deviations from the conditional mean with those equal to 2 s.d. in the appropriate direction (4.0% of total values). Only latencies for trials which resulted in correct facial affect decisions were analysed. The mean error rates and response times for 'true' and 'false' facial expressions of emotion are summarized in Table 2, as a function of relevant prime and target variables.

3.1. Processing emotional prosody in a foreign language

To test whether the emotional meaning and duration of Arabic speech influenced decisions about 'true' face targets, accuracy and latency responses corresponding to YES trials were examined through two separate 2 × 2 × 2 ANOVAs with repeated measures on Face (*happy, sad*), Prosody (congruent, incongruent), and Prosody Duration (600, 1000). Data for *neutral* prosody were omitted to focus these analyses on the nature and time course of emotion congruency effects for *happy* and *sad* expressions which were of greater theoretical interest (although see Table 2

Table 2
Mean errors (%) and latencies (ms) to make facial affect decisions when preceded by foreign (Arabic) prosody when the primes measured 600 or 1000 ms in duration, according to emotional attributes of the prime-target

Measure	Emotion of Face	Prosody duration	Emotion of prosody		
			Happiness	Sadness	Neutral
Errors	Happiness	600	0.9	1.4	1.5
		1000	0.9	1.2	1.7
	Sadness	600	9.4	10.0	12.4
		1000	13.8	9.9	12.2
	'False'	600	5.8	7.7	4.6
		1000	4.6	9.6	5.9
Latencies	Happiness	600	600	620	609
		1000	604	606	595
	Sadness	600	723	730	737
		1000	736	729	721
	'False'	600	713	703	705
		1000	705	690	699

Note. Conditions of congruent prosody-face emotional attributes are indicated in shaded cells.

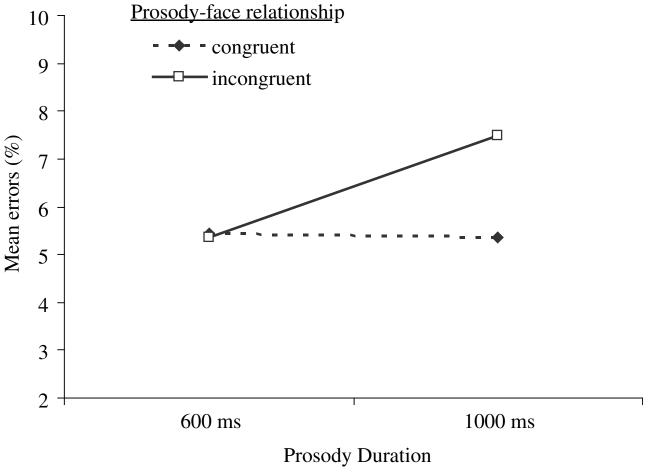

Fig. 2. Influence of congruent versus incongruent emotional features of prosody on facial affect decision errors as a function of the prosody duration.

and Footnote 4 for details about the neutral stimuli). All significant main and interactive effects from the ANOVAs were elaborated *post hoc* using Tukey's HSD comparisons ($p < .05$) and estimates of effect size were computed as prescribed by Rosenthal (1991).

Errors—The ANOVA performed on the conditional error proportions indicated that the emotional value of the Face had a significant main effect on decision accuracy, $F(1,\ 48) = 36.87$, $p < .001$, $r = 0.66$. Post hoc analyses established that decisions about *sad* face targets were considerably more error-prone (10.76%) than those about *happy* face targets (1.06%). In addition, the accuracy of facial affect decisions varied significantly according to the emotional relationship and duration of the prosody in the forms of a Prosody × Prosody Duration interaction, $F(1,\ 48) = 4.86$, $p < .05$, $r = 0.30$. Post hoc inspection of the interaction revealed that when prosodic information was emotionally congruent rather than incongruent with the target face, participants made significantly fewer errors when processing face targets; however, this emotion congruency effect was isolated to the condition when prosody endured for 1000 ms. When prosodic primes lasted 600 ms there was practically no difference in the effect of congruent versus incongruent emotional features on facial affect decisions, as illustrated in Fig. 2. No further main or interactive effects were produced by the analysis (all p's > .05).

Latencies—The ANOVA on response latencies indicated that the speed of facial affect decisions was also influenced by the emotion of the Face in the form of a significant main effect, $F(1,48) = 135.07$, $p < .0001$, $r = 0.86$. Post hoc comparisons indicated that *happy* faces ($M = 607$ ms) were correctly judged to be 'true' emotional expressions more quickly than *sad* faces ($M = 730$ ms) which took more than 120 ms longer on average. The ANOVA yielded no significant main or interactive effects which would imply that the prosody-face relationship or Prosody Duration were predictors of latency performance when listening to Arabic utterances (all p's > .05).[4]

3.2. Effects of language experience on vocal emotion processing

Our main analysis tested whether emotional features of prosody in a foreign language (Arabic) systematically influence emotional face processing for English-speaking listeners, revealing conditions in which this in fact appeared to be true. Since these methods and analyses largely replicated those of Pell, 2005b who presented *native* prosody to a comparable group of 45 English speakers, we undertook secondary analyses to directly compare the performance measures obtained in the two studies. These comparisons represent a unique opportunity to inform whether emotional meanings of prosody are processed in a similar manner and according to a similar time course between the two languages of interest.

[4] To exemplify how *neutral* prosody in a foreign language influenced facial affect decisions, a separate 2 × 3 × 2 ANOVA involving Face (happy, sad), Prosody Emotion (happy, sad, neutral), and Prosody Duration (600, 1000) was performed on the response errors and latencies. For *errors*, the significant main effect of Face (happy < sad) was reproduced by this analysis, $F(1,48) = 37.56$, $p < .001$. There was also a significant interaction of Face, Prosody Emotion, and Prosody Duration on error patterns, $F(2,96) = 3.12$, $p = .05$. Post hoc (Tukey's) comparisons indicated that decisions about sad faces led to greater errors when preceded by happy prosody in the 1000 ms versus 600 ms condition. In the 1000 ms condition only, sad faces were judged less accurately when preceded by happy than sad prosody (there were no accuracy differences between neutral prosody and either happy or sad prosody). For *latencies*, in addition to the main effect of Face, $F(1,48) = 145.18$, $p < .0001$, a significant interaction of Prosody Emotion and Prosody Duration emerged, $F(2,96) = 4.41$, $p = .02$. Post hoc comparisons showed that facial affect decisions were significantly faster in the 1000 ms versus the 600 ms condition when preceded by neutral prosody (no differences were observed across Prosody Duration conditions for happy or sad prosody). No further main or interactive effects were significant for either analysis.

Table 3
Mean errors (%) and latencies (ms) to make facial affect decisions when preceded by native (English) prosody when the primes measured 600 or 1000 ms in duration, according to emotional attributes of the prime-target (reproduced from Pell, 2005b)

Measure	Emotion of Face	Prosody duration	Emotion of prosody		
			Happiness	Sadness	Neutral
Errors	Happiness	600	2.0	1.4	2.8
		1000	2.4	2.4	2.8
	Sadness	600	19.2	14.3	15.5
		1000	17.3	14.1	15.9
	'False'	600	4.5	5.9	5.9
		1000	5.8	5.3	4.4
Latencies	Happiness	600	533	568	570
		1000	571	589	602
	Sadness	600	695	666	683
		1000	736	704	718
	'False'	600	639	657	655
		1000	666	678	686

Note. Conditions of congruent prosody-face emotional attributes are indicated in shaded cells.

To compare the effects of foreign (Arabic) versus native (English) prosody on facial affect decisions, two further ANOVAs were run on the error and latency measures, respectively; these analyses employed the current data as the foreign language condition (reported in Table 2) and the relevant data reported by Pell (2005b) as the native language condition (these measures are replicated in Table 3). As reported by Pell (2005b), data analysed in the native English condition were based on the performance of 41 participants (21 female, 20 male) for errors and 34 participants (17 female, 17 male) for latencies, prepared in the exact same manner as conducted here. Separate *t*-tests indicated that there were no significant differences in the mean age, $t(89) = -0.51$, $p = .61$, or mean education, $t(89) = -1.59$, $p = .12$, of the 42 English participants who contributed to measures reported by Pell (2005b) and the 49 English participants included in analyses here. Separate $2 \times 2 \times 2 \times 2$ mixed ANOVAs then considered the fixed effect of Language (native, foreign) on repeated measures of Face (*happy, sad*), Prosody (congruent, incongruent), and Prosody Duration (600 ms, 1000 ms) for each dependent measure.

Errors—The ANOVA on errors reconfirmed the pattern observed separately for each dataset indicating that *happy* faces were always judged more accurately than *sad* faces, representing a main effect for Face, $F(1,88) = 53.84$, $p < .001$, $r = 0.62$. Overall, the emotional relationship of the prosody to the face had a significant influence on the accuracy of facial affect decisions in the form of a Prosody main effect, $F(1,88) = 10.30$, $p < .01$, $r = 0.32$. In addition, there were significant interactions of Face x Prosody, $F(1, 88) = 11.54$, $p < .001$, $r = 0.34$, and Face × Prosody × Language, $F(1, 88) = 3.98$, $p < .05$, $r = 0.21$. Post hoc inspection of the three-way interaction revealed that judgments about *sad* (but not *happy*) facial expressions were significantly more error-prone in the native compared to the foreign language condition, irrespective of the prosody value. Also, *sad* faces preceded by an emotionally congruent rather than incongruent prosody led to significantly fewer errors, although this emotion congruency effect was only witnessed when the prosody was native (English) rather than foreign (Arabic).

Finally, the interaction of Prosody × Prosody Duration × Language was significant for this analysis, $F(1, 88) = 4.62$, $p < .05$, $r = 0.22$. Somewhat surprisingly, post hoc comparisons showed that facial affect decisions were always significantly more accurate when preceded by foreign than native prosody as the prime stimulus (with the exception of emotionally incongruent trials presented for 1000 ms, where a corresponding trend was observed). Of even greater interest, the amount of exposure to prosodic primes necessary for emotion congruency effects to emerge in the accuracy measures varied significantly by language: when participants heard native prosody, significant priming of the face occurred when the prosody lasted 600 ms in duration; in contrast, when participants heard foreign prosody, significant priming occurred only when the prosody lasted 1000 ms. The impact of language on these patterns is illustrated in Fig. 3a.

Latencies—The $2 \times 2 \times 2 \times 2$ ANOVA performed on response latencies yielded a main effect of Face, $F(1, 81) = 254.24$, $p < .0001$, $r = 0.87$, Prosody, $F(1, 81) = 19.93$, $p < .001$, $r = 0.44$, and Prosody Duration, $F(1, 81) = 7.47$,

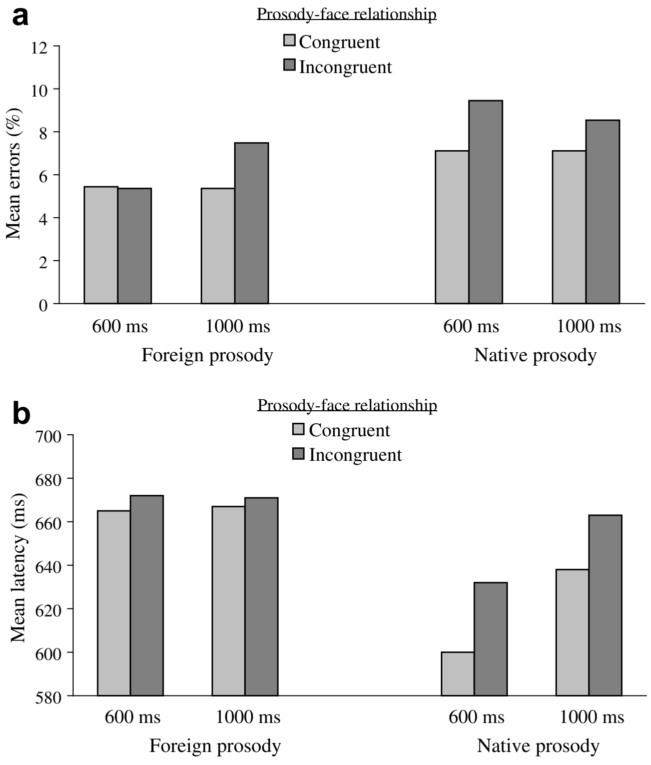

Fig. 3. Mean (a) errors and (b) latencies of facial affect decisions when primed by foreign (Arabic) versus native (English) prosody, according to the emotional relationship of the prosody and face and relevant features of the prime.

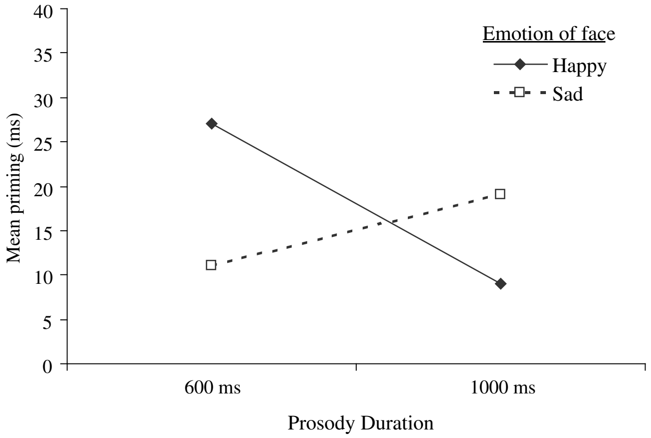

Fig. 4. Amount of emotional priming observed for happy and sad face targets according to the prosody duration (across languages, where priming = incongruent–congruent).

$p < .01$, $r = 0.29$. In addition, there was an interaction of Prosody Duration and Language, $F(1, 81) = 7.16$, $p < .01$, $r = 0.28$. When the emotional prosody was native, correct facial affect decisions were rendered significantly faster when the primes lasted 600 ms ($M = 616$ ms) as opposed to 1000 ms ($M = 650$ ms). However, when the emotional prosody was foreign, there was no difference in decision speed as a function of prosody duration ($M_{600} = 668$ ms; $M_{1000} = 669$ ms).

The interaction of Prosody × Language was also significant for response latencies, $F(1, 81) = 8.94$, $p < .01$, $r = 0.32$. Facial emotions were always judged to be 'true' expressions more quickly when preceded by native (English) than foreign (Arabic) prosody (contrary to the influence of Language on certain patterns in the accuracy data). More critically, emotion congruency effects on decision latencies were only significant when listeners heard native rather than foreign prosody as the prime stimulus; these patterns are demonstrated in Fig. 3b. Finally, the ANOVA yielded an interaction of Face × Prosody × Prosody Duration, $F(1, 81) = 5.04$, $p < .05$, $r = 0.24$. Irrespective of language, *happy* faces showed priming by an emotionally congruent prosody only when the prosodic information endured for 600 ms, whereas *sad* faces showed priming only when the prosody endured for 1000 ms. The impact of specific emotions on the time course of emotion congruency effects is shown in Fig. 4.

4. Discussion

To date, there is little empirical evidence that prosodic inflections of a language unknown to a listener convey meaningful emotional information between cultures when the foreign language is processed in an implicit manner (i.e., "on-line"). The Facial Affect Decision Task (FADT) has proven successful at indexing the emotional meanings of prosody activated when listeners are exposed to their *native* language, as inferred by emotion congruency effects when listeners judge an emotionally congruent target face (Pell, 2005a,b). In keeping with these previous methods, we limited the scope of our new study to two relatively non-controversial emotions, *happy* and *sad*, to focus discussion on how English listeners are influenced by the emotional value *and* duration of vocal emotion expressions encoded in a foreign language. Our new results, which can be readily compared to published data on English prosody (Pell, 2005b), permit a number of insights about the role of language and culture on prosody processing, and on the time course of vocal emotion processing in a native versus an unfamiliar language.

4.1. Universal and cultural factors in vocal emotion processing from speech

There were strong indications in our new data that English listeners engaged in a meaningful analysis of emotional prosody and accessed underlying features of these expressions in certain conditions, despite the fact that prosodic cues were embedded in an unfamiliar and typologically distinct language such as Arabic. Principally, we found that the accuracy of facial affect decisions was systematically biased by the emotional relationship of the prosody to the face, yielding reliable effects of emotion congruency in key conditions (i.e., when prosody endured for 1000 ms). This effect, while clearly influenced by our experimental manipulations as elaborated below, establishes that English listeners did engage in emotional processing of vocal cues when presented in a foreign language context. As such, our findings imply that vocal expressions of emotion possess certain universally recognizable characteristics (Scherer et al., 2001) which are implicitly detected during on-line speech perception, even when listeners lack familiarity with the input language and its conventions for cultural expression.

Although English listeners demonstrated sensitivity to the emotional content of Arabic utterance primes, it was obvious that emotion congruency effects produced by foreign prosody were less pervasive in our dependent measures than in previous FADT studies which presented the listeners' native language, English (Pell, 2005a,b). When exposed to Arabic speech, emotion congruency between the prosody and the face facilitated the accuracy of facial affect decisions only when at least 1000 ms of the prosodic information was encountered; in fact, the influence of congruent versus incongruent prosody on face judgements was indistinguishable in the condition that presented Arabic prosody for 600 ms in duration (review Fig. 2). These patterns contrast significantly from those corresponding to native prosody (Pell, 2005b) where maximal congruency effects were associated with prosodic primes lasting 600 ms in duration.

As a starting point, our direct comparisons between languages therefore argue that when vocal expressions are encoded in an unfamiliar language, underlying knowledge about emotional prosody is not sufficiently activated by a relatively short (600 ms) speech sample, precluding emotion-based priming in this environment (and presumably, hampering cross-cultural recognition of emotions encoded by foreign language speakers when additional cues are unavailable). This conclusion, while in need of verification, fits with data reported by Beier and Zautra (1972) who argued that recognition of vocal emotions in a foreign language requires increased exposure to relevant cues pertaining to vocal emotions than when processing the listener's native language. In line with comments by Scherer et al. (2001), one can speculate that presentation of a highly disparate language such as Arabic in the current study led to marked delays in the application of universal "inference rules" for understanding vocal emotions in our English listeners who had no experience with members of this culture.

In further support of this contention, it is perhaps relevant that Arabic prosody did not serve to prime facial decision *latencies* in either the 600 ms or the 1000 ms prosody duration conditions (which differed minimally), and correct facial affect decisions were generally slower following non-native prosody overall. In contrast, recall that an emotion congruency effect was strongly reflected in facial decision latencies following native prosody in both duration conditions (Pell, 2005b). These comparative findings indicate that, although the linguistic and cultural distance between Arabic and English did not prevent activation of emotional information when Arabic sentences were heard (as dictated by emotional congruency effects on accuracy rates), the cross-cultural transmission of emotion during speech processing can not be carried out without a certain level of 'noise' entering the system (Beier and Zautra, 1972). Part of this 'noise' could pertain to interference from acquired, culturally-appropriate models for expressing vocal emotions in speech which can not be applied as efficiently to non-native expressions, precluding differences in our response time measures. Alternatively, or in addition, noise induced by a foreign language could be introduced by *non-emotional* factors such as differences in the phonological or intonational structure of the exposed language which hinder auditory speech processing at a more basic level (Mesquita and Frijda, 1992; Scherer et al., 2001). These language-related differences would yield subsequent costs on the speed for applying inference rules dedicated to recognizing vocal emotions, which incidentally, would not be detectable in studies that have employed a forced-choice ("off-line") recognition format.

Interestingly, we noted that English listeners tended to make fewer facial affect decision errors overall in the foreign rather than in the native language condition, although this advantage began to disappear as the prosodic stimulus became longer (i.e., at 1000 ms, where priming began to emerge for foreign prosody). In keeping with our claim that vocal emotion processing for Arabic was prone to 'noise' and therefore delayed, it is possible that procedures for engaging in an emotional analysis of the face target were subject to less interference by prosodic features in the foreign language condition because the underlying emotional meanings were not fully activated in memory. As well, listeners may have made fewer errors in the foreign as opposed to the native language condition because only the pseudo-words in the native language condition would be perceived as "odd" in relation to known word candidates in English, promoting some interference at the stage of lexical-semantic processing. The precise meaning of these patterns cannot be stated with certainty until more cross-cultural data are gathered. In particular, it would be insightful to examine whether Arabic *listeners* display similar language-related differences in the accuracy and latency measures when presented with English and Arabic expressions of emotion if the design were to be reversed. Research of this nature is now ongoing in our lab.

4.2. Language-independent effects in the FADT

Although our principal focus was to understand how vocal *primes* conveying emotion were processed in different conditions, dependent measures in the FADT centre on conscious decisions about the emotional "acceptability" of a target facial expression. A ubiquitous finding in our FADT experiments is that face targets associated with a *negative* evaluation promote markedly greater errors and extended response latencies than face targets associated with a positive evaluation, irrespective of the emotion category (Pell, 2005a,b). There is a known tendency for negatively-evaluated or "threatening" events to temporarily interrupt cognitive–interpretive processes by holding or diverting attention away from immediate task goals (Dimberg and Ohman, 1996; Juth et al., 2005; Mathews and Mackintosh, 1998; Mogg and Bradley, 1998). As discussed in our earlier work, conscious attention to face targets in the FADT may index these attentional-evaluative processes to some extent, promoting "impaired" facial affect decisions in response to negative faces (especially expressions of *anger*, see Pell, 2005a).

There is also considerable evidence that during tasks of recognizing or *categorizing* facial expressions—which is the goal of the FADT—emotionally positive facial expressions are processed substantially faster than emotionally negative facial expressions (see Leppänen et al., 2003 for an overview). Arguably, the advantage for positive facial expressions is due to the differential effects of event valence on cognitive processes necessary for recognition or categorization of the target event; negatively valenced cues require more extensive cognitive analysis than positive events which is time consuming (Leppänen and Hietanen, 2004). This account is potentially consistent with our data which demonstrate the "happy" face advantage in a novel evaluative context: one in which participants render a facial affect decision about the target expression (*Is the facial expression an emotion?*). As demonstrated previously (Pell, 2005a,b), these general response tendencies associated with the valence of the face target showed little impact of the emotional value or duration of prosodic primes, whether the prime stimuli were native, English utterances or produced in a foreign language.

Interestingly, in addition to these anticipated effects of our face stimuli, the comparative analysis of foreign versus native prosody revealed a significant three-way interaction involving the emotional value of the prosody and the face (happy or sad) with prosody duration; these patterns hint at potential differences in the time course for activating knowledge about discrete vocal expressions of emotion independent of language. This possibility was suggested by several patterns in the dependent measures, especially response latencies which demonstrated prosody-face priming only for *happy* in the 600 ms condition and only for *sad* in the 1000 ms condition (see also Pell, 2005b). This pattern of priming implies that recognizing *sadness* from vocal cues evolves somewhat more slowly than understanding *happiness*, a hypothesis that can not be evaluated based on existing data derived from most forced-choice recognition tasks. Certainly, it is well accepted that acoustic-perceptual parameters that signal *joy* and *sadness* in the voice are highly distinct, and one of the hallmark difference of *sad* vocalizations is a relatively slow speaking rate (Banse and Scherer, 1996; Pell, 2001; Juslin and Laukka, 2003; Williams and Stevens, 1981). Thus, representative changes in pitch, intensity, and other prosodic elements corresponding to *sad* expressions tend to be modulated over an extended time frame. It is possible that recognizing sadness in the isolated vocal channel takes more time than recognizing joy or some other emotions, pending further experiments which test this idea in an *apriori* manner employing on-line behavioral or electrophysiological approaches for studying vocal emotion recognition.

4.3. Recognizing emotional prosody across languages: future directions

Researchers have long debated whether facial expressions of emotion display pan-cultural elements in recognition (Ekman, 1994; Izard, 1994; Russell, 1994) and this debate is equally germane to the question of how humans interpret vocal expressions of emotion embedded in spoken language. The behavioral measures we report here which revolve around *implicit* processing of the prosodic stimulus reinforce the likelihood that vocal cues to emotion, even in a foreign language, contain recognizable elements which signal whether a speaker is *happy* or *sad* (Scherer et al., 2001), in spite of the fact that the vocal expressions were encoded according to norms shared by a distinct linguistic-cultural group. Nonetheless, processes for understanding vocal emotions in a foreign language appear to require more exposure to prosodic cues for underlying meanings to be detected, and unfamiliar linguistic and paralinguistic features of the target language appear to have a general cost on the speed for inferring emotions cross-culturally. These claims, while advanced cautiously in the current context and in need of replication, encourage new research which considers the conjoint influences of ‘universal’ as well as culturally-determined processes in vocal emotion processing from speech.

Our study can be elaborated in the future by testing listeners from different linguistic backgrounds and by presenting a larger array of emotional expressions, rather than only two presumably “opposing” emotions (*happiness* and *sadness*). In the latter case, this meant that our experiment was not constructed to differentiate what *form* of representational detail(s) about our prime stimuli may have been activated by English listeners when exposed to vocal expressions of emotion in a foreign language. One can speculate that in our experimental conditions which showed significant cross-cultural priming, English listeners registered emotion-specific details when listening to Arabic speech, since only these features would be likely to promote more efficient analysis of the *representational status* of prototypical expressions of emotion in the face (see Pell, 2005a for further arguments based on results for four distinct emotion categories). For example, it has been proposed that discrete emotions impose an organizational structure on semantic memory (Bower, 1981; Niedenthal et al., 1994) and one can argue that this knowledge is used to infer the intended meaning of emotion displays irrespective of the language which carries the vocal expression. If true, our data imply that procedures or “inference rules” (Scherer et al., 2001) for accessing this knowledge are susceptible to interference and delay when a foreign language is encountered, owing to cultural and language-related differences which shape how vocal expressions of emotion are communicated and understood.

Acknowledgement

This research was financed by the Natural Sciences and Engineering Research Council of Canada (Discovery Grant to M.D.P.) and a summer bursary from the McGill Faculty of Medicine (to V.S.). The support of McGill University to the first author is gratefully acknowledged (William Dawson Scholar award).

References

Albas, D., McCluskey, K., Albas, C., 1976. Perception of the emotional content of speech: a comparison of two Canadian groups. J. Cross-Cultural Psychol. 7, 481–489.

Bachorowski, J., 1999. Vocal expression and perception of emotion. Curr. Directions Psychol. Sci. 8 (2), 53–57.

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

Beier, E., Zautra, A., 1972. Identification of vocal communication of emotions across cultures. J. Consult. Clin. Psychol. 39 (1), 166.

Bower, G.H., 1981. Mood and memory. Amer. Psychol. 36 (2), 129–148.

Breitenstein, C., Van Lancker, D., Daum, I., 2001. The contribution of speech rate and pitch variation to the perception of vocal emotions in a German and an American sample. Cognition Emotion 15 (1), 57–79.

de Gelder, B., Vroomen, J., 2000. The perception of emotions by ear and by eye. Cognition Emotion 14 (3), 289–311.

Dimberg, U., Ohman, A., 1996. Beyond the wrath: Psychophysiological responses to facial stimuli. Motiv. Emotion 20 (2), 149–182.

Ekman, P., 1994. Strong evidence for universals in facial expressions: a reply to Russell's mistaken critique. Psychol. Bull. 115, 268–287.

Ekman, P., Friesen, W., 1971. Constants across cultures in the face and emotion. J. Personality Soc. Psychol. 17 (2), 124–129.

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

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. J. Personality Soc. Psychol. 53 (4), 712–717.

Elfenbein, H., Ambady, N., 2002. On the universality and cultural specificity of emotion recognition: a meta-analysis. Psychol. Bull. 128 (2), 203–235.

Frick, R.W., 1985. Communicating emotion: the role of prosodic features. Psychol. Bull. 97 (3), 412–429.

Innes-Ker, A., Niedenthal, P., 2002. Emotion concepts and emotional states in social judgment and categorization. J. Personality Soc. Psychol. 83 (4), 804–816.

Izard, C.E., 1977. Human Emotions. Plenum Press, New York.

Izard, C.E., 1994. Innate and universal facial expressions: evidence from developmental and cross-cultural research. Psychol. Bull. 115 (2), 288–299.

Juslin, P., Laukka, P., 2003. Communication of emotions in vocal expression and music performance: different channels same, code? Psychol. Bull. 129, 770–814.

Juth, P., Lundqvist, D., Karlsson, A., Ohman, A., 2005. Looking for foes and friends: perceptual and emotional factors when finding a face in the crowd. Emotion 5 (4), 379–395.

Kramer, E., 1964. Elimination of verbal cues in judgments of emotion from voice. J. Abnormal Soc. Psychol. 68 (4), 390–396.

Laukka, P., 2005. Categorical perception of vocal emotion expressions. Emotion 5 (3), 277–295.

Leppänen, J., Hietanen, J., 2004. Positive facial expressions are recognized faster than negative facial expressions but why? Psychol. Res. 69, 22–29.

Leppänen, J., Tenhunen, M., Hietanen, J., 2003. Faster choice-reaction times to positive than to negative facial expressions: the role of cognitive and motor processes. J. Psychophysiol. 17, 113–123.

Massaro, D., Egan, P., 1996. Perceiving affect from the voice and the face. Psychonom. Bull. Rev. 3 (2), 215–221.

Mathews, A., Mackintosh, B., 1998. A cognitive model of selective processing in anxiety. Cognitive Therapy Res. 22 (6), 539–560.

McCluskey, K., Albas, D., Niemi, R., Cuevas, C., Ferrer, C., 1975. Cross-cultural differences in the perception of the emotional content of speech: a study of the development of sensitivity in Canadian and Mexican children. Develop. Psychol. 11, 551–555.

Mesquita, B., Frijda, N., 1992. Cultural variations in emotions: a review. Psychol. Bull. 112 (2), 179–204.

Mogg, K., Bradley, B., 1998. A cognitive-motivational analysis of anxiety. Behaviour Res. Therapy 36, 809–848.

Niedenthal, P., Setterlund, M., Jones, D., 1994. Emotional organization of perceptual memory. In: Niedenthal, P., Kitayama, S. (Eds.), The Heart's Eye: Emotional Influences in Perception and Attention. Academic Press, New York, pp. 87–113.

Onifer, W., Swinney, D., 1981. Accessing lexical ambiguities during sentence comprehension: effects of frequency of meaning and contextual bias. Memory Cognition 9, 225–236.

Pell, M.D., 2001. Influence of emotion and focus location on prosody in matched statements and questions. J. Acoust. Soc. Amer. 109 (4), 1668–1680.

Pell, M.D., 2002. Evaluation of nonverbal emotion in face and voice: some preliminary findings on a new battery of tests. Brain Cognition 48, 499–504.

Pell, M.D., 2005a. Nonverbal emotion priming: evidence from the'facial affect decision task'. J. Nonverbal Behav. 29 (1), 45–73.

Pell, M.D., 2005b. Prosody-face interactions in emotional processing as revealed by the facial affect decision task. J. Nonverbal Behav. 29 (4), 193–215.

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. Abstr. Psychonom. Soc. 46th Ann. Meet. 10, 98.

Pell, M.D., Monetta, L., Paulmann, S., Kotz, S.A., submitted for publication-a. Recognizing emotions in a foreign language: a "case study" of monolingual Spanish listeners.

Pell, M.D., Paulmann, S., Dara, C., Alasseri, A., Kotz, S.A., submitted for publication-b. Factors in the recognition of vocally expressed emotions: a comparison of four languages.

Rosenthal, R., 1991. Meta-analytic procedures for social research. Appl. Soc. Res. Methods 6, 19.

Rossell, S.L., Nobre, A., 2004. Semantic priming of different affective categories. Emotion 4 (4), 354–363.

Russell, J.A., 1994. Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. Psychol. Bull. 115 (1), 102–141.

Scherer, K.R., 1986. Vocal affect expression: a review and a model for future research. Psychol. Bull. 99 (2), 143–165.

Scherer, K.R., Banse, R., Wallbott, H.G., Goldbeck, T., 1991. Vocal cues in emotion encoding and decoding. Motiv. Emotion 15 (2), 123–148.

Scherer, K.R., Banse, R., Wallbott, H., 2001. Emotion inferences from vocal expression correlate across languages and cultures. J. Cross-cultural Psychol. 32, 76–92.

Schroder, M., 2003. Experimental study of affect bursts. Speech Comm. 40, 99–116.

Swinney, D., 1979. Lexical access during sentence comprehension: (Re)consideration of contextual effects. J. Verbal Learning Verbal Behav. 18, 645–659.

Thompson, W., Balkwill, L.-L., 2006. Decoding speech prosody in five languages. Semiotica 158, 407–424.

Van Bezooijen, R., Otto, S., Heenan, T., 1983. Recognition of vocal expressions of emotion: a three-nation study to identify universal characteristics. J. Cross-Cultural Psychol. 14 (4), 387–406.

Vroomen, J., Driver, J., de Gelder, B., 2001. Is cross-modal integration of emotional expressions independent of attentional resources. Cognitive Affective Behav. Neurosci. 1, 382–387.

Williams, C.E., Stevens, K.N., 1972. Emotions and speech: Some acoustical correlates. J. Acoust. Soc. Amer. 52, 1238–1250.

Williams, C.E., Stevens, K.N., 1981. Vocal correlates of emotional states. In: Darby, J.K. (Ed.), Speech Evaluation in Psychiatry. Grune and Stratton, New York, pp. 221–240.

Wilson, D., Wharton, T., 2006. Relevance and prosody. J. Pragmatics 38, 1559–1579.