2005

Recognition of affective prosody by speakers of English as a first or foreign language

Christopher Dromey
Brigham Young University, dromey@byu.edu

Jose Silveira
University of Toronto

Paul Sandor
University of Toronto

Original Publication Citation

Dromey, C., Silveira, J. & Sandor, P. (2005). Recognition of affective prosody by speakers of English as a first or foreign language. Speech Communication, 47, 351-359

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Speech Communication 47 (2005) 351–359
SPEECH COMMUNICATION
www.elsevier.com/locate/specom

Recognition of affective prosody by speakers of English as a first or foreign language

Christopher Dromey [a,*], Jose Silveira [b,c], Paul Sandor [b,c]

[a] Department of Audiology and Speech-Language Pathology, Brigham Young University, 133 Taylor Building, Provo, UT 84602, USA
[b] Department of Psychiatry, Toronto Western Hospital, 399 Bathurst Street, Toronto, ON, Canada M5T 2S8
[c] Department of Psychiatry, University of Toronto, 250 College Street, Toronto, ON, Canada M5T 1R8

Received 27 February 2004; received in revised form 14 September 2004; accepted 28 September 2004

Abstract

Adults who were fluent in English, and who grew up speaking English or one of 21 other languages listened to words spoken with angry or neutral intonation. We measured the accuracy with which the listeners identified the intended emotion. English mother tongue (EMT) polyglots scored higher than other mother tongue (OMT) listeners, whereas EMT monoglots did not. Women were significantly more accurate than men across the three listener groups. There was a modest inverse correlation between accuracy and age. The learning of a second language may have helped the EMT polyglots develop additional perceptual skills in decoding speech emotion in their native language.

Keywords: Prosody; Emotion; Second language; Perception

1. Introduction

Most listeners recognize that the way in which something is said can be as important as the words that are used. A speaker's tone of voice carries a wealth of subtle acoustic cues about the emotion that accompanies the message. The non-verbal acoustic cues that convey the emotional valence of a speaker's communication are often referred to as affective prosody.

Over the past few decades, numerous studies have been conducted to evaluate the way speakers and listeners typically encode and decode emotional cues in speech. Some of these studies have focused on the ability of listeners to reliably detect differences between specific emotions, while others have sought to identify the acoustic cues that correspond to individual emotional states (Cummings and Clements, 1995; Frick, 1985). Experiments have involved the simulation of emotion by

* Corresponding author. Tel.: +1 801 422 6461; fax: +1 801 422 0197.
E-mail address: dromey@byu.edu (C. Dromey).

0167-6393/$ - see front matter
doi:10.1016/j.specom.2004.09.010

professional actors (Johnson et al., 1986) or untrained individuals (Pell, 2001). Other studies have been based on experiments with synthesized speech (Johnson et al., 1986; Murray and Arnott, 1993).

Studies of healthy listeners (Buchanan et al., 2000; Herrero and Hillix, 1990; McNeely and Parlow, 2001), as well as individuals with neural damage following stroke (Heilman and Gilmore, 1998; Pell and Baum, 1997; Ross et al., 1997) have led to the conclusion that the right hemisphere plays an important role in the processing of prosody as it conveys both linguistic and affective information. While both hemispheres are actively involved in processing emotional speech, a recent report described a pattern of lateralization that involves left hemisphere dominance for primarily linguistic tasks, and right hemisphere dominance for the emotional aspects of speech (Buchanan et al., 2000). While patients with deficits in the recognition of affective prosody may often have difficulty recognizing the facial expression of emotion, the neural pathways involved appear to be different (Starkstein et al., 1994). Experiments with individuals who have become globally aphasic due to unilateral left hemisphere pathology have shown that the ability to interpret affective prosody can be relatively spared, despite the loss of linguistic skills (Barrett et al., 1999).

The ability to recognize affect from prosody has been reported to change with age. Children gradually improve in their recognition skills, and in older adults this ability declines (Brosgole and Weisman, 1995; McCluskey and Albas, 1981). Indeed, an inability to recognize emotion from both facial and auditory cues has been suggested as a sign of incipient dementia (Brosgole and Weisman, 1995). A listener's mental health appears to influence the recognition of affective prosody. A study of school-aged boys diagnosed with depression found that they were less able than their healthy peers to recognize emotion in speech (Emerson et al., 1999).

A number of studies have shown that women generally perform better than men in decoding emotions. This has been found with the recognition of emotion from facial expression (Kirouac and Dore, 1985; Rotter and Rotter, 1988), as well as lexical emotion—the recognition of the inherent emotional quality of words (Grunwald et al., 1999). Women have also outperformed men in recognizing emotions portrayed via visual or auditory modalities, as well as when the two are combined (Crucian and Berenbaum, 1998; Gitter et al., 1972). A study of vocal affect, where actors read neutral passages in a variety of emotional tones, found that women were more accurate than men in their judgments (Bonebright et al., 1996).

Only a few reports have dealt with the recognition of vocal emotion across cultures. Some have found that individuals tend to be more accurate at recognizing emotional prosody in their own language (Albas et al., 1976; Beier and Zautra, 1972), as well as in the speech of more familiar speakers (Bachorowski, 1999), but that 'foreign' listeners are still able to reliably recognize emotional states from speech at levels far greater than chance (Beier and Zautra, 1972). One study that involved speech that had been low-pass filtered to remove all linguistic information—leaving the prosodic contours intact—revealed that Mexican listeners were better able to recognize emotion in speech than were their Canadian counterparts. This was true regardless of whether the speech they heard came from Mexican or Canadian speakers (McCluskey and Albas, 1981). This suggests that there may be important cultural differences that make some listeners more sensitive to emotional cues in speech. The ability to recognize emotion in a foreign language at all has led to the suggestion that in the vocal expression of emotion—as appears to be the case for facial emotion—there may be universal features that cross cultures (Van Bezooijen et al., 1983).

The prosodic features of speech may be significant during language acquisition. Linguistic prosody can focus the listener's attention on the most grammatically salient words, while affective prosody can reinforce the semantics of the message. Mehler and Christophe (1994) summarized evidence that infants as young as six months of age can distinguish between different languages on the basis of prosodic cues for utterances as short as two syllables in length. Since the rhythmic structure of languages can differ (e.g., syllables for romance languages, moras in Japanese, and strong/weak stress in English), it seems reasonable that individuals who have grown up in different language environments may be sensitive to different types of acoustic cues in the interpretation not only of the linguistic aspects of what they hear, but also the prosodic patterns in their own language.

The present study arose from the observation that some individuals from our ethnically diverse pool of listeners were more accurate than others during a listening task where they were asked to choose which emotion was being conveyed by an English speaker. This observation led to a decision to increase the sample size in order to evaluate the effect of language background on the performance of the task. The goal was to measure accuracy rates in the perception of affective prosody across listeners with diverse linguistic backgrounds compared with English mother tongue listeners. Because the listeners who grew up speaking languages other than English were by definition multilingual, the native English speakers who also spoke a foreign language were considered separately from those who spoke only English.

2. Method

2.1. Participants

The listeners were 142 adults who were fluent in both spoken and written English. They volunteered to participate by responding to English advertisements that were posted at the Toronto Western Hospital (Toronto, Ontario, Canada) which serves an ethnically diverse population. Some were hospital employees, while others were students in health-related professions or visitors in the hospital. Before performing the listening task, each participant completed questionnaires to determine the mother tongue and currently preferred language, age of English acquisition, spheres of use of English, and other languages spoken. The listeners also responded to questions evaluating educational background, mood and medication use.

Table 1 reports listener age by gender and language group: English mother tongue (EMT) monoglot, EMT polyglot or some other mother tongue (OMT). A one-way ANOVA revealed no differences in age between the groups. Table 2 reports the educational background of the listeners in the three groups. A one-way ANOVA revealed a significant difference in estimated years of education by group ($F = 3.592$, $p = .03$). A Bonferroni corrected post hoc test found that the length of education was higher for the EMT polyglots (15.5 years) than for the EMT monoglots (13.8 years). Other group differences were not significant. Appendix A reports the specific mother tongue languages of the listeners. The proximity of the hospital to Toronto's Chinatown and Portuguese neighborhoods is reflected in the distribution of languages. Potential participants were asked whether they had any history of speech, hearing or learning disorders, and any responding affirmatively were excluded from the study.

2.2. Equipment

The stimulus recording and listening tasks were conducted in a sound-attenuating booth. The words were recorded with a head-mounted microphone (AKG C-420) into a digital audio tape recorder (Panasonic SV-3800). The recorded words were subsequently re-digitized at 25 kHz into a Kay Elemetrics Computerized Speech Lab (CSL—model 4300B). The CSL unit was used

Table 1
Listener demographics, indicating frequency and age by gender and language group

	English mother tongue monoglot		English mother tongue polyglot		Other mother tongue	
	Male	Female	Male	Female	Male	Female
Count	16	37	11	21	24	33
Mean age (SD)	37.8 (8.2)	40.7 (11.5)	37.1 (10.4)	37.0 (10.1)	37.5 (15.6)	35.9 (11.1)

Table 2
Educational background of the listeners

	Estimated years[a]	English mother tongue monoglot (%)	English mother tongue polyglot (%)	Other mother tongue (%)
Some elementary	3	0	0	2
Elementary grad	5	0	3	0
Some high school	10	12	0	6
High school grad	12	14	3	8
Some post-secondary	13	18	25	16
College grad	14	25	16	17
University grad	16	23	13	22
Graduate degree	18	8	34	27
Postgraduate degree	21	0	6	2

Percentage of listeners in each group who reached each level of education.
[a] Participants reported education level by category, rather than by number of years. The "Estimated years" column shows the estimated equivalent number of years of schooling corresponding to each category, in order to allow correlation with accuracy scores.

for playback of the words via a loudspeaker at a comfortable listening level.

2.3. Speech materials

Single words, rather than sentences, were used in order to decrease the impact of sentence level prosodic cues on the listeners' perception of the stimuli. One to three syllable words were selected from a list of low imageability words, with most being one or two syllables. Because of their low imageability, the stimuli had a reduced likelihood of carrying an affective valence in and of themselves.

A professional actor, who was an EMT monoglot, spoke the words. He read through the same list of 117 selected words once in a neutral and once in an angry tone of voice. The authors listened to his speech during the recordings, and requested occasional repetitions until the specified affect was deemed convincing. Each word was saved as a separate computer file, and the amplitude of the words was normalized by digitally adjusting the decibel level up or down in the CSL program to avoid biasing listener response with differences in intensity. All words in both the angry and neutral condition were randomized together and played back to each listener in an individual session. A total of 234 words (with word and emotion occurring randomly) were presented in isolation with 2-second pauses between them in the same sequence for all listeners.

2.4. Procedure

The listeners were given the same written and spoken instructions. They were asked to listen to the words presented to them and to simply check off from a list the emotion they perceived the speaker to be communicating for that word. They were not told which emotions might be involved but were asked to select from the options of *anger*, *disgust*, *joy*, *fear*, *surprise* and *neutral*. Subjects thus selected from a total of six primary emotional states, two of which were present in the recorded stimuli, and four of which were foil options. Although the term 'neutral' does not refer to a specific emotion, this term has been used in previous work to denote the absence of any discernible affect (Bonebright et al., 1996).

2.5. Data analysis

The dependent variable was the number of correct responses, which was defined as the number of responses marked 'neutral' or 'angry' which matched the intended affect of the stimulus. The primary analysis goal was to determine the differences in response accuracy between EMT monoglots, EMT polyglots and the OMT listeners. Because previous studies have identified superior female performance in the identification of emotional stimuli (Gitter et al., 1972; Grunwald et al., 1999; Rotter and Rotter, 1988), the effect of listener sex on accuracy was tested. Since age has been linked with the ability to decode emotions in previous experiments, it was included as a covariate in the present analysis. To examine the nature and strength of the association between demographic variables and performance, Pearson correlations were computed between accuracy scores and the age and education level of all participants.

3. Results

A 3 × 2 ANOVA (3 listener groups: EMT monoglot, EMT polyglot, OMT; and 2 sexes) with age as a covariate revealed main effects for listener group ($F = 3.201$, $p = .044$), for sex ($F = 12.351$, $p = .001$) and the covariate of age ($F = 11.221$, $p = .001$). Sex was found not to interact with listener group. The number of correct responses for listeners in each group, separated by sex, are plotted in Fig. 1. Bonferroni corrected pairwise comparisons revealed a significant difference ($p = .039$) between the EMT polyglots and the OMT listeners. No other comparisons were significant. The EMT polyglots performed better than the OMT listeners, but the EMT monoglots did not differ significantly from the other two groups.

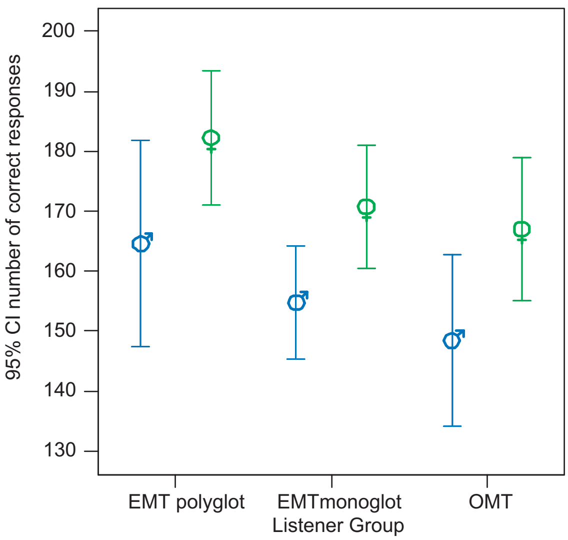

Fig. 1. Means and 95% confidence intervals for the number of correctly identified items for men and women in each listener group.

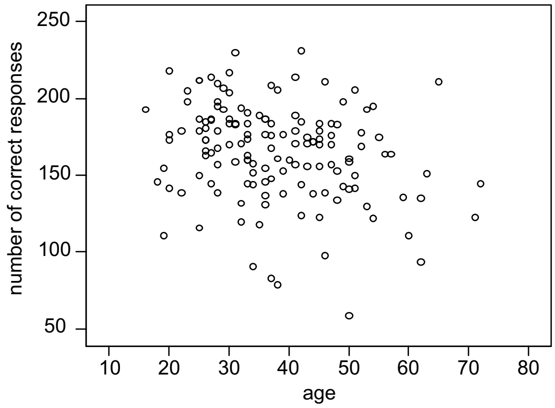

Fig. 2. Scatterplot of the number of correct responses as a function of listener age.

Pearson correlations examining the relationship between the number of correct responses and other participant variables yielded the following results. There was a modest but significant negative association between age and task performance (r −.249, $p = .003$), with younger listeners scoring slightly higher (see Fig. 2). There was no correlation between the length of education and accuracy scores.

4. Discussion

The present study differs from several earlier investigations of the perception of affective prosody across cultures. Previous work has shown that individuals who do not speak a language can still reliably identify the affective state of a speaker, even if their accuracy is lower than that of native speakers. This has been found to be true for both natural (Beier and Zautra, 1972) and filtered (Albas et al., 1976) speech. The listeners in the present study were individuals who were either native English speakers, or who had acquired English as a second language. As such, traditional concepts of fluency in the language itself should not have presented a barrier, as might have been the case with a completely foreign language.

Studies of the production and recognition of emotion have employed a variety of stimuli in order to answer their respective research questions. The use of an actor to produce emotional speech avoids the practical and ethical problems associated with inducing strong emotions in speakers in a laboratory setting (Scherer, 1995). However, some authors have cautioned that acted emotions may differ acoustically from those that occur naturally, and there is a risk of their becoming caricatures of the intended state (Greasley et al., 2000). On the other hand, recent work with evoked potentials has shown that listeners respond physiologically to brief simulated emotional vocalizations, even in the absence of other contextual cues (Bostanov and Kotchoubey, 2004). In a comprehensive review of the literature, Scherer (2003) noted that even though the portrayal of emotions by actors may be influenced by popular stereotypes of an emotion, listeners tend to be quite reliable in recognizing them: "Stereotypes are often generalized inference rules, based on past experience or cultural learning, that allow the cognitive economy of making inferences on the basis of very little information." (p. 237). However, it cannot be concluded that these experiences are necessarily comparable with more natural events. Therefore, care must be taken in drawing inferences from the present findings, because laboratory responses to word-level stimuli produced by an actor may not reflect fully the processes involved in everyday situations.

The results of the present study are consistent with previous accounts of female superiority in decoding emotion (Grunwald et al., 1999; Kirouac and Dore, 1985; Rotter and Rotter, 1988). As seen in Fig. 1, this pattern was consistent across listener groups, suggesting that language background and second language acquisition do not influence the ability of women to outperform men in recognizing vocal affect.

The EMT polyglot listeners in the present study were more accurate than the OMT listeners in their identification of the intended affect in spoken English words. The OMT listeners performed at a level well above chance, indicating that individuals who are not native speakers of a language can detect affective prosody at the single word level. It has been suggested that listeners use the sound stream segmentation approaches from their native language when learning a foreign language (Mehler et al., 1994). The tendency to apply familiar approaches to the interpretation of affective prosody may also explain why the OMT listeners scored lower than the EMT participants.

Our findings differ somewhat from those of McCluskey and Albas (1981) who suggested that there might be important cultural differences that make some listeners more sensitive to emotional cues regardless of language discordance. Their conclusions were drawn from a protocol that involved speech that had been low-pass filtered to remove all linguistic information—leaving the prosodic contours intact. It may be that their Mexican listeners were better than the Canadians in the task because the brain processes phrase level prosodic contours differently from spoken language, and the mechanism of prosody interpretation may not be the same for filtered speech as it is for natural speech. Consequently, extensive modulation of natural speech to the point of masking linguistic information may not allow the language centers of the brain to be operating to the extent that we intend to test them. The present findings do not allow us to address the suggestion that some cultures or language groups are more sensitive than others to emotional cues in speech, because the wide cultural diversity of our listeners may have prevented any clear patterns from becoming apparent.

Importantly, our findings also reveal that EMT polyglots were somewhat more accurate in identifying the speaker's intended emotional communication than were their monoglot counterparts (see Fig. 1), although this difference was not statistically significant. Indeed, this finding refutes the potential hypothesis that the OMT listeners were less accurate on the basis of "signal confusion" because of speaking more than one language. There appears to be an advantage in recognizing vocal affect for individuals who grew up with the native language of the speaker, and who also have learned a foreign language. It could be speculated that the experience of acquiring a second language develops in a person additional sensitivity to certain aspects of speech that carries over to native language tasks. Since phonology, semantics, and syntax vary across languages, a person learning a new language must attend to sounds that he or she did not grow up with. Rhythmic and prosodic differences also distinguish language groups (Mehler and Christophe, 1994), which suggest that individuals who have mastered a foreign tongue have acquired sensitivity to its prosody as well as the linguistic differences. This developed ability to attend more carefully to speech sounds could be responsible for the superior scores of the EMT polyglots in the present study, but would not explain why this improved sensitivity is not seen in OMT listeners. Alternatively, the superiority of EMT polyglots may reveal some individuals' inherent ability to identify emotional signals of speech, which may have contributed to their capacity and willingness to master a second language.

The literature on psychoacoustics in musicians suggests that their experience and training allow them to outperform individuals without musical training. The additional sensitivity that develops with musical training may result from focusing one's attention on sound. One study revealed that experienced orchestral conductors had superior ability to localize sounds coming from the periphery, which the authors attributed to the conductors' extensive training and experience (Muente et al., 2001). Recent experimental work indicates that the electrical response of the cortex to piano sounds was greater than to pure tones for piano players but not for control subjects who had no musical experience (Pantev et al., 2001). The EMT polyglots in the present study had the advantage of learning English early and may have developed additional sensitivity to the sounds of human communication when they learned a second language, that was not available to the EMT monoglots, allowing the polyglots to perform more accurately in identifying emotion from prosody. A psychoacoustic investigation of musical pitch perception found that listeners who grew up in England performed very differently from Californian listeners on the same task, leading the author to conclude that a person's language environment during language acquisition can influence musical perception (Deutsch, 1991). It seems reasonable to speculate that the language environments in which our listeners grew up could have influenced them in their perception of affective prosody.

The present work has several limitations that could be overcome in future research. The volunteer listeners varied greatly in their cultural and language backgrounds, and a more systematic sampling of perhaps a smaller number of language groups would allow greater control and a more straightforward interpretation of the data. The number of male and female listeners was not equal, and neither was there a balance in the educational level of the participants. Although level of education did not correlate with accuracy scores, the EMT polyglots nevertheless had on average more years of education than the other listeners. The present data do not allow definitive conclusions regarding the role of early second language exposure compared with formal classroom pedagogy in acquiring second language skills, nor which of these might have exerted a stronger role in developing sensitivity to vocal affect. Nevertheless, useful insights have been gained into the perception of emotional prosody in native and non-native speakers of English, including the relatively less accurate perception of intended emotional communications by learners of English as a second language and the perceptual advantage conferred upon native speakers who have acquired an additional language.

Appendix A

Mother tongue of the listeners

	Frequency	Percent
Afrikaans	1	0.7
Cantonese	17	12.0
English/Hindi[a]	1	0.7
English	81	57.0
Filipino/English[a]	1	0.7
Filipino	4	2.8
French	1	0.7
German	1	0.7
Gujarati	1	0.7
Hebrew	1	0.7
Hungarian	1	0.7
Indian/French	1	0.7
Italian	5	3.5

(*continued on next page*)

Appendix A (*continued*)

	Frequency	Percent
Italian/English[a]	1	0.7
Mandarin	1	0.7
Polish	4	2.8
Portuguese	9	6.3
Punjabi/English[a]	1	0.7
Punjabi	2	1.4
Russian	1	0.7
Slovak	1	0.7
Spanish	2	1.4
Tamil	2	1.4
Ukrainian, Polish	1	0.7
Urdu	1	0.7
Total	142	100

[a] Despite having two mother tongues, these individuals were counted as English mother tongue, since English was spoken from the age of language acquisition.

References

Albas, D.C., McCluskey, K.W., Albas, C.A., 1976. Perception of the emotional content of speech: a comparison of two Canadian groups. J. Cross Cult. Psychol. 7, 481–490.

Bachorowski, J., 1999. Vocal expression and perception of emotion. Curr. Dir. Psychol. Sci. 8, 53–57.

Barrett, A.M., Crucian, G.P., Raymer, A.M., Heilman, K.M., 1999. Spared comprehension of emotional prosody in a patient with global aphasia. Neuropsy. Neuropsychol. Behav. Neurol. 12, 117–120.

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

Bonebright, T.L., Thompson, J.L., Leger, D.W., 1996. Gender stereotypes in the expression and perception of vocal affect. Sex Roles 34, 429–445.

Bostanov, V., Kotchoubey, B., 2004. Recognition of affective prosody: continuous wavelet measures of event-related brain potentials to emotional exclamations. Psychophysiology 41, 259–268.

Brosgole, L., Weisman, J., 1995. Mood recognition across the ages. Internat. J. Neurosci. 82, 169–189.

Buchanan, T.W., Lutz, K., Mirzazade, S., Specht, K., Shah, N.J., Zilles, K. et al., 2000. Recognition of emotional prosody and verbal components of spoken language: an fMRI study. Cogn. Brain Res. 9, 227–238.

Crucian, G.P., Berenbaum, S.A., 1998. Sex differences in right hemisphere tasks. Brain Cogn. 36, 377–389.

Cummings, K.E., Clements, M.A., 1995. Analysis of the glottal excitation of emotionally styled and stressed speech. J. Acoust. Soc. Amer. 98, 88–98.

Deutsch, D., 1991. The tritone paradox: an influence of language on music perception. Music Percept. 8, 335–347.

Emerson, C.S., Harrison, D.W., Everhart, D.E., 1999. Investigation of receptive affective prosodic ability in school-aged boys with and without depression. Neuropsy. Neuropsychol. Behav. Neurol. 12, 102–109.

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

Gitter, A.G., Kozel, N.J., Mostofsky, D.I., 1972. Perception of emotion: the role of race, sex, and presentation mode. J. Soc. Psychol. 88, 213–222.

Greasley, P., Sherrard, C., Waterman, M., 2000. Emotion in language and speech: methodological issues in naturalistic approaches. Lang. Speech 43, 355–375.

Grunwald, I.S., Borod, J.C., Obler, L.K., Erhan, H.M., Pick, L.H., Welkowitz, J. et al., 1999. The effects of age and gender on the perception of lexical emotion. Appl. Neuropsychol. 6, 226–238.

Heilman, K.M., Gilmore, R.L., 1998. Cortical influences in emotion. J. Clin. Neurophysiol. 15, 409–423.

Herrero, J.V., Hillix, W.A., 1990. Hemispheric performance in detecting prosody: a competitive dichotic listening task. Percept. Motor Skills 71, 479–486.

Johnson, W.F., Emde, R.N., Scherer, K.R., Klinnert, M.D., 1986. Recognition of emotion from vocal cues. Arch. Gen. Psychiat. 43, 280–283.

Kirouac, G., Dore, F.Y., 1985. Accuracy of the judgment of facial expression of emotions as a function of sex and level of education. J. Nonverbal Behav. 9, 3–7.

McCluskey, K.W., Albas, D.C., 1981. Perception of the emotional content of speech by Canadian and Mexican children, adolescents, and adults. Internat. J. Psychol. 16, 119–132.

McNeely, H.E., Parlow, S.E., 2001. Complementarity of linguistic and prosodic processes in the intact brain. Brain Lang. 79, 473–481.

Mehler, J., Christophe, A., 1994. Language in the infant's mind. Philos. T. Roy. Soc. B 346, 13–20.

Mehler, J., Dupoux, E., Pallier, C., Dehaene-Lambertz, G., 1994. Cross-linguistic approaches to speech processing. Curr. Opin. Neurobiol. 4, 171–176.

Muente, T.F., Kohlmetz, C., Nager, W., Altenmueller, E., 2001. Superior auditory spatial tuning in conductors. Nature 409, 580.

Murray, I.R., Arnott, J.L., 1993. Toward the simulation of emotion in synthetic speech: a review of the literature on human vocal emotion. J. Acoust. Soc. Amer. 93, 1097–1108.

Pantev, C., Engelien, A., Candia, V., Elbert, T., 2001. Representational cortex in musicians: plastic alterations in response to musical practice. In: Zatorre, R.J., Peretz, I. (Eds.), The Biological Foundations of Music. New York Academy of Sciences, New York, pp. 300–314.

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

Pell, M.D., Baum, S.R., 1997. The ability to perceive and comprehend intonation in linguistic and affective contexts by brain-damaged adults. Brain Lang. 57, 80–99.

Ross, E.D., Thompson, R.D., Yenkosky, J., 1997. Lateralization of affective prosody in brain and the callosal integration of hemispheric language functions. Brain Lang. 56, 27–54.

Rotter, N.G., Rotter, G.S., 1988. Sex differences in the encoding and decoding of negative facial emotions. J. Nonverbal Behav. 12, 139–148.

Scherer, K.R., 2003. Vocal communication of emotion: a review of research paradigms. Speech Comm. 40, 227–256.

Scherer, K.R., 1995. Expression of emotion in voice and music. J. Voice 9, 235–248.

Starkstein, S.E., Federoff, J.P., Price, T.R., Leiguarda, R.C., Robinson, R.G., 1994. Neuropsychological and neuroradiologic correlates of emotional prosody comprehension. Neurology 44, 515–522.

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