

# Recognizing sarcasm without language

## A cross-linguistic study of English and Cantonese\*

Henry S. Cheang and Marc D. Pell

McGill University

The goal of the present research was to determine whether certain speaker intentions conveyed through prosody in an unfamiliar language can be accurately recognized. English and Cantonese utterances expressing sarcasm, sincerity, humorous irony, or neutrality through prosody were presented to English and Cantonese listeners unfamiliar with the other language. Listeners identified the communicative intent of utterances in both languages in a crossed design. Participants successfully identified sarcasm spoken in their native language but identified sarcasm at near-chance levels in the unfamiliar language. Both groups were relatively more successful at recognizing the other attitudes when listening to the unfamiliar language (in addition to the native language). Our data suggest that while sarcastic utterances in Cantonese and English share certain acoustic features, these cues are insufficient to recognize sarcasm between languages; rather, this ability depends on (native) language experience.

**Keywords:** Cantonese, communicative intentions, cross-linguistic, sarcasm, speech perception

### 1. Introduction

Sarcasm can be described as a negative critical attitude held by speakers that is expressed to mock and criticize other persons or events (Kreuz and Glucksberg 1989; Lee and Katz 1998). Like other forms of verbal irony, the expression of sarcasm in speech is characterized by indirect language meant to be interpreted non-literally by the listener; specific contexts, particular vocabulary, and a number of acoustic cues appear to contribute in a unique manner to sarcastic interpretation (Utsumi 2000). Although many studies to date have focused on the contextual mechanisms that drive sarcastic interpretations during human communication, a few have examined the role of prosody in this communicative context. For example, Anolli et al. (2002) have shown indications that the acoustic cues that convey sarcasm are

different from those that convey positive, humorous forms of verbal irony (henceforth referred to as “humorous irony” or “humor” for brevity, although note that humorous irony does not encompass all forms of humor). The goal of this study was to advance the literature by investigating whether listeners can use prosody to accurately recognize sarcasm and other commonly-expressed speaker attitudes in their native language and in a completely foreign language. Details of our rationale and approach are provided in what follows.

## 2. Acoustic-perceptual correlates of sarcasm in speech

There is a body of literature that links sarcasm to characteristic shifts in several acoustic parameters of spoken language. Various investigators have furnished evidence that speakers convey sarcasm through manipulations of fundamental frequency (F0), amplitude, speech rate, voice quality, and/or nasal resonance (e.g., Anolli et al. 2002; Rockwell 2000a, 2005, 2007; Schaffer 1982). However, the specific patterns associated with sarcastic utterances, such as whether speakers tend to raise (Anolli et al. 2002; Attardo, Eisterhold, Hay, and Poggi 2003) or lower (Rockwell 2000a; Schaffer 1982) their voice pitch/mean F0 to mark this attitude, are not always reported consistently. In addition, the available literature that is based on sarcastic expressions produced in English, French, Italian, and Japanese reveals both similarities and differences in the use of prosody among languages (Adachi 1996; Anolli et al. 2002; Laval and Bert-Erboul 2005; Rockwell 2000a). The most frequent points of commonality in sarcasm expression across languages involve speaker manipulation of F0/pitch and speech rate, whereas a more inconsistent pattern has been reported for other acoustic parameters, such as changes in voice quality (see Haiman 1998, for an overview).

Recently, we reported two complementary studies that describe the acoustic features associated with sarcastic utterances in English and in Cantonese, and which directly compare these features between the two languages (Cheang and Pell 2008, 2009). In each of our two language conditions, a comparable set of utterances (e.g., “*She is a healthy lady*”) was elicited from six native speakers of each language to convey four distinct attitudes: sarcasm, sincerity, positive/humorous irony, and neutrality. A number of acoustic measures were then taken from each recorded utterance (e.g., F0 mean and range, amplitude mean and range, speech rate, harmonics to noise ratio) for cross-linguistic comparison (Cheang and Pell 2009). In general, our data show that there are reliable, text-independent acoustic changes associated with the vocal expression of sarcasm in both English and Cantonese. For English, sarcastic utterances exhibited a significantly lower F0 mean, restricted F0 variability, heightened levels of noise (i.e., reduced harmonics

to noise ratio), and distinct resonance patterns from the other attitudes (Cheang and Pell 2008). For Cantonese, sarcasm was again acoustically distinct from the other attitudes but signalled with a significantly higher mean F0, restricted F0 variability, and restricted amplitude variability (Cheang and Pell 2009). Together, these studies support the argument that sarcasm in both English and Cantonese is marked by specific, albeit not identical, patterns of prosodic cues (Cheang and Pell 2008, 2009). The observation that acoustic profiles associated with sarcasm were not identical in English and Cantonese is perhaps not surprising, given that previous acoustic evaluations of sarcasm expressed in Japanese and French (among other languages) also report acoustic differences in this speech context (e.g., Adachi 1996; Laval and Bert-Eboul 2005).

Upon further examination of our data, mean F0 emerged as an acoustic parameter of particular importance for differentiating sarcasm from sincerity, humorous irony, and neutrality in the two languages, although this acoustic cue was employed differently by English versus Cantonese speakers: sarcasm in English displayed a lower F0 relative to the comparison attitudes, whereas sarcasm in Cantonese exhibited the highest F0 mean (Cheang and Pell 2009). Thus, global settings of mean F0 appear to be critical for highlighting the sarcastic intent of an utterance to listeners. Another key finding was that for both languages, the prosodic features associated with sarcastic expressions differentiated most clearly from those of sincere expressions; when the mean F0 of sarcastic expressions was lowered, the mean F0 of sincere expressions was raised and *vice versa* for the two languages (Cheang and Pell 2008, 2009). Finally, it is noteworthy that certain acoustic cues were exploited in the same manner by speakers of English and Cantonese to convey sarcasm: speakers of both languages tended to restrict F0 variation within sarcastic utterances and to express sarcasm at a slower rate than the other attitudes. Thus, there are notable similarities in how speakers of English and Cantonese communicate sarcasm (i.e., through reduced F0 variation, reduced speech rate), as well as pronounced cross-language differences in how certain, potentially critical parameters are employed in this context (i.e., concerning the directionality of changes in mean F0).

It is recognized that many acoustic differences observed in speech do not have a direct or proportional influence on the perception of intended meanings, including sarcasm (Rockwell 2007). As such, it is unclear how different conventions for marking sarcasm through prosody observed between languages (e.g., Cheang and Pell 2008, 2009) would affect the *recognition* of speaker intentions if presented in a cross-linguistic setting. It has even been suggested that verbal cues in sarcastic speech could transcend language boundaries (Haiman 1990) with a potential impact on sarcasm perception between languages. The question of whether sarcastic intentions can be accurately detected by listeners exposed to a foreign language

has not been tested to date (although cf. Bryant and Barrett 2007 for a related study which tested recognition of other speaker intentions in a cross-linguistic setting). It would be worthwhile to characterize the relationship between acoustic and perceptual measures of sarcasm in natural speech communication. As well, such research is of direct functional relevance to individuals in multi-cultural societies who increasingly interact with people from different linguistic backgrounds and must learn to recognize negative intentions in the absence of native language experience

### 3. On the cross-linguistic recognition of speaker attitudes

To our knowledge, no studies have looked at the cross-linguistic recognition of sarcasm/irony from prosody, although recent work has shown that some speaker intentions (marking attention and comfort) can be correctly inferred by adults listening to a foreign language (Bryant and Barrett 2007). A more established literature has investigated how basic emotions (e.g., joy, anger) are understood from prosody; if one looks at this work, there is consistent evidence that listeners exposed to a foreign language can accurately recognize a speaker's emotion strictly from prosodic attributes of speech at levels well exceeding chance (Albas, McCluskey, and Albas 1976; Beier and Zautra 1972; Kramer 1964; Pell, Monetta, Paulman, and Kotz 2009; Scherer et al. 2001; Thompson and Balkwill 2006; van Bezooijen, Otto, and Heenan 1983). Vocal emotion expressions may be recognized well across cultures because they are associated with common psycho-physiological responses to experiencing an emotion that impact on the vocal apparatus (Frick 1985; Scherer 1986); these reactions promote modal tendencies in the acoustic structure of vocal emotion expressions which are detectable across languages (Pell, Paulmann, Dara, Alasseri, and Kotz 2009; Scherer, Banse, and Walbott 2001). For example, exposure to unpleasant (e.g., disgust-inducing) stimuli is associated with heightened tension in the orofacial region (among other behaviors) that evoke spitting or regurgitation; these gestures contribute to predictable changes in resonance and voice quality when a speaker expresses disgust while speaking (Scherer 1986).

Although sarcasm assumes a more interpersonal function in communication and is not dependent of basic emotional processes, it remains possible that the inherently negative attitude expressed in sarcasm enacts physiological processes similar to those experienced when one encounters certain negative stimuli (Fongy 1971; Rockwell 2000a, 2005). Alternately (or concurrently), sarcastic messages could somehow encode information that bears a resemblance to (but is by no means identical to) certain "universal" emotion features (Haiman 1990, 1998).

If true, it is possible that listeners exposed to a foreign language could infer sarcastic intent when exposed to these more basic emotive features (in addition to the possibility that there is a distinct “ironic tone of voice” that is similar across languages). However, even in the cross-linguistic literature on emotion processing, it should be underlined that adult listeners typically demonstrate an “in-group advantage” for recognizing emotions produced by persons who share the same linguistic and cultural background (see Elfenbein and Ambady 2002 for a review). These latter findings argue that despite modal tendencies in how emotions are expressed through prosody, social conventions continue to play an important role in how meanings are inferred from prosody within and across language groups. One might expect that social conventions would play an even stronger role in the cross-linguistic processing of speaker attitudes and intentions such as sarcasm, especially since no consistent acoustic profile has yet been associated with sarcastic speech across languages. Unfortunately, there is little research to inform these predictions to date.

#### 4. The present study

Our present goal was to test whether speaker attitudes such as sarcasm, which are commonly expressed in most cultures, can be understood from their vocal expression in a foreign language. This aim arose in light of the fact that previous work, though few in number, have suggested that prosodic cues mark sarcasm differently across languages (cf. Adachi 1996; Anolli et al. 2002; Cheang and Pell 2008, 2009). This is a significant point, given the inherently negative role that sarcasm plays in communication (i.e., a mocking form of criticism). In particular, results from our previous studies of sarcasm have indicated a profile of sarcastic prosody in one language that is quite comparable to the profile of sincere prosody in another; such a pattern implies perceptual confusability of these two clearly opposing attitudes across distinct languages (Cheang and Pell 2008, 2009). Mistaking sincerity for sarcasm across interlocutors who speak different languages could have important social consequences; whether this is a genuine tendency therefore merits consideration.

Thus, Cantonese and English utterances conveying sarcasm, sincerity, humorous irony, and neutrality that were found to be acoustically distinct from one another in our previous acoustic studies (Cheang and Pell 2008, 2009) were presented to native listeners of both Cantonese and English in a cross-linguistic perceptual study. In light of our data which show that the Cantonese and English exemplars of sarcasm exhibit important acoustic differences, especially in the directionality of pitch register adopted in this context (Cheang and Pell 2009), we anticipated

that each listener group would have significantly more difficulty to recognize sarcastic intent from vocal cues present in the foreign versus native language due to the salience of pitch/F0 cues. In addition, given that sarcasm and sincerity appear to be strongly contrasted by Cantonese and English speakers using mean F0 but in the *opposite* direction (Cheang and Pell 2009), we speculated that listeners might confuse these particular intentions if they base their responses strongly on global F0 settings appropriate to their native language. The extent to which other acoustic parameters which are sometimes shared by sarcastic utterances in both languages (e.g., reduced F0 variation, reduced speech rate) would offset language-related differences in mean F0 to promote accurate cross-linguistic recognition of sarcasm could not be predicted with any certainty. As well, no firm predictions could be made about the ability to recognize humorous irony in a foreign language, although there is some evidence that neutral prosody is distinctive and leads to reliable cross-linguistic recognition in many instances (e.g., Pell, Monetta et al. 2009; Pell, Paulmann et al. 2009).

## 5. Method

### 5.1 Participants

We recruited 20 native English speakers (mean age in years: 22.6, SD: 3.5; mean years of education: 16.4, SD: 2.0) and 20 native Cantonese speakers (mean age in years: 34.7, SD: 5.3; mean years of education: 16.5, SD: 2.3) to participate as listeners. To be included in the study, listeners could not have any functional ability or protracted exposure to the non-native language as determined by an initial screening interview (which was always carried out in the participant's native language). All English participants were native speakers of Canadian English from Montreal and southern Ontario, and were undergraduate students attending McGill University. All Cantonese participants were born, raised, and educated either in the city of Hong Kong or Guangzhou (i.e., Cantonese environments) and each was a recent immigrant to the province of Quebec (Canada). All Cantonese participants continued to carry out their daily activities predominantly or exclusively in the Cantonese language.

### 5.2 Materials

The stimuli were a subset of recorded utterances taken from our previous studies (see Cheang and Pell 2008, 2009 for complete details regarding stimulus rationale and construction). Stimulus elicitation, recording, and perceptual validation

procedures were highly comparable in each of the two language conditions and are only summarized briefly here.

a. *Stimulus elicitation.* For each language, six young adults (three male, three female) were recruited as native speakers to enact each of the four target attitudes (sarcasm, sincerity, humorous irony, and neutrality) in their respective native language. The speakers produced short target sentences as part of a scripted dialogue; these sentences were semantically and syntactically comparable in the two languages and the text of each utterance allowed the speakers to produce the same item to express each of the four attitudes on separate occasions during the recording session. The text of the tokens consisted of the following English sentences and their Cantonese analogues: “I suppose; it’s a respectful gesture/ 係啱,呢個係個好客氣嘅表示”; “Is that so; she is a healthy lady./ 係咩;佢係個好健康嘅女人”; “Oh boy; he is a superior chef/ 嘩哎;佢係個好鬼叻嘅廚師”; “Yeah, right; what a spectacular result/ 係囉; 呢個係個犀利嘅結果”. A pilot reading study involving native speakers of the respective target language was run to establish that the text of each utterance did not strongly bias one of the target attitudes (Cheang and Pell 2008, 2009). Each speaker produced 96 recorded utterances. Recordings were conducted in a sound-attenuated booth using a high quality head-mounted mono microphone positioned approximately one inch from the speaker’s mouth (sampling rate of recordings: 44.1 kHz, 16 bit, mono).

b. *Stimulus validation and selection.* For the purpose of our acoustic studies (Cheang and Pell 2008, 2009), a separate group of English and Cantonese listeners were recruited from the same populations as the speakers to verify the intended attitudes expressed in the recordings (prior to submitting the tokens to acoustic analyses). None of these participants was the same as those who participated in the current study. In each language condition, 16 native English or Cantonese speakers were presented all of the items recorded *in the same language* and were required to identify the attitude conveyed by each utterance from among the four possible alternatives (25% recognition represents chance performance). This allowed us to estimate how accurately the target attitude was encoded by each recorded utterance.

These perceptual data were used as a basis from which to select utterances that were recognized as the target attitude. To keep the task manageable for participants, only 15% of the best validated utterances were selected as stimuli in the present experiment. These tokens were recognized as conveying a given attitude by a minimum of 57% of the native listener group (i.e., more than two times chance). Note that the items initially constructed for acoustic analysis in each language varied in linguistic structure and syllable length (i.e., utterances were two, seven, or eleven syllables in length, Cheang and Pell 2008, 2009). In the present experiment, in

order to provide the participants increased exposure to acoustic information upon which to base their recognition, only the 11-syllable tokens that met or exceeded the recognition criteria were entered as stimuli for cross-linguistic recognition.

In total, 79 English utterances (20 exemplars conveying sarcasm, sincerity, and neutrality and 19 exemplars of humorous irony) and 77 Cantonese utterances (20 exemplars conveying sarcasm, sincerity, and humorous irony and 17 exemplars of neutrality) served as the experimental stimuli. As these stimuli represent the best exemplars of utterances conveying each attitude described in our previous work (Cheang and Pell 2008, 2009), acoustic features of the selected items mirrored the major patterns reported in our earlier studies. For example, sarcastic utterances spoken in Cantonese were marked by higher mean F0 values than corresponding sincere, humorous, or neutral utterances, whereas sarcastic utterances in English displayed lower mean F0 values than the other attitudes; in each language, sincere utterances demonstrated the opposite setting in mean F0 making them most distinct from sarcasm for this acoustic parameter (see Cheang and Pell 2009 for complete details).

### 5.3 Experimental tasks/procedure

The English and Cantonese utterances were blocked for presentation in two separate tasks according to the respective language condition. Each of the 40 participants (20 English-speaking, 20 Cantonese-speaking) completed both the English and the Cantonese task during a single testing session. The order in which the two language tasks were presented varied evenly within each participant group and the sequence of individual trials was always randomized within each task. A total of 156 experimental trials (79 English, 77 Cantonese stimuli) were judged by each listener. The experiment was presented by a computer using Superlab 2.0 presentation software (Cedrus, USA) which also recorded the participants' responses.

Testing was conducted on an individual basis at McGill University or in a quiet room in the participant's home. In all cases, communication between the examiner and participants was carried out entirely in the native language of the participant. Participants were informed that they would be listening to individual utterances, spoken in either English or Cantonese, and that they should judge the attitude of the speaker in each case from four alternatives: sarcasm, sincerity, humor, and neutral. Listeners were always instructed to attend to *how* the sentences were spoken, since in half of the cases they would not understand the language. After listening to each sentence, written labels appeared on the computer screen (in the native language) and the participant used a mouse click response to indicate their judgement. Before beginning the experiment, definitions and short descriptions of each attitude and the situations under which the attitudes might

be produced were given. Following these examples and the administration of instructions, listeners then completed two blocks of practice trials which were not included in the experiment to get accustomed to the experimental procedure and the sound of the stimuli. The experiment began when all questions regarding the procedure had been addressed. Each participant was paid \$20 CDN after completing both tasks.

#### 5.4 Statistical procedure

The dependent variable of interest was response accuracy. Data for each attitude (sarcasm, sincerity, humorous irony, and neutrality) were examined in two ways. First, responses to stimuli of each attitude from both listener groups were subjected to separate single-sample t-tests; these analyses were conducted to determine whether listener responses for each attitude category differed significantly from chance (i.e., chance = 0.25). Second, the data for each attitude were then submitted to separate analyses of variance (ANOVA) with a fixed factor of LANGUAGE (Cantonese, English) and a repeated factor of LISTENER GROUP (Cantonese, English). We conducted separate ANOVAs on each attitude in an attempt to focus our findings on identification differences across listener groups per attitude, as this was the comparison of greatest theoretical interest. All significant main and interactive effects were elaborated using Tukey's HSD criteria ( $\alpha = 0.05$ ). Main effects subsumed by higher-order interactions are reported but not described.

## 6. Results

The ability of English and Cantonese listeners to correctly identify each of the four target attitudes when spoken in English and Cantonese is summarized in Table 1, which also demonstrates patterns of confusion among the four response categories.

### 6.1 Response patterns

The results of the series of single-sample t-tests conducted on proportions of responses as a function of attitude type revealed that listeners in both groups identified the attitude tokens spoken in both their native and non-native languages significantly above chance levels in the majority of cases ( $p < 0.0001$ ); in the native language conditions, these findings attest to the construct validity of the stimulus materials. Identification was found to be at chance levels in only three conditions: English listeners identified Cantonese tokens of humor and sarcasm at chance

**Table 1.** Mean recognition (%) of sarcasm, sincerity, humor, and neutrality expressed in Cantonese and English by native listeners of each language (correct target recognition is indicated in bold).

Listener Group	Language of Expression									
	Cantonese					English				
	Attitude	Sarcasm	Sincerity	Humor	Neutrality	Attitude	Sarcasm	Sincerity	Humor	Neutrality
Cantonese	Sarcasm	46	22	11	21	Sarcasm	27	41	16	16
	Sincerity	10	<b>63</b>	16	12	Sincerity	28	<b>49</b>	11	12
	Humor	29	25	<b>40</b>	6	Humor	33	11	<b>52</b>	3
	Neutrality	7	15	4	<b>74</b>	Neutrality	5	32	0	<b>62</b>
English	Sarcasm	<b>24</b>	45	14	17	Sarcasm	<b>85</b>	8	3	4
	Sincerity	21	<b>50</b>	16	14	Sincerity	2	<b>91</b>	4	4
	Humor	35	24	<b>29</b>	12	Humor	33	13	<b>53</b>	1
	Neutrality	14	20	4	<b>62</b>	Neutrality	6	8	0	<b>87</b>

levels ( $p=0.12$ ,  $p=0.44$ , respectively), whereas Cantonese listeners identified English tokens of sarcasm at chance level ( $p=0.54$ ). An elaboration of the response patterns follows.

In general, it can be seen that English listeners were quite successful at identifying sarcasm (85% correct), sincerity (91%), and neutrality (87%) expressed in English, although recognition of humor in this condition was less precise (53%). One-third of the humorous expressions in English were identified as sarcasm by English listeners, indicating a degree of overlap between these two categories in English. The identification of attitudes spoken in Cantonese by English listeners was much less exact overall (ranging from 24% — 62%) and was notably poor for sarcasm and humor which were, as outlined previously, identified at chance accuracy level (24% and 29%, respectively). Interestingly, nearly one-half (45%) of the sarcastic expressions produced in Cantonese were identified as “sincere” by English listeners. Sentences conveying humor in Cantonese were more frequently categorized as sarcasm (35% of all responses). English listeners were most accurate in recognizing neutrality in Cantonese (62% correct).

For the Cantonese listeners, accuracy in the *native* language tended to be lower overall when compared to the English listeners, although recognition of each attitude in Cantonese was still reliable: sarcasm (46%), sincerity (63%), neutrality (74%), and humor (40%). Of particular note here, Cantonese listeners often mistakenly identified sarcastic utterances as being neutral or sincere (21% and 22% of responses to sarcasm, respectively). Humor expressions in Cantonese were also highly confusable for Cantonese listeners, being identified as sarcasm (29%) or sincerity (25%). When listening to English speech, Cantonese listeners were more like the English listeners in their judgment of attitudes portrayed in the non-native language. Correct attribution of English sarcasm by Cantonese listeners was at chance levels (27% correct) and a plurality of responses in this condition marked the utterances as expressing sincerity (41% of responses to English sarcasm). While responses to English sincerity exemplars by Cantonese listeners were fairly accurate (49% correct), many responses (28%) erroneously indicated sarcasm as the intent. Interestingly, Cantonese listeners were *more* accurate in identifying humor spoken in English than in Cantonese (52% vs. 40% correct). The most accurate cross-linguistic identification of attitudes for the Cantonese listeners could be seen for neutrality expressed in English (62% correct).

## 6.2 Cross-linguistic analysis of accuracy data for each attitude

For sarcasm, the 2 x 2 ANOVA yielded significant main effects of LISTENER GROUP ( $F(1, 38) = 37.18$ ,  $p < .001$ ) and LANGUAGE ( $F(1, 38) = 53.97$ ,  $p < .0001$ ) and a significant interaction of these two factors ( $F(1, 38) = 189.88$ ,  $p < .0001$ ). The

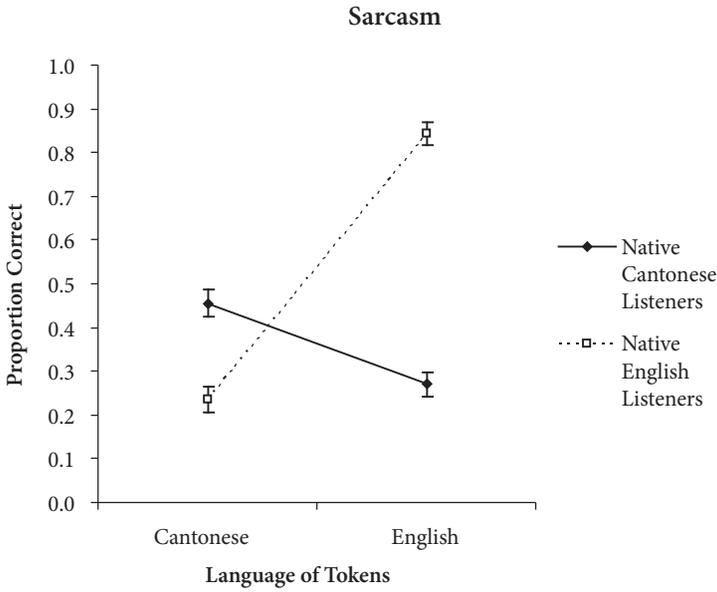


Figure 1a.

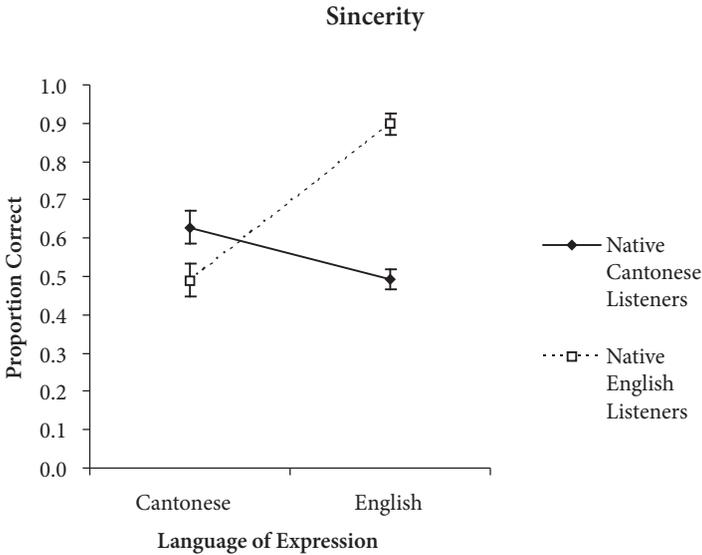


Figure 1b.

interaction was explained by the fact that the two listener groups were always significantly more accurate at identifying sarcastic sentences spoken in their native language than in a foreign language. Cantonese listeners recognized the Cantonese exemplars of sarcasm significantly better than English listeners, and English

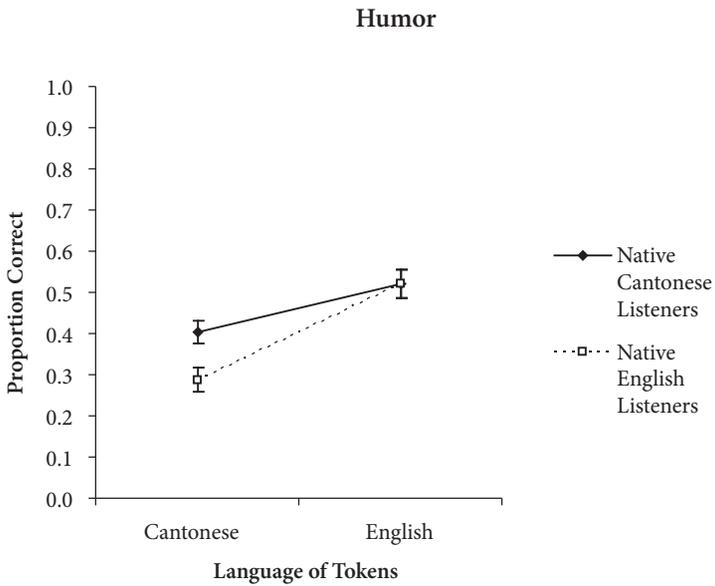


Figure 1c.

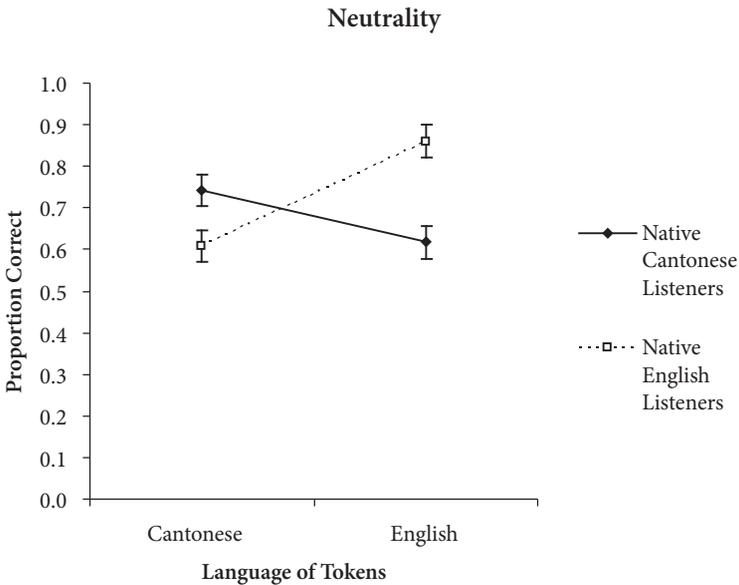


Figure 1d.

**Figure 1.** Effects of language and listener group on the recognition accuracy of (a) sarcasm (b) sincerity (c) humor and (d) neutrality. Vertical lines depict standard errors of the means.

listeners recognized the English exemplars of sarcasm significantly better than Cantonese listeners. These patterns are illustrated in Figure 1a.

For sincerity, the ANOVA yielded main effects of LISTENER GROUP ( $F(1,38)=11.66, p=.0002$ ) and LANGUAGE ( $F(1,38)=17.34, p=.0002$ ), as well as a significant interaction between the two factors ( $F(1,38)=68.74, p<.0001$ ). Posthoc tests established that participants in each listener group were always significantly better at identifying sincerity when spoken in their native language when compared to the foreign language. Expressions of sincerity in English were recognized more accurately by English than Cantonese listeners, whereas there were no significant group differences in the recognition of sincerity from Cantonese (although there was a trend for Cantonese listeners to be more accurate in this condition, see Figure 1b). For humor, the ANOVA produced a significant main effect of LANGUAGE ( $F(1,38)=38.88, p<.0001$ ) and a significant LANGUAGE by LISTENER GROUP interaction ( $F(1,38)=4.15, p=.0487$ ). Surprisingly, the interaction demonstrated that Cantonese listeners identified humor significantly less accurately when listening to Cantonese versus English tokens. As one might expect, English listeners were better at recognizing humor in English sentences versus Cantonese sentences (see Figure 1c).

Finally, for neutrality there was a significant main effect of LANGUAGE ( $F(1,38)=4.62, p=.0381$ ) and a significant interaction between LANGUAGE and LISTENER GROUP ( $F(1,38)=39.90, p<.0001$ ). The interaction was explained by the fact that like sincerity, neutral sentences spoken in English facilitated the performance of the English listeners rather than the Cantonese listeners. There were no significant differences between the listener groups in the recognition of neutrality from Cantonese utterances (Figure 1d).

## 7. Discussion

In this study we investigated how speaker attitudes are recognized in a listener's native language and in a foreign language, using a fully crossed research design involving English and Cantonese speakers and listeners. Our data imply that sarcastic intentions are processed in a distinct manner from sincerity, neutrality, and humorous irony; listeners in both groups recognized sarcasm expressed in the unfamiliar language at a level approximating chance for this task, whereas the cross-linguistic identification of the other attitudes was generally more successful. In most cases, recognition of speaker attitudes was facilitated when these expressions were produced in the native language of listeners, although there were some exceptions; these points are elaborated further below. On a practical level, our study highlights the fact that one must be sensitive to potential extra-linguistic

difficulties that may arise in the vocal channel while communicating with interlocutors from different cultural backgrounds.

Focusing on the identification of sarcasm, we found that both groups of listeners were sensitive to the sarcastic intent of utterances spoken in their native language but were highly error-prone for identifying sarcasm expressed in an unfamiliar language. Broadly speaking, these results argue that (native) experience with a language is essential for recognizing sarcastic intentions, as listeners had little ability to recognize this attitude from prosody in a foreign language. The recognition patterns observed could reflect underlying differences in the acoustic structure of sarcastic expressions produced in Cantonese and English; as noted earlier, we reported that English sarcasm is marked by significantly reduced mean F0, restricted F0 variability, and heightened levels of noise, whereas Cantonese sarcasm is marked by significantly greater mean F0, restricted F0 variability, and restricted amplitude variability (Cheang and Pell 2008, 2009). It is possible that listeners in each group employed knowledge of how sarcasm is expressed in their own language system as a model for identifying all instances of sarcasm, leading to predictable misattribution errors in the cross-linguistic context.

To elaborate on this idea, in our cross-linguistic conditions we noted that sarcasm was most frequently misidentified as “sincerity” by both English and Cantonese listeners. While our data show that listeners in both groups were relatively successful at identifying sincere utterances spoken in their native language (63 and 91% correct for Cantonese and English participants, respectively), both groups concurrently made similar identification errors for sincere and sarcastic sentences spoken in their non-native language. The observation that sarcasm was confused for sincerity in the non-native language condition by each listener group is predicted by differences in our acoustic data for the English and Cantonese tokens. When the acoustic profile of our four attitudes was compared, the prosodic characteristics of sincere expressions were most strongly distinguished from sarcasm in both English and Cantonese, and we reported that sentences projecting sarcastic and sincere attitudes in Cantonese exhibited mean F0 levels that were *opposite* from their respective English analogues (Cheang and Pell 2009). That is, whereas Cantonese speakers tended to adopt a relatively high F0 register (mean F0) to convey sarcasm and a low F0 register to convey sincerity, English speakers demonstrated the opposite tendency. Different expectations about how mean F0 is employed to express sincere versus sarcastic intentions held by Cantonese and English listeners could well explain the error patterns noted in our cross-linguistic conditions. At the same time, these data emphasize that mean F0 serves an important pragmatic function in both English and Cantonese and that speakers of both languages accord considerable weight to these cues when inferring speaker intentions and attitudes.

The misuse of particular acoustic features is unlikely to explain all of the findings vis-à-vis sarcasm and sincerity, since sincere utterances expressed in English and Cantonese could be recognized at two times chance level by listeners not fluent in these languages. It has been argued that interlocutors typically expect their communicative partners to express sincere sentiments (Bryant in press; Gibbs Jr. 2000) which is also in accordance with Grice's Cooperative Principle and communicative maxims (i.e., listeners and speakers cooperate to exchange information as effectively as possible by being efficient, truthful, relevant, and perspicuous; Grice 1975). Hence, it is possible that when listening to a foreign language and faced with utterances for which the acoustic cues marking speaker intentions are ambiguous, listeners simply erred towards the interpretation which is more frequent and expected, i.e. that the speaker is being sincere (Gibbs Jr. 2000). These issues will require further analysis to determine how speaker intentions are judged in a cross-linguistic setting.

The identification of humorous irony in our study merits some commentary, since recognition of this intention demonstrated a unique pattern in our four listening conditions. First, it should be underscored that humorous irony was the most poorly-recognized attitude for both the English and Cantonese listeners when judging their native language. One possible reason for this finding is that the construction of our stimuli made it difficult for listeners to fully appreciate humorous intent. Typically, humor appreciation requires the presence of a "play cue", or meta-message that indicates to listeners that the speaker intends to engage in non-literal communication, which is followed by violations of expectations generated by the speaker (Berger 1987; Berlyne 1972; Suls 1983). Since we presented only short utterances that were devoid of context, our stimuli could not promote expectations which would be violated to convey humor, and as such, the prosodic features of these utterances may have had little value as "play cues" in the native language.

Also of note was the finding that, irrespective of language or listener group, the humorous irony tokens were commonly identified as sarcasm in all of our listening conditions. By contrast, the opposite was not true, i.e., the present listeners did not tend to identify sarcasm as humor. This pattern may be accounted for by overlap in the functional role of sarcasm and humorous irony during communication; aside from its principal role of negative criticism, sarcasm can often be humorous, although not necessarily (Colston and O'Brien 2000). By contrast, humor seldom shares the critical intent of sarcastic messages (Anolli et al. 2002), which could explain why the tokens of humorous irony were often interpreted as sarcasm (and not *vice versa*). Interestingly, the fact that humor stimuli were usually interpreted as being non-literal, or conveying some form of irony (i.e., humor or sarcasm) across our different listening conditions, suggests that there were certain prosodic cues in these stimuli which broadly signalled ironic intentions that can

be recognized independent of language experience (Haiman 1990). This statement contrasts with our conclusion that recognizing sarcasm is specifically dependent on language experience; these issues will need to be explored in future research.

More generally, our results show that there were overall differences in how well each listener group identified attitudes in their native language. While the performance of each group was facilitated by the presentation of native-language stimuli, it is clear that the English listeners benefited more in this condition than the Cantonese listeners. Research suggests that there is considerable individual variability in the ability to successfully perceive affective and attitudinal states (Ivanko, Pexman, and Olinek 2004; Rockwell 2000b; Schaffer 1982; Toplak and Katz 2000) and this could have contributed in part to the group differences. However, our data imply that variation in the recognition of particular attitudes was relatively comparable for our two listener groups (review error bars in Figure 1).

Perhaps a better explanation for these group differences relates to basic differences between the two languages of interest here; specifically, Cantonese is known to employ additional means of signalling sarcasm in speech that are not used in English and were not present in our materials. In Cantonese, particles (i.e., utterance final non-word suffixes) are important for certain pragmatic and syntactic functions performed by acoustic cues in English. For example, the addition of /mae/ at the end of an utterance changes the mode from declarative to interrogative in Cantonese (among other dialects of Chinese, Chao 1968; Matthews and Yip 1994). In other contexts, critical attitudes such as sarcasm can be marked by specific particles in Cantonese (Chan 2002; Matthews and Yip 1994), although these particles were not present at the ends of the Cantonese utterances in our study. Although particles are not compulsory in most contexts in which they occur (as prosodic features do nonetheless perform prominent signalling functions, Fok 1974; Vance 1976), these may have been expected by Cantonese listeners in many instances and their absence may have influenced the data in some manner.

## 7.1 Future directions

This investigation represents an initial attempt to gauge whether speakers of two highly distinct languages, English and Cantonese, can use prosodic information to infer the meaning of commonly-expressed attitudes in their native and a foreign language; as such, future refinements of the current work can be envisioned. Constructing new and more diverse stimulus materials would be useful; for example, the number of speakers who produce the stimuli should be increased as a means of capturing the wide diversity of postulated cues for communicating sarcasm in speech (Haiman 1998). Studying further languages and additional attitudes of interest would also be constructive to build on our findings.

Finally, future research could benefit by presenting “speech filtered” utterances which convey specific intentions through prosody in spontaneously produced utterances which still control for the presence of semantic meaning. Such a manipulation would permit a fuller appreciation of the separate contributions of semantic cues and of prosodic cues in marking sarcasm. While there were firm indications that our listeners could employ only acoustic information to decode the attitudes spoken in their non-native language in certain conditions, there is some possibility that the text of the utterances (either semantically or through differentially frequent associations with sarcastic contexts, Bryant *in press*) played a role in the recognition of attitudes conveyed in their native language (despite our attempt to control for any influences of semantic information on target recognition in the native language conditions, Cheang and Pell 2008, 2009). Although the potential textual influence of the utterances is a significant consideration, it is still important to note that our earlier acoustic analyses of the tokens employed in the current study have found that certain prosodic cues of sarcasm (F0 in particular) mark this attitude independently of the semantic cues (Cheang and Pell 2009). Regardless, this is a limitation of the current study for which an experimental manipulation such as filtering speech would address.

## 7.2 Conclusion

By examining the relationship between acoustic and perceptual measures of sarcastic utterances, the present work represents a starting point for elucidating the factors that govern the expression of speaker attitudes and intentions, and how such factors differ across languages. We conclude that sarcasm has a unique expression in speech that consistently differentiates it from other attitudes, but that the conventions for signaling sarcasm in the voice vary in important ways across languages, yielding poor cross-linguistic recognition of this attitude. At the same time, it would seem that sarcasm plays an important role in many, if not all, languages (Haiman 1998) and that speakers/listeners routinely use prosody to distinguish this intention from related attitudes. However, they must first acquire knowledge about how specific changes in the acoustic code refer to signaling functions accepted by the language community. Thus, while the defining characteristics of sarcastic prosody may not be universal as is often claimed of emotional expressions in the voice (e.g., Pell et al. 2009), the goal of communicating the message of sarcasm using salient features of the voice is likely to be.

## Acknowledgment

\* This research was supported by a Canadian Institutes of Health Research — K.M. Hunter Doctoral Training Award and a Bridge Funding Award from the Center for Research on Language, Mind, and Brain (McGill University) to the first author, and operating funds from the Social Sciences and Humanities Research Council of Canada (to the second author).

## References

- Adachi, T. 1996. "Sarcasm in Japanese". *Studies in Language* 19: 1–36.
- Albas, D.C., McCluskey, K.W., and Albas, C.A. 1976. "Perception of the emotional content of speech: A comparison of two Canadian groups". *Journal of Cross-linguistic Psychology* 7: 481–490.
- Anolli, L., Ciceri, R., and Infantino, M.G. 2002. "From 'blame by praise' to 'praise by blame': Analysis of vocal patterns in ironic communication". *International Journal of Psychology* 37: 266–276.
- Attardo, S., Eisterhold, J., Hay, J., and Poggi, I. 2003. "Multimodal markers of irony and sarcasm". *HUMOR: International Journal of Humor Research* 16: 243–260.
- Beier, E.G. and Zautra, A.J. 1972. "Identification of vocal communication of emotions across cultures". *Journal of Consulting and Clinical Psychology* 39: 166.
- Berger, A.A. 1987. "Humor: An introduction". *American Behavioral Scientist* 30: 6–15.
- Berlyne, D.E. 1972. "Humor and its kin". In P.E. McGhee and J.H. Goldstein (eds), *The Psychology of Humor*. New York: Springer-Verlag, 43–60.
- van Bezooijen, R., Otto, S.A., and Heenan, T.A. 1983. "Recognition of vocal expressions of emotion: A three-nation study to identify universal characteristics". *Journal of Cross-linguistic Psychology* 14: 387–406.
- Bryant, G.A. In press. "Prosodic contrasts in ironic speech". *Discourse Processes*.
- Bryant, G.A. and Barrett, H.C. 2007. "Recognizing intentions in infant-directed speech: Evidence for universals". *Psychological Science* 18: 746–751.
- Bryant, G.A. and Fox Tree, J.E. 2005. "Is there an ironic tone of voice?". *Language and Speech* 48: 257–277.
- Chan, M. 2002. "Gender-related use of sentence-final particles in Cantonese". In M. Hellinger and H. Bussmann (eds), *Gender Across Languages: The Linguistic Representation of Women and Men*. Amsterdam: John Benjamins, 57–72.
- Chao, Y.R. 1968. *A Grammar of Spoken Chinese*. Los Angeles: University of California Press.
- Cheang, H.S. and Pell, M.D. 2008. "The sound of sarcasm". *Speech Communication* 50: 366–381.
- Cheang, H.S. and Pell, M.D. 2009. "Acoustic markers of sarcasm in Cantonese and English". *Journal of the Acoustical Society of America* 126(3): 1394–1405.
- Colston, H.L. and O'Brien, J. 2000. "Contrast of kind versus contrast of magnitude: The pragmatic accomplishments of irony and hyperbole". *Discourse Processes* 30: 179–199.
- Elfenbein, H.A. and Ambady, N. 2002. "On the universality and cultural specificity of emotion recognition: A meta-analysis". *Psychological Bulletin* 128: 203–235.
- Flege, J.E., Bohn, O.-S., and Jang, S. 1997. "Effects of experience on non-native speakers' production and perception of English vowels". *Journal of Phonetics* 25: 437–470.

- Fok, C.Y.-Y. 1974. *A Perceptual Study of Tones in Cantonese*. (vols. 18). Hong Kong: Centre of Asian Studies, University of Hong Kong.
- Fonagy, I. 1971. "Synthèse de l'ironie". *Phonetica* 23: 42–51.
- Frick, R.W. 1985. "Communicating emotion: The role of prosodic features". *Psychological Bulletin* 97(3): 412–429.
- Gibbs, R.W., Jr. 2000. "Irony in talk among friends". *Metaphor and Symbol* 15: 5–27.
- Grice, H.P. 1975. "Logic and conversation". In P. Cole and J.L. Morgan, *Speech Acts*, vol. 3: *Syntax and Semantics*. New York: Academic Press: 41–58.
- Haiman, J. 1990. "Sarcasm as theatre". *Cognitive Linguistics* 1–2: 181–205.
- Haiman, J. 1998. *Talk Is Cheap: Sarcasm, Alienation, and the Evolution of Language*. Oxford: Oxford University Press.
- Ivanko, S.L., Pexman, P.M., and Olineck, K.M. 2004. "How sarcastic are you? Individual differences and verbal irony". *Journal of Language and Social Psychology* 23: 244–271.
- Kramer, E. 1964. "Elimination of verbal cues in judgments of emotion from voice". *Journal of Abnormal and Social Psychology* 68: 390–396.
- Kreuz, R.J. and Glucksberg, S. 1989. "How to be sarcastic: The echoic reminder theory of verbal irony". *Journal of Experimental Psychology: General* 118: 374–386.
- Laval, V. and Bert-Erboul, A. 2005. "French-speaking children's understanding of sarcasm: The role of intonation and context". *Journal of Speech, Language, and Hearing Research* 48: 610–620.
- Lee, C.J. and Katz, A.N. 1998. "The differential role of ridicule in sarcasm and irony". *Metaphor and Symbol* 13: 1–15.
- Matthews, S. and Yip, V. 1994. *Cantonese: A Comprehensive Grammar*. London: Routledge.
- Pell, M.D., Monetta, L., Paulmann, S., and Kotz, S.A. 2009. "Recognizing emotions in a foreign language". *Journal of Nonverbal Behavior* 33: 107–120.
- Pell, M.D., Paulmann, S., Dara, C., Allasseri, A., and Kotz, S.A. 2009. "Factors in the recognition of vocally expressed emotions: A comparison of four languages". *Journal of Phonetics* 37: 417–435.
- Rockwell, P. 2000a. "Lower, slower, louder: Vocal cues of sarcasm". *Journal of Psycholinguistic Research* 29: 483–495.
- Rockwell, P. 2000b. "Actors', partners', and observers' perceptions of sarcasm". *Perceptual and Motor Skills* 91: 665–668.
- Rockwell, P. 2005. "Sarcasm on television talk shows: Determining speaker intent through verbal and nonverbal cues". In A. Clark (ed), *Psychology of Moods*. New York: Nova Science Publishers, 109–140.
- Rockwell, P. 2007. "Vocal features of conversational sarcasm: A comparison of methods". *Journal of Psycholinguistic Research* 36(5): 361–369.
- Schaffer, R. 1982. "Are there consistent vocal clues for irony?". In C.S. Masek, R.A. Hendrick, and M.F. Miller (eds), *Parasession on Language and Behavior*. Chicago: Chicago Linguistic Society, 204–210.
- Scherer, K.R. 1986. "Vocal affect expression: A review and a model for future research". *Psychological Bulletin* 99: 143–165.
- Scherer, K.R., Banse, R., and Wallbott, H.G. 2001. "Emotion inferences from vocal expression correlate across languages and cultures". *Journal of Cross-linguistic Psychology* 32: 76–92.
- Suls, J. 1983. "Cognitive processes in humor appreciation". In P.E. McGhee and J.H. Goldstein (eds), *Handbook of Humor Research*. New York: Springer Verlag, 39–57.
- Superlab (Version 4.0) [Computer software] 2007. San Pedro, CA: Cedrus.

- Thompson, W.F. and Balkwill, L.L. 2006. "Decoding speech prosody in five languages". *Semiotica* 158: 407–424.
- Toplak, M. and Katz, A. 2000. "On the uses of sarcastic irony". *Journal of Pragmatics* 32: 1467–1488.
- Utsumi, A. 2000. "Verbal irony as implicit display of ironic environment: Distinguishing ironic utterances from nonirony". *Journal of Pragmatics* 32: 1777–1806.
- Vance, T.J. 1976. "An experimental investigation of tone and intonation in Cantonese". *Phonetica* 33: 368–392.

### *Authors' addresses*

Henry S. Cheang  
 School of Communication Sciences and  
 Disorders  
 McGill University, 1266 Pine Avenue West  
 Montréal, Quebec H3G 1A8  
 Canada  
 henry.cheang@mail.mcgill.ca

(Current Affiliation):  
 Laboratoires CITÉ  
 Université de Montréal  
 Pavillon Roger-Gaudry (V-13-1)  
 2900, boul. Édouard-Montpetit  
 Montréal, Québec, H3T 1J4  
 Canada

Marc D. Pell  
 School of Communication Sciences and Disorders  
 McGill University, 1266 Pine Avenue West  
 Montréal, Quebec, H3G 1A8  
 Canada

marc.pell@mcgill.ca

### *About the authors*

**Henry S. Cheang** has conducted acoustic and behavioral research into the comprehension and production of language, emotional states, and attitudinal expression across typical adults and clinical populations (including persons presenting with right hemisphere stroke or Parkinson's disease). At present, he is a postdoctoral fellow in the Department of Psychiatry at the University of Montréal where he currently employs electroencephalography to study visual perceptual/attentional processes. He is currently affiliated with the CITÉ Laboratories at the University of Montréal.

**Marc D. Pell** has a broad interest in how adults communicate their emotions and intentions in speech, and how these abilities are affected by acquired brain disease (e.g., stroke, Parkinson's disease). Much of his research focuses on how speech *prosody*, or a speaker's tone of voice, is used in communication; current studies use methods from social psychology, cognitive neuropsychology, and neuroimaging to explore related issues. He is an Associate Professor and holds a research chair at McGill University.

Copyright of Pragmatics & Cognition is the property of John Benjamins Publishing Co. and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.