Acoustic markers of sarcasm in Cantonese and English

Henry S. Cheang[a] and Marc D. Pell
School of Communication Sciences and Disorders, McGill University, 1266 Pine Avenue West, Montreal, Quebec H3G 1A8, Canada

(Received 19 June 2008; revised 16 June 2009; accepted 18 June 2009)

The goal of this study was to identify acoustic parameters associated with the expression of sarcasm by Cantonese speakers, and to compare the observed features to similar data on English [Cheang, H. S. and Pell, M. D. (2008). Speech Commun. **50**, 366–381]. Six native Cantonese speakers produced utterances to express sarcasm, humorous irony, sincerity, and neutrality. Each utterance was analyzed to determine the mean fundamental frequency (F0), F0-range, mean amplitude, amplitude-range, speech rate, and harmonics-to-noise ratio (HNR) (to probe voice quality changes). Results showed that sarcastic utterances in Cantonese were produced with an elevated mean F0, and reductions in amplitude- and F0-range, which differentiated them most from sincere utterances. Sarcasm was also spoken with a slower speech rate and a higher HNR (i.e., less vocal noise) than the other attitudes in certain linguistic contexts. Direct Cantonese-English comparisons revealed one major distinction in the acoustic pattern for communicating sarcasm across the two languages: Cantonese speakers *raised* mean F0 to mark sarcasm, whereas English speakers *lowered* mean F0 in this context. These findings emphasize that prosody is instrumental for marking non-literal intentions in speech such as sarcasm in Cantonese as well as in other languages. However, the specific acoustic conventions for communicating sarcasm seem to vary among languages.
 [DOI: 10.1121/1.3177275]

PACS number(s): 43.70.Kv, 43.70.Fq, 43.70.Mn [AL] Pages: 1394–1405

I. INTRODUCTION

Speech prosody (i.e., intonation and stress patterns) conveys many types of information to listeners, including whether or not a speaker intends to be ironic or sarcastic (the latter being a subtype of verbal irony). Verbal irony occurs when the intended meaning of statements is opposite to, or different from, the literal sense of the words used (see Gibbs, 2000 for a detailed description). The unique feature of sarcasm as a form of verbal irony is that it is chiefly used to express *negative* critical attitudes. Several research sources have highlighted the importance of prosody as a cue for detecting sarcasm; for example, adult listeners have been found to identify sarcastic intent in content-filtered utterances (Bryant and Fox Tree, 2005; Rockwell, 2000a). It has also been shown that young children can recognize the intonational markers of sarcasm, and this ability is developmentally distinct from the ability to recognize sarcasm through semantic or contextual cues of speech (Ackerman, 1983, 1986; Capelli *et al.*, 1990; Laval and Bert-Erboul, 2005; Winner and Leekman, 1991).

Based on acoustic studies, it is likely that sarcasm is encoded in speech through various global manipulations in acoustic parameters such as fundamental frequency (F0) and F0 variability, amplitude and amplitude variability, speech rate, voice quality, and resonance (Attardo *et al.*, 2003; Bryant and Fox Tree, 2005; Cutler, 1974, 1976; Haiman, 1998; Mueke, 1969, 1978; Myers-Roy, 1976; Rockwell, 2007, 2000a, 2005; Schaffer, 1982). However, owing to methodological differences across studies, the available data are marked by uncertainty, and the relative importance of particular acoustic parameters for signaling sarcasm and their directionality cannot be fully determined. One limitation in literature on the expression of various attitudinal states is that some studies did not control for the possibility that speakers employ semantic cues (i.e., words or phrases; henceforth, "keyphrases") independently of, or in conjunction with, prosodic cues to express such information (see Scherer *et al.*, 1984). With particular respect to sarcasm, there is evidence that some keyphrases are used so frequently by speakers to convey sarcasm that these keyphrases become semantic markers of sarcasm regardless of overlaid prosodic features (i.e., "enantiosemantic," Haiman, 1998, p. 39). Another possible shortcoming in literature is that early acoustic descriptions of sarcastic prosody may be conflated with the acoustic markers of other forms of verbal irony; for example, Anolli *et al.* (2002) found significant acoustic differences between sentences, which were meant to convey positive verbal irony (i.e., a playful, humorous attitude) and negative verbal irony (i.e., sarcasm). However, this distinction is often not controlled in literature, promoting some uncertainty about the true prosodic markers of sarcasm.

Recently, Cheang and Pell (2008) reported a detailed acoustic investigation of sarcasm expressed in English. In that study, six native English speakers produced a common set of utterances to convey sarcasm, positive humorous irony, sincerity, or neutrality. Some of the sentences included keyphrases that previous authors claimed to be enantiosemantic of sarcasm in English (e.g., "I suppose."), whereas other sen-

[a]Author to whom correspondence should be addressed. Present address: Laboratoire de Recherche en Neurosciences et Électrophysiologie Cognitive, Service de Recherche, Hôpital Rivière-des-Prairies, 7070 Boulevard Perras, Montréal, Québec H1E 1A4, Canada. Electronic mail: henry.cheang@mail.mcgill.ca

tences could only be understood as sarcastic from the speaker's prosody. Following acoustic analyses of the recordings, it was found that sarcastic utterances were associated with a significantly lower mean F0 than utterances, which conveyed humorous irony or sincerity. Moreover, sarcasm was characterized by reductions in F0 variation (standard deviation) and in the harmonics-to-noise ratio (HNR) (a measure of voice quality) when compared to sincerity. Since these acoustic changes were observed for sentences with and without enantiosemantic keyphrases, they appear to be central vocal cues for encoding sarcasm in English, which are text-independent (Cheang and Pell, 2008). Speech rate and resonance changes also distinguished sarcasm from humor and sincerity but only in the context of specific sentence types. Cheang and Pell (2008) concluded that certain prosodic cues (e.g., reduced mean F0) may be central for expressing sarcasm in English, whereas speakers employ other acoustic cues to signal sarcasm in particular linguistic contexts.

However, these findings are restricted to how sarcasm is expressed in English and do not address potential cross-language differences in how prosody is used to convey intentions such as sarcasm. There is evidence that culture norms promote differences in how particular affective or attitudinal states are communicated across languages (Grabe *et al.*, 2003; McCluskey *et al.*, 1975). In fact, careful inspection of literature on sarcasm implies that there are certain cross-linguistic differences in the acoustic cues to sarcasm. French sarcastic utterances have been associated with higher F0, restricted F0-range, and a slower speech rate (Laval and Bert-Erboul, 2005). In contrast, sarcasm in Italian has been characterized as having a higher F0, increased F0-range, and greater amplitude (Anolli *et al.*, 2002). Despite differences in how acoustic cues appear to be used across languages (especially F0), it is possible that some of the similarities noted may reflect common physiological tendencies associated with sarcastic speech. While sarcasm cannot be considered an affective state, this intention is inherently marked by its negative valence; certain vocal and facial gestures associated with the expression of negative affect (such as a disgusted sneer) sometimes accompany the expression of sarcasm in speech (Cutler, 1974; Haiman, 1998). One can assume that any effects of these negative physiological responses on prosody would be relatively comparable across languages. For example, the disgusted sneer is associated with heightened tension in the orofacial region, which contributes to predictable changes in resonance and voice quality (Scherer, 1986).

Further studies of how sarcasm is conveyed in different languages, and especially in a language, which is highly distinct from English, would help to reconcile the observed differences in literature, and identify possible similarities in the acoustic expression of sarcasm across languages. Cantonese is a good language to study for this purpose because Cantonese and English have no common linguistic roots, and there are enormous cultural distance and communication style differences between speakers of these languages (Bond, 1991; Ho, 1986; Huang and Kok, 1999; Snow, 2004). Little research has been done to quantify the acoustic and linguistic markers of sarcasm in Cantonese; however, it appears that some idioms or syntactic forms predispose listeners to recognize sarcastic intentions in this language (Chan, 2001; Killingley, 1986; Matthews and Yip, 1994). There are hints that Cantonese speakers also use acoustic cues to signal sarcasm [Kwok and Luke, 1986, as cited by Bauer and Benedict (1997)], but no empirical evidence is available and it is difficult to dissociate potential prosodic markers from cues derived from semantics, connotation, or communication rules. A prosodic evaluation of Cantonese sarcasm should shed light on its acoustic specifications, which can also be usefully compared to known acoustic cues of sarcasm in other languages.

Following Cheang and Pell (2008), the major goal of this study was to evaluate whether Cantonese speakers utilize specific acoustic cues to express sarcasm when compared to other attitudes, and whether these acoustic features interact with specific phrase types (i.e., keyphrases) associated with sarcasm. A second aim was to compare these new acoustic findings for Cantonese with the data on English sarcastic speech (Cheang and Pell, 2008). To allow for these comparisons, a procedure highly similar to that of Cheang and Pell's (2008) investigation was adopted: Native (Cantonese) speakers were recruited to convey sarcasm, humor, sincerity, and neutrality in simple sentences using only prosodic cues. Following a perceptual validation procedure, involving a separate group of native Cantonese-speaking listeners, acoustic analyses of the sentences were conducted to determine whether the four attitudes were characterized by specific patterns of prosodic cues. The Cantonese sentences were then acoustically compared with the English sentences analyzed previously (Cheang and Pell, 2008). In a broad context, this research will contribute to a new understanding of how attitudes are conveyed extra-linguistically and across languages, and could represent an important step toward bridging potential cross-cultural misunderstandings in verbal communication.

As the context and approach of this study were relatively novel, firm predictions about the Cantonese speakers could not always be made, and cross-language comparisons of sarcastic speech should be considered exploratory in nature. Nonetheless, based on the literature cited above, it was hypothesized that sarcastic utterances in Cantonese would differ significantly from utterances conveying positive, humorous irony, as well as sincerity on measures of F0 [Kwok and Luke, 1986, as cited by Bauer and Benedict (1997)]. Also, since speaker manipulations of amplitude, speech rate, and voice quality are frequently cited as markers of sarcasm (Cutler, 1974, 1976; Haiman, 1998; Mueke, 1969, 1978; Myers-Roy, 1976; Rockwell, 2007, 2000a, 2005), it was anticipated that some of these features would also distinguish sarcasm from the other attitudes in Cantonese. Given the considerable linguistic and cultural distinctions between Cantonese and English, no strong predictions about the nature of Cantonese-English differences in sarcastic prosody could be made, although it was speculated that F0 would be used in some way to communicate sarcasm in both languages.

TABLE I. Text through which speakers articulated the four target attitudes (sarcasm, humor, sincerity, and neutrality). "A" items denote combined sentences, "B" items denote single sentences, and "C" items denote keyphrases. The text of the exemplars across the Cantonese and English samples were highly comparable. Biasing sentences for the Cantonese tokens are not presented, as they were close translations of the English materials. As there are presently no truly universal transcription schemes for Cantonese, the transcriptions of the text are based on those summarized in Matthews and Yip (1994), where words are transcribed with the phonemic tone identified following vowel identification. Note that there are six Cantonese vowel tones: high level (55), mid level (33), low level (22), high rising (35), low rising (23), and low falling (21).

Item	Attitude	Biasing sentence	Target utterance (Cantonese)	Target utterance (English)
1	Sarcasm	Don't you just love how your stupid mother-in-law always smirks and snorts loudly when you misspeak?	A: 係^{口}卦，呢個係個好客氣^{口}既表示。 hai22 gwa33, lei55 go33 hai22 go33 hou35 haak33 hei33 ge5 biu35 si22.	A: I suppose; it's a respectful gesture.
	Humor	Not everyday that you see a priest give the finger, is it?	B: 呢個係個好客氣^{口}既表示。 lei55 go33 hai22 go33 hou35 haak33 hei33 ge5 biu35 si22.	B: It's a respectful gesture.
	Sincerity	It was nice of your supervisor to send flowers.	C: 係^{口}卦。 hai22 gwa33.	C: I suppose.
2	Sarcasm	That horrid woman smokes a pack a day.	A: 係咩; 佢係個好健康^{口}既女人。 hai22 me55, keui5 hai22 go33 hou35 gin22 hong55 ge23 neui23 yan35.	A: Is that so; she is a healthy lady.
	Humor	Your friend Shelley can't even do a single pushup.	B: 佢係個好健康^{口}既女人。 Keui23 hai22 go33 hou35 gin22 hong55 ge23 neui23 yan35.	B: She is a healthy lady.
	Sincerity	She runs 10 miles everyday.	C: 係咩。 hai22 me55	C: Is that so?
3	Sarcasm	Our moronic boss gave us all food poisoning.	A: 嘩哎；佢係個好鬼叻^{口}既廚師。 wa55 aai55, keui23 hai22 go33 hou35 gwai35 lek55 ge23 cheui21 si55.	A: Oh boy; he is a superior chef.
	Humor	Your brother singed his eyebrows while making toast.	B: 佢係個好鬼叻^{口}既廚師。 Keui23 hai22 go33 hou35 gwai35 lek55 ge23 cheui21 si55.	B: He is a superior chef.
	Sincerity	Butch just won another cooking contest; this is his twentieth win in the last three years.	C: 嘩哎。 wa55 aai55.	C: Oh boy.
4	Sarcasm	The arrogant front-runner finished dead last.	A: 係囉； 呢個係個犀利^{口}既結果。 hai22 lo55, lei55 go33 hai22 go33 sai55 lei22 ge5 git33 gwo35.	A: Yeah, right; what a spectacular result.
	Humor	Fascinating how she lost the eating contest to someone half her size huh?	B: 呢個係個犀利^{口}既結果。 lei55 go33 hai22 go33 sai55 lei22 ge23 git33 gwo35.	B: What a spectacular result.
	Sincerity	He broke three records in that race!	C: 係囉。 hai22 lo55.	C: Yeah, right.

II. METHOD

A. Stimulus production: Encoders and materials

The "encoders" were six native Cantonese speakers who were living in Montreal, Canada (3 males, 3 females; mean age: 22.7 years, SD: 3.2 years; mean education: 17.0 years, SD: 2.0 years). All encoders were born, raised, and educated in Hong Kong or Guangzhou (i.e., cities where Cantonese is primarily spoken) and moved to Canada as young adults. They were late learners of additional languages (English and/or French) and used Cantonese exclusively when at home. As with the six English encoders studied by Cheang and Pell (2008), the present encoders had no formal training in acting.

A set of 96 Cantonese utterances was recorded from each encoder—24 utterances representing each of the 4 attitudes investigated by Cheang and Pell (2008): sarcasm, positive humorous irony, sincerity, and neutrality. Henceforth, positive humorous irony will be referred to as "humor" for brevity; while it is acknowledged that humor does not have to be ironic in nature, here "humor" will always refer to a positive form of irony associated with playful, humorous intent (Anolli *et al.*, 2002). Sincerity typically refers to a speaker's attempt to reinforce the literal meaning of their utterance using prosody or other cues. Neutrality, which marks instances where the speaker does not wish to convey obvious emotions or intentions through prosody, is known to have a distinct prosodic form in different languages (e.g., Pell *et al.*, 2009) and was selected as a baseline category for data interpretation. Given the secondary objective of evaluating whether Cantonese speakers modulate acoustic cues in conjunction with particular semantic cues, a subset of the items included Cantonese idioms ("keyphrases") associated with sarcastic messages. Overall, the 96 utterances consisted of an equal number of keyphrases, sentences, and "combined sentences," which were composed of the keyphrases and the sentences. An example of a combined sentence is "係咩; 佢係個好健康口既女人" (English meaning: "Is that so; she is a healthy lady"), for which "係咩" ("Is that so") is the keyphrase and "佢係個好健康口既女人"("She is a healthy lady") is the sentence. The text of the items ranged from 2 syllables for keyphrases to 9–11 syllables for combined sentences. Four such sets of utterance forms were created (see Table I).

In all cases, materials were devised to be as semanti-

cally, syntactically, and syllabically comparable as possible to the English materials constructed in Cheang and Pell, 2008. All items could be produced by the encoders to express different attitudes using identical text. Common words and idioms were used in the construction of all stimuli.[1] Also, care was taken to evenly distribute the six Cantonese phonemic vowel tones across the text of the tokens, except for the low falling tone, which occurred less frequently than the other tones (see transcriptions in Table I). This limitation is unlikely to be critical, as earlier work on Cantonese suggests that cues used to mark semantic information through phonemic tone shape are relatively independent of global cues, which mark extra-linguistic information in the sentence (Vance, 1976).

Each target utterance was produced in response to a biasing sentence, except in the neutral condition where encoders simply produced the utterance in isolation. For sarcasm, humor, and sincerity, the encoders produced each target utterance as if they were engaging in a scripted dialog. Biasing sentences provided a context that would facilitate production of the associated attitude; in the case of sarcasm, these sentences included insulting, cruel, or unfairly critical cues (e.g., "That horrid woman smokes a pack a day" to bias a sarcastic rendition of "She is a healthy lady"). Sentences biasing humor production included overt declaration of a friendly relation or playful cues (e.g., "Your friend can't even do a single push-up"). Biasing sentences used to help elicit sincerity did not contain information that suggested positive, negative, or other possible associations (e.g., "She runs 10 miles every day"). Moreover, encoders were presented with detailed explanations regarding the attitudes they were to reproduce prior to the recording of each set of attitudes, although specific acoustic attributes were intentionally not highlighted to them. For each attitude, identical biasing sentences were used for each item of the three phrase types. Each utterance was recorded twice non-sequentially from each encoder (review Table I).

As part of stimulus development, the text of the recording materials was presented to four native Cantonese speakers in a pilot reading study to judge the suitability of the sentence pairs in portraying situational contexts appropriate to each target attitude. Raters had to classify the pairs as being "unnatural," "somewhat natural," or "natural." Target utterances and biasing sentences were refined based on these ratings and presented to different raters. Refinement continued in this fashion until there was a minimum of 75% agreement across raters that sentence pairs reflected natural interactions.

B. Recording procedure

Interactions between the experimenter and participants were conducted exclusively in Cantonese. The entire set of neutral utterances was always elicited first from each encoder. All encoders were instructed to read aloud target sentences printed on cards in a neutral voice, devoid of affect. After recording the full set of neutral utterances, three separate sets of utterances (each conveying only one of the remaining attitudes) were recorded one at a time in random sequence across encoders. The recording of each attitude was blocked to help encoders to successfully adopt each expressive mode. Within each attitude set, target sentences were presented in a fixed random order.

Prior to producing sentences representing each attitude, encoders were provided definitions of each attitude and short, standardized descriptions of situations under which these attitudes are expressed (e.g., "people use sarcastic utterances to respond to insulting comments directed at them;" "people use humorous statements to be playful with friends"). Encoders were given no indication of which acoustic cues to employ during the recording procedure; rather, they were instructed to use these descriptions and the biasing sentences to facilitate their enactments of the target attitude. Sentence pairs consisting of a biasing sentence and an associate phrase type (i.e., keyphrase, sentence, or combined sentence) were presented to the encoder on printed cards. Encoders read the biasing sentence silently and then produced the target sentence aloud to communicate the target attitude. Encoders were given practice trials for each attitude prior to recording the experimental items. Encoders were neither coached nor given feedback regarding their renditions of the target attitudes, although they were allowed to repeat their productions if desired (the final exemplar was always retained for analysis). Recordings were conducted in a sound-attenuated booth, captured by a AKG C-420 head-mounted professional microphone onto a Sony TCD-D100 digital audio tape recorder (sampling rate: 44.1 kHz, 16 bits, mono). The digital recordings were transferred directly to a computer (downsampled to 24 kHz but unfiltered) for editing and acoustic analyses using PRAAT software (Boersma and Weenink, 2006). A total of 576 utterances were recorded (4 attitudes $\times$ 4 items $\times$ 3 phrase types $\times$ 2 repetitions $\times$ 6 encoders).

C. Perceptual validation study

In this study, attitudes were "posed" to control for the linguistic-semantic structure of the items submitted to acoustic analysis. Given this approach, it was necessary to first establish the representativeness and validity of each token, to ensure that acoustic measures referred to perceptually identifiable exemplars of each attitude. Therefore, a perceptual validation study was conducted prior to the acoustic analyses, involving a separate group of 16 native Cantonese "decoders" (8 males, 8 females; mean age: 24.9 years, SD: 6.0 years; mean education: 17.9 years, SD: 3.5 years). The decoders were recruited from the same population as the encoders (i.e., were born, raised, and educated in Hong Kong or Guangzhou and were living in Montreal). Each decoder was required to identify the attitude expressed by each utterance, based on their knowledge of how prosodic cues (and when applicable, enantiosemantic terms) are used to convey sarcasm, humor, sincerity, and neutrality. This procedure was designed to eliminate utterances, which did not reliably convey the intended meanings due to difficulties at the stage of *simulating* attitudes, while establishing the perceptual validity of utterances identified by a majority of decoders as representing one of the target attitudes.

TABLE II. Number of utterances retained from the perceptual validation study (with percent agreement across decoders in parentheses) for each attitude and phrase type condition.

	Attitude				
Phrase type	Sarcasm	Sincerity	Humor	Neutrality	Total
Keyphrase	58 (65%)	8 (57%)	26 (69%)	23 (57%)	115
Sentence	5 (60%)	89 (73%)	22 (62%)	33 (74%)	149
Combined sentence	22 (59%)	44 (60%)	25 (55%)	30 (69%)	121
Total	85	141	73	86	385

Note: Items were retained and/or reclassified based on a minimum 50% agreement about the intended attitude when judged by the 16 decoders.

For the validation experiment, all 576 utterances were presented in 19 blocks (i.e., approximately 30 utterances per block), and each decoder identified the intended attitude of the speaker in a forced choice decision task (sarcasm, humor, sincerity, and neutrality). Following previous methods (Cheang and Pell, 2008), utterances were retained as "valid" exemplars for acoustic analysis if the final recognition consensus of the 16 decoders was at least two times chance performance levels (i.e., 50% or greater, where 25% consensus represented chance levels). When a speaker's intended attitude was strongly recognized as a different attitude, the item was reclassified in favor of the perceptual ratings; this procedure helped to control for possible mismatches between perceived and intended meaning in utterances and to retain the maximum number of items for acoustic analysis. Following the perceptual validation study, 385 (67%) of the initial utterances were retained as valid exemplars of the four attitudes, as summarized in Table II.

D. Acoustic analyses

The choice of acoustic parameters was guided by the authors' previous findings on English sarcasm (Cheang and Pell, 2008), trends in sarcasm research (e.g., Anolli *et al.*, 2002; Attardo *et al.*, 2003; Cutler, 1974, 1976; Haiman, 1998; Rockwell, 2000a), and empirical observations of acoustic patterns that characterize Cantonese (Bauer and Benedict, 1997; Fok, 1974; Vance, 1976). All acoustic analyses were performed using PRAAT; for F0/pitch, these analyses were conducted automatically using an autocorrelation method, the results were smoothed, and the output was visually inspected and manually corrected in the event of "halving" and "doubling" errors in the data. The specific acoustic measures derived for each utterance were given as follows:

(1) mean F0—measured in hertz for each utterance as a whole;
(2) F0-range—computed by subtracting the minimum F0 value from the maximum F0 value of the full utterance, to estimate the degree of F0 variation;[2]
(3) mean amplitude—measured in decibel for each utterance as a whole;
(4) amplitude-range—computed by subtracting the minimum amplitude value from the maximum amplitude value of each exemplar, to estimate amplitude variation;
(5) speech rate—calculated in syllables/s by dividing the number of syllables by the total utterance duration; and
(6) HNR—computed as the ratio of the averaged periodic component of a sound signal to the corresponding averaged noise component, expressed in decibel (Yumoto *et al.*, 1982); HNR measures were taken from 50-ms stable central portions of vowels segmented from stressed syllables (most vowels in unstressed syllables were shorter than 50 ms).

E. Statistical procedure

Prior to statistical analysis, all acoustic measures pertaining to tokens produced by a single encoder (irrespective of attitude) were converted into *z*-scores to allow comparisons across items and encoders. All individual values of F0, amplitude, and speech rate were standardized separately per encoder in reference to his or her entire set of productions by dividing the difference between the averaged value of all exemplars from an individual data point by the standard deviation of all exemplars [e.g., ([mean F0(one exemplar of sarcasm) − mean F0(all exemplars)])/F0 SD(all exemplars)]. For values of HNR only, all HNR values taken from the stressed vowels of a given exemplar were first averaged prior to standardization as described above.

Following normalization, *z*-scores of the acoustic data from the Cantonese exemplars were subjected to separate analyses of variance (ANOVAs) involving the factors of attitude (sarcasm, humor, sincerity, and neutrality) and phrase type (keyphrase, sentence, and combined sentence), independently of each normalized acoustic measure. After characterizing which acoustic cues contributed to the expression of different attitudes in Cantonese, a second set of ANOVAs directly compared acoustic values of Cantonese versus English; these analyses considered the fixed variable of language (Cantonese and English) with repeated measures on attitude (sarcasm, humor, sincerity, and neutrality). Phrase type was omitted in the cross-language comparison to concentrate analyses on how speakers of the two different languages vary in their acoustic expression of sarcasm. In all cases, significant effects and interactions were explored *post hoc* using Tukey's honestly significant difference (HSD) method ($\alpha = 0.05$). Significant main effects that were subsumed by higher-order interactions are reported but not described in the text.

TABLE III. Mean normalized acoustic measures (and standard deviation) of Cantonese utterances expressing each of the four attitudes, divided by phrase type.

Phrase type	Attitude	Mean F0 (Hz)	F0-range (Hz)[a]	Mean amplitude (dB)	Amplitude-range (dB)[a]	Speech rate (syllables/s)	HNR[b] (dB)
Keyphrase	Sarcasm	0.64(1.40)	−0.36(1.17)	0.66(0.90)	−0.89(0.73)	−0.96(0.69)	0.61(1.01)
	Sincerity	−0.29(0.74)	−0.88(0.46)	0.50(0.99)	−1.23(0.6)	−0.79(0.87)	0.66(1.22)
	Humor	0.25(1.26)	−0.16(1.00)	0.49(0.84)	−1.33(0.70)	−1.27(0.62)	−0.93(1.20)
	Neutrality	−0.50(0.89)	−0.68(1.28)	−0.27(0.98)	−0.64(0.91)	−0.87(0.78)	0.27(1.62)
Sentence	Sarcasm	−0.03(0.84)	−0.12(0.43)	0.25(0.39)	−0.14(0.49)	0.04(0.45)	0.04(0.90)
	Sincerity	−0.01(0.72)	0.15(0.82)	0.33(0.71)	−0.07(0.70)	1.03(0.47)	−0.17(0.71)
	Humor	0.03(1.07)	0.07(1.08)	0.12(0.89)	−0.01(0.61)	0.96(0.51)	−0.81(1.59)
	Neutrality	−0.74(0.39)	−0.33(0.55)	−0.38(0.84)	0.47(0.58)	0.05(0.67)	−0.04(0.85)
Combined sentence	Sarcasm	0.12(0.69)	0.47(0.70)	−0.34(0.91)	0.92(0.61)	−0.04(0.45)	−0.06(0.63)
	Sincerity	0.15(0.69)	0.54(0.85)	−0.42(0.89)	0.78(0.64)	0.22(0.59)	−0.02(0.73)
	Humor	0.34(1.02)	0.58(1.21)	−0.41(0.90)	0.65(0.60)	0.16(0.67)	0.24(0.97)
	Neutrality	−0.74(0.29)	−0.13(0.60)	−1.13(1.00)	1.06(0.52)	−0.41(0.69)	0.01(0.53)

[a]Range=maximum−minimum.
[b]HNR=harmonics-to-noise ratio.

III. RESULTS

Table III summarizes the (normalized) acoustic features of Cantonese expressions of sarcasm, humor, sincerity, and neutrality, separately by phrase type.

A. Fundamental frequency: Mean and range

A 4×3 (attitude×phrase type) repeated-measures ANOVA performed on mean F0 yielded a significant main effect for attitude, $F(3,373)=14.91$, $p<0.0001$. *Post hoc* comparisons indicated that utterances conveying sarcasm were produced with a significantly higher mean F0 than sincerity overall. In addition, neutrality was spoken with a significantly lower mean F0 than all other attitudes. No significant interactions were found. Patterns of mean F0 are illustrated in Fig. 1.

Analysis of F0-range yielded main effects for attitude, $F(3,373)=4.66$, $p=0.003$, and phrase type, $F(2,373)=20.12$, $p<0.0001$. The F0-range of sarcastic utterances was significantly narrower than that of sincere tokens, and neutral utterances were produced with a significantly smaller F0-range than humor and sincerity. For the phrase type main effect, *post hoc* comparisons showed that combined sentences were produced with the widest F0-range overall, which was greater than sentences, followed by keyphrases, which exhibited a relatively narrow F0-range. No significant interactions were found.

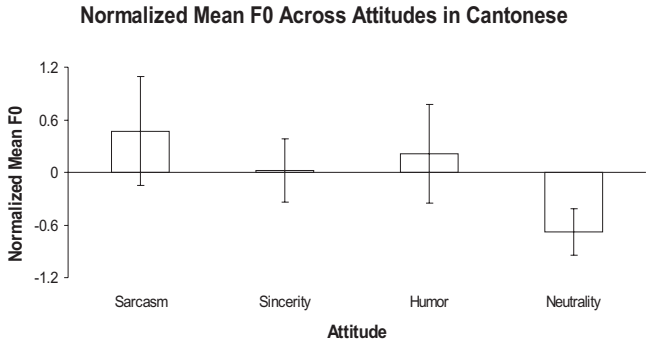

FIG. 1. Mean fundamental frequency (F0) of Cantonese utterances conveying sarcasm, sincerity, humor, and neutrality. Error bars represent standard deviations.

B. Amplitude: Mean and range

The 4×3 ANOVA on mean amplitude yielded a significant main effect of attitude, $F(3,373)=12.89$, $p<0.0001$. This effect was explained by the fact that neutrality was produced with a lower mean amplitude than the other three attitudes. There was also a significant main effect of phrase type, $F(2,373)=27.62$, $p<0.0001$, which revealed that keyphrases were spoken with the greatest amplitude, followed by sentences, which displayed greater amplitude than combined sentences. No significant interactions were found.

Analyses of amplitude-range yielded a significant main effect of attitude, $F(3,373)=9.61$, $p<0.0001$. Sarcasm and humor were produced with a significantly more restricted amplitude-range than sincerity, which in turn showed less amplitude variability than neutrality. A main effect of phrase type, $F(2,373)=171.62$, $p<0.0001$, could be explained by the fact that keyphrase exemplars displayed the most restricted amplitude-range and sentence exemplars showed less amplitude variation than combined sentence exemplars. No significant interactions were found.

C. Speech rate

Analysis of normalized speech rate yielded significant main effects for attitude, $F(3,373)=15.05$, $p<0.0001$, and phrase type, $F(2,373)=84.42$, $p<0.0001$, and a significant interaction of these factors, $F(6,373)=4.05$, $p=0.0006$. *Post hoc* tests on the interaction revealed that when producing sentences, sarcasm was expressed at a significantly slower speech rate than sincerity; also, neutrality was produced more slowly than humor and sincerity. For combined sentences, humor and sincerity were marked by a significantly faster speech rate than neutrality. There were no speech rate

TABLE IV. Mean normalized acoustic measures (and standard deviation) of Cantonese and English utterances expressing each of the four attitudes.

	Sarcasm		Sincerity		Humor		Neutrality	
Measure	Cantonese	English	Cantonese	English	Cantonese	English	Cantonese	English
Mean F0 (Hz)	0.47(1.25)	−0.45(0.72)	0.02(0.71)	0.42(1.05)	0.21(1.12)	0.34(1.17)	−0.68(0.55)	−0.49(0.48)
F0-range (Hz)	−0.13(1.09)	−0.05(1.02)	0.21(0.87)	0.25(0.91)	0.16(1.13)	0.09(0.96)	−0.36(0.84)	−0.44(0.99)
Mean amplitude (dB)	0.38(0.98)	0.14(1.24)	0.11(0.86)	0.22(0.72)	0.07(0.95)	0.33(0.72)	−0.61(1.00)	−0.65(0.89)
Amplitude-range (dB)	−0.38(1.05)	0.00(1.27)	0.13(0.84)	−0.10(0.90)	−0.25(1.06)	0.10(1.05)	0.38(0.94)	0.09(0.82)
Speech rate (syllables/s)	−0.67(0.76)	−0.80(0.99)	0.67(0.74)	0.46(0.84)	−0.11(1.11)	−0.03(1.00)	−0.36(0.79)	−0.02(0.72)
HNR (dB)	0.41(0.96)	−0.09(1.00)	−0.08(0.77)	−0.08(0.97)	−0.48(1.36)	0.18(0.96)	0.06(1.03)	0.10(1.06)

distinctions across the attitudes when speakers produced keyphrases.

D. HNR

The ANOVA on HNR produced a significant main effect of attitude, $F(3,365)=7.14$, $p=0.0001$, and a significant attitude by phrase type interaction, $F(6,365)=5.11$, $p=0.000\,05$.[3] For keyphrases only, humor exhibited significantly lower HNR values than utterances conveying all other attitudes. Both humorous keyphrases and humorous sentences also demonstrated significantly lower HNR values than humorous combined sentences.

E. Expressing sarcasm in Cantonese versus English

The second major goal was to highlight similarities and differences in the expression of sarcasm between languages by directly comparing the present acoustic data on Cantonese with published data on English (Cheang and Pell, 2008). This was possible as the Cantonese exemplars were devised, produced, validated, and measured acoustically in a manner identical to that of the prior study, although English exemplars needed to be renormalized in the same fashion as the Cantonese tokens to allow comparisons across data sets.[4] As noted earlier, phrase type was eliminated from these analyses by collapsing data along this factor to focus exclusively on the prosodic differences among attitudes; this manipulation is justified on the basis of the previous finding of global acoustic cues marking sarcasm in English (Cheang and Pell, 2008). The English and Cantonese data were entered into a series of ANOVAs involving the between-subjects factor of language (Cantonese and English) with repeated measures on attitude (sarcasm, sincerity, humor, and neutrality). Normalized data, which entered into the cross-language comparison, are furnished in Table IV (see Cheang and Pell, 2008 for the raw acoustic measures pertaining to English).

F. Fundamental frequency: Mean and range

Cross-language analysis of mean F0 data yielded a significant main effect of attitude, $F(3,866)=37.88$, $p<0.0001$, and an interaction of language by attitude, $F(3,866)=22.76$, $p<0.0001$. *Post hoc* elaboration of the interaction revealed that for Cantonese, sarcasm was produced with higher F0 levels than sincerity and neutrality (which was significantly lower than all attitudes). In contrast, English exemplars of sarcasm and neutrality exhibited significantly *lower* mean F0 than sincerity and humor. Across languages, sarcasm exhibited a significantly higher mean F0 in Cantonese than in English, and sincerity exhibited a significantly lower mean F0 in Cantonese than in English. There were no cross-linguistic differences in the mean F0 for neutral or humorous utterances (see Fig. 2).

The ANOVA of F0-range yielded a significant main effect of attitude, $F(3,866)=18.88$, $p<0.0001$. Irrespective of language, sarcasm was produced with a greater F0-range than neutral utterances, but with a narrower F0-range than sincere utterances. Neutrality exhibited the smallest F0-range, differentiating these utterances from all other attitudes. There was no interaction of attitude and language.

G. Amplitude: Mean and range

For mean amplitude, statistical analyses revealed a significant main effect of attitude, $F(3,865)=34.85$, $p<0.0001$, and a significant interaction of attitude and language, $F(3,865)=2.82$, $p=0.038$. The interaction was accounted for by the observation that neutral exemplars were produced with lower amplitude than all other attitudes in both languages. Also, there were differences between sarcasm and humor, which varied by language: In English, sarcasm was spoken with *lower* amplitude than humor, whereas in Cantonese sarcasm was spoken with *greater* amplitude than humor.

For amplitude-range, a significant main effect of attitude, $F(3,865)=6.48$, $p=0.0002$, and a significant language by attitude interaction, $F(3,865)=6.56$, $p=0.0002$, were found. In Cantonese, sarcasm was produced with a significantly restricted amplitude-range relative to neutral and sin-

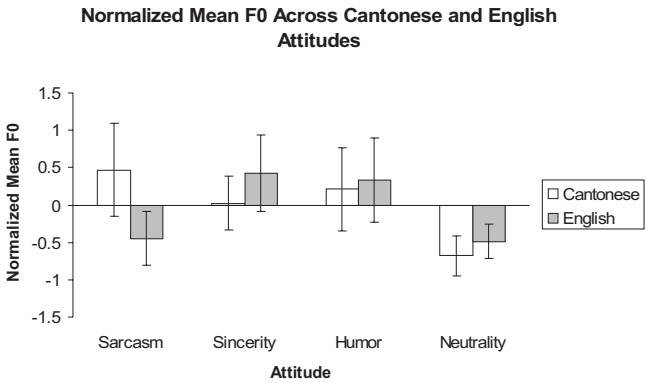

FIG. 2. Mean fundamental frequency (F0) of Cantonese versus English utterances conveying sarcasm, sincerity, humor, and neutrality. Error bars represent standard deviations.

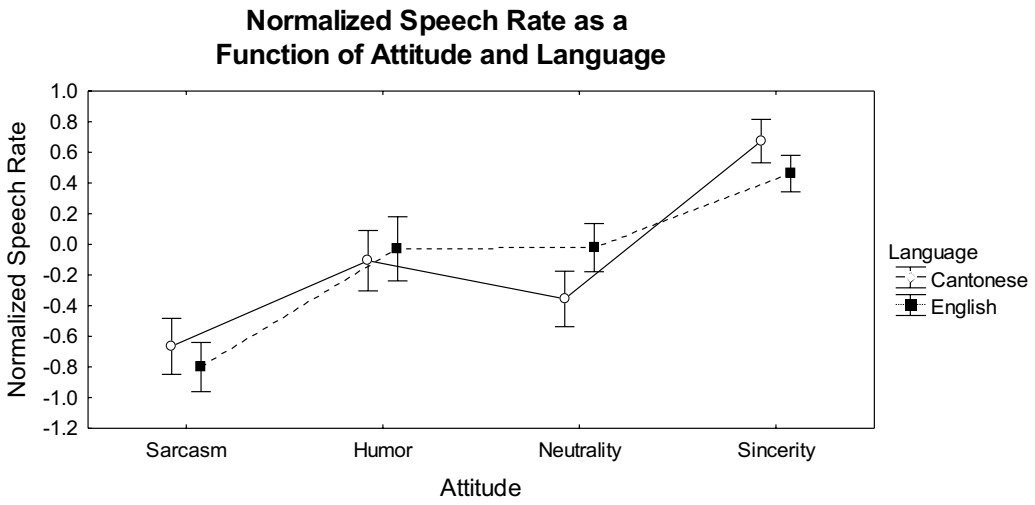

FIG. 3. Mean speech rate (syllables/s) of Cantonese versus English utterances conveying sarcasm, sincerity, humor, and neutrality. Error bars represent standard deviations.

cere utterances. Humor demonstrated a smaller amplitude-range than neutrality. English exemplars did not differ in amplitude-range as a function of attitude. Between languages, sarcastic utterances in Cantonese displayed a more restricted amplitude-range than in English.

H. Speech rate

Analyses of speech rate revealed a significant main effect of attitude, $F(3,866)=98.62$, $p<0.0001$, and an interaction between language and attitude, $F(3,866)=4.63$, $p=0.003$. In both Cantonese and English, sarcasm was produced with a significantly slower speech rate than all other attitudes (with the exception of neutrality in Cantonese). Sincerity was associated with the fastest articulation rate when compared to the other attitudes in both languages. In English, neutral utterances were spoken at a faster rate than humor, whereas in Cantonese neutral utterances were spoken more slowly than humor (see Fig. 3).

I. HNR

A significant main effect of attitude, $F(3,853)=3.71$, $p=0.01$, and a significant interaction of language and attitude, $F(3,853)=9.09$, $p<0.0001$, were found for measures of HNR. For Cantonese, the HNR of sarcastic utterances was significantly greater than that of sincere and humorous utterances, which in turn were greater than neutrality. There were no HNR differences as a function of attitude for English.[5] Between languages, Cantonese sarcasm displayed a higher HNR than English sarcasm, and Cantonese humor was produced with a lower HNR than in English.

IV. DISCUSSION

A. Overview

The current results provide novel data, which show that the expression of sarcasm in Cantonese speech is associated with a specific set of acoustic cues. Moreover, comparative analyses between languages show that the pattern in Cantonese is somewhat different from that used by English speakers to convey sarcasm (Cheang and Pell, 2008), although particular acoustic cues appear to share similar signaling functions between languages as discussed below. In Cantonese, the acoustic features of sarcastic utterances differentiated most clearly from sincere utterances: For this contrast, sarcasm displayed a higher mean F0, and narrower range in both F0 and amplitude, than utterances perceived as sincere in Cantonese speech. The data imply that Cantonese speakers encoded these particular features to convey sarcasm irrespective of the type/length of utterance produced (i.e., there was no influence of different phrase types on these acoustic variables). Other acoustic parameters were sometimes associated with sarcastic speech but only for specific phrase types; for example, sarcasm displayed higher HNR values (i.e., lower noise levels) than humor when keyphrase stimuli were produced.

Here, speaker mean F0 appeared to be the most important acoustic parameter for marking sarcasm. This finding is consistent with previous studies of sarcasm or verbal irony (Cutler, 1974, 1976; Haiman, 1998; Mueke, 1969, 1978; Myers-Roy, 1976; Rockwell, 2007, 2000a, 2005; Schaffer, 1982). This observation also fits the broader literature, which argues that mean F0 is a pivotal cue for signaling a variety of affective and attitudinal states across cultures and languages (see Banse and Scherer, 1996, for an overview). However, there were cross-language differences in how mean F0 was used to convey sarcasm; in Cantonese, mean F0 tended to be raised, whereas this parameter was lowered in the same context in English. Interestingly, Cantonese demonstrates similar patterns to Italian and French, which have both been associated with elevated mean F0 levels when speakers of these languages express sarcasm (Anolli *et al.*, 2002; Laval and Bert-Erboul, 2005). Collectively, the data suggest that the manner in which speakers exploit mean F0 to communicate sarcasm across languages is dictated to a considerable extent by social conventions. This differentiates how speakers use F0 to mark social intentions in spoken language, such as sarcasm, from how they use F0 to express basic emotions

such as fear, which demonstrate similar tendencies in speech irrespective of the speaker's language or culture (Pell *et al.*, 2009).

The finding that sarcasm is produced with less amplitude variation (range) in Cantonese, and frequently with a slower speech rate, coincides with data on English and other languages such as Japanese, Italian, and French (e.g., Adachi, 1996; Anolli *et al.*, 2002; Bryant and Fox Tree, 2005; Laval and Bert-Erboul, 2005; Rockwell, 2000a). It has been suggested that a reduced speech rate has the effect of drawing listener focus to a particular excerpt of discourse for a number of communicative purposes (Haiman, 1998; Kreuz and Roberts, 1995). The present results contribute to the argument that speakers of many languages use reduced speech rate to alert the listener to the intended, sarcastic meaning of utterances. It is possible that restrictions of amplitude variability serve a similar goal of the speaker as reduced speech rate to focus the listener to the fact that a non-literal message is intended. These suggestions could be usefully tested in future studies and in other contexts in which non-literal meanings are marked through prosody.

HNR was a significant cue, which differentiated expressions of sarcasm versus humor in certain contexts (when keyphrases were produced), indicating that speakers may modulate their voice quality at times to communicate sarcasm or irony. Attempts to convey a humorous attitude were associated with reliably greater amounts of noise (lower mean HNR values) when compared to analogous exemplars of sarcasm. Possibly, HNR fluctuations reflect changes in facial gestures, which occur during the expression of a playful intent; for example, smiling can be audibly detected by listeners (Auberge and Cathiard, 2003), although the mechanism through which this is accomplished is not yet fully specified (Tartter and Braun, 1994). Alternately, speakers may purposely deviate from their normal registers in attempting to convey humorous attitudes (Norrick, 2004), and this could introduce more noise into the speech signal. Interestingly, no other prosodic cues in these data clearly differentiated humorous utterances from the other attitudes overall; distinctions between humor and the other attitudes arose in unpredictable combinations of attitude and phrase type relative to sarcasm and sincerity. Nonetheless, data presented in Tables II–IV indicate that the two separate types of verbal irony targeted in this study show distinctions and should therefore not be treated as a single context with a unitary expression (Gibbs, 2000).

B. Distinctions between sarcasm and related attitudes

When expressions of sarcasm in Cantonese and English are compared to other attitudes, the clearest acoustic distinctions that emerged in both languages were between sarcasm and sincerity. This systematic distinction hints at a common communicative principle operating in both Cantonese and English: Sarcastic speech tends to be marked by a "play" cue, or a (meta-)message that signals to the listener that the speaker does not mean what they say, which is encoded by particular acoustic changes (Haiman, 1998). In contrast, sincere declarations are statements that are meant to reinforce the speaker's intended, literal meaning (even if implicature is required in some cases). Sincere utterances occur more frequently than sarcastic ones, and one can speculate that this context represents the "unmarked" mode of expression for most speakers. In the absence of linguistic cues to understand speaker intentions, it would be essential for sincere utterances to be highly distinct from (sarcastic) utterances that contain the play cue, which appears to be the case in both of the languages studied here.

When compared to sarcasm and sincerity, sarcasm and humor exhibited relatively few acoustic differences. This may be due to the fact that the sarcastic and humorous utterances evaluated here both represent instances of subtypes of verbal irony that share conceptual features (such as requiring an extra play cue to be recognized, Gibbs, 2000; Haiman, 1998). Rather than differences in acoustic features, contextual factors may be more important for recognizing humorous intent when compared to sarcasm. It has been argued that following the introduction of a play cue, a speaker must create a surprising incongruity in discourse and cohesively resolve it in order to successfully convey various forms of humor, ironic or otherwise (Berger, 1987; Berlyne, 1972; Brownell *et al.*, 1983; Cunningham and Derks, 2005; McGhee, 1976; Shultz and Horibe, 1974; Shultz, 1972; Suls, 1983; Wicker *et al.*, 1980, 1981). In the validation study, it may have been more difficult for decoders to appreciate the intended humorous intent of utterances because they did not have a context for interpreting humor. This likelihood is suggested by the relatively small number of humor exemplars retained for acoustic analysis (review Table II). Nonetheless, the present results do capture some of the possible acoustic correlates of humor, since many of these tokens were perceived correctly and subjected to acoustic analyses.

Finally, it was obvious that neutral utterances were acoustically distinct in nearly every respect from the other three attitudes. This point is important because many previous studies have simply divided utterances into the categories of "sarcastic (or ironic)" and "non-sarcastic (or non-ironic)," without particular attention to whether the comparison tokens were examples of neutrality or sincerity (e.g., Bryant, 2007; Bryant and Fox Tree, 2002; Rockwell, 2000a). The fact that previous studies may have inadvertently conflated neutral and sincere utterances as a baseline for characterizing sarcasm may underlie some of the disparities in the acoustic literature on sarcasm.

C. Issues of language

As reported previously for English (Cheang and Pell, 2008), the current data on Cantonese indicate that phrase type plays an important role in how prosody is used to convey speaker attitudes, especially for sarcasm and sincerity. A review of Table II shows that the presence of Cantonese keyphrases greatly increased the likelihood of decoders identifying sarcasm, whereas the absence of such phrases biased a sincere interpretation. This finding confirms that certain phrases or words can be largely associated with sarcastic meaning in both English and Cantonese (Haiman, 1998; Matthews and Yip, 1994). The present results may be a start-

ing point from which to detail the range of semantic terms that predispose sarcastic intentions in future studies, although they do not negate the fact that speakers use prosody in isolation to express sarcasm.

In the case of tone languages such as Cantonese, suggestions have been made that the set of acoustic cues available for expressing global meanings through prosody may be restricted in tonal languages due to the important role of pitch for signaling lexico-semantic information (Ross *et al.*, 1986). This notion was not borne out in the current study. Despite certain cross-language differences in the directionality of acoustic cues used to mark attitudes in Cantonese and English, speakers of both languages tended to exploit the *same* prosodic markers to convey attitudes (particularly mean F0). This finding suggests that local pitch phenomena in Cantonese do not hinder the use of F0 for other signaling functions over longer time domains (Vance, 1976).

D. Future directions

While great efforts were made to perceptually validate the present materials prior to acoustic analysis, the present findings are somewhat limited by the small number of items used and the small number of encoders who produced exemplars of the four attitudes. The generalizability of the present findings may be also somewhat restricted given that the present acoustic analyses were based on non-spontaneous speech. Since the encoders were intentionally given no explicit instructions as to how to acoustically pattern their speech productions, the encoders may have attempted to produce stylized "folk models" of sarcastic speech in addition to simulations of their personal style of articulating sarcasm in natural discourse. As such, the present results may suggest that speakers have less variable modulation of the individual acoustic components of sarcastic prosody than would be naturally observed in spontaneous conversation. Regardless, recognizably sarcastic prosody must have been adequately emulated by the encoders in the speech tokens, given the converging identification rates of a majority of naïve decoders during validation. Furthermore, speakers do use a diverse array of cues to signal sarcasm in natural speech, including highly stylized speech (Haiman, 1998). Therefore, although the current acoustic analyses are based on non-spontaneous tokens, the findings should reflect aspects of sarcastic prosody that occur in typical discourse.

Nonetheless, future studies would benefit from analyzing spontaneous excerpts of sarcasm elicited from more encoders; this should mitigate the relatively high number of recordings which were excluded from the present analyses and provide an even more comprehensive profile of sarcastic prosody. The nature of excluded tokens may also be of interest in future work: Since there appears to be a certain flexibility in how speakers communicate sarcasm (Haiman, 1998), analyzing tokens that are ambiguous or poorly recognized as sarcasm may be of value for revealing additional cues tied to this context. Future research should also consider an even larger set of acoustic parameters, which may be associated with sarcastic speech; some potential measures to examine include the number and length of pauses, different voice registers, and resonance changes (see Haiman, 1998 for an overview). The role of non-verbal cues such as facial expressions, and how these cues interact with prosody in the context of sarcasm, is also another avenue to explore (cf. Rockwell, 2000b, 2001, 2005).

Finally, given the differences and similarities noted here in how sarcasm is encoded in Cantonese versus English, these data raise the question of whether sarcasm can be detected from prosody in the speaker's non-native language. There is evidence that widely disparate cultures base their recognition of attitudes or emotions on culturally-defined factors such as overall communication style, at least in part (Elfenbein and Ambady, 2002, 2003; Kitayama and Ishii, 2002). Given the observed language-related differences in the direction of mean F0 change, which cues sarcasm in Cantonese versus English, one can speculate that listeners of one language would experience difficulties using cues in the other language correctly to infer when speakers intend to be sarcastic. This hypothesis is being tested by presenting both the Cantonese and English exemplars to monolingual speakers of each language (Cheang and Pell, in preparation).

ACKNOWLEDGMENTS

Support for this work was provided by doctoral training scholarships awarded to H.S.C. in the form of a CIHR-K.M. Hunter Doctoral Training Award from the Canadian Institutes of Health Research (CIHR) and a Bridge Funding Award from the Centre for Research on Language, Mind and Brain (CRLMB), an interdisciplinary unit in McGill University; and a Discovery Grant from the National Sciences and Engineering Research Council of Canada (NSERC) to M.D.P. The authors would also like to thank Elmira Chan and Marie Desmarteau for assistance in data collection.

[1]Cantonese speakers may add particles (that are constrained only by having to be in utterance-final position) that freely vary (i.e., can be paired with any context) to utterances in addition to (or instead of) modifying prosodic cues over the course of a sentence to help convey various attitudes or serve grammatical functions such as specifying utterance mode (Matthews and Yip, 1994). Inclusion of such particles in the present materials would limit comparability with earlier English materials devised by the current authors as there are no English equivalents. Perhaps, more importantly, acoustic variation may be (but not necessarily) concentrated disproportionately in particles for the purposes of attitude expression when such particles are present (Chan, 2001; Matthews and Yip, 1994). Such possible interactions between particles and prosody are not the present focus and, hence, particles were excluded from the ends of sentence exemplars. It should be noted that keyphrases could not be translated without the inclusion of particles, although particle inclusion here was justified as the keyphrases conveyed meaning as a unit rather than having the bulk of meaning expressed disproportionately by the particle.

[2]In Cheang and Pell, 2008, both F0 SD and F0-range were included as measures of F0 variation whereas the current study only included F0-range. F0 SD may be influenced by the very rapid lexical tonal shifts in Cantonese vowels (Bauer, 1998; Bauer and Benedict, 1997; Fok, 1974; Khouw and Ciocca, 2007; Vance, 1976, 1977) rather than from attitude expression. To avoid this confound F0 SD was dropped from consideration.

[3]The PRAAT program could not derive acoustic measurements for several tokens, hence the somewhat discrepant degrees of freedom reported across acoustic analyses.

[4]The current Cantonese utterances were subjected to *z*-score standardization. By contrast, all individual acoustic data points (from separate English utterances) were normalized per encoder in reference to the set of neutral exemplars spoken by that encoder in the authors' previous study {e.g.,

[mean F0(one keyphrase exemplar of humor) − mean F0(all keyphrase exemplars of neutrality)]/mean F0(all keyphrase exemplars of neutrality)} (Cheang and Pell, 2008). This change in strategy was adopted because naive Cantonese raters identified nearly none of the current utterances produced by one Cantonese encoder as conveying neutrality. However, the results (with respect to the acoustic cues of sarcasm for English) are very comparable across studies, suggesting that the prosodic features that emerged in the previous study were not spurious.

[5]The present lack of significant HNR differences as a function of attitude type in the English exemplars is at odds with the earlier finding of significantly reduced HNR values in English sarcasm (Cheang and Pell, 2008). The discrepancy stems from the fact that using the previous method of normalization, $p=0.046$ for the HNR attitude effect. Given the divergence, the renormalized English exemplars were analyzed in isolation using the same factors as in the ANOVAs of Cantonese tokens. Significant main effects of attitude for mean F0, significant effects of attitude and phrase type for F0 standard deviation, a significant effect of phrase type for mean amplitude and amplitude-range, and a significant attitude by phrase type interaction for speech rate were found. These effects were all attained in previous analyses of the English tokens normalized in reference to the neutral exemplars, and *post hoc* analyses yielded highly comparable results (Cheang and Pell, 2008). The great similarity in results for the analysis of the other acoustic parameters suggests that standardization was comparable across studies; the previous effect of HNR was genuine, albeit statistically weaker.

Ackerman, B. (**1983**). "Form and function in children's understanding of ironic utterances," J. Exp. Child Psychol. **35**, 487–508.

Ackerman, B. (**1986**). "Children's sensitivity to comprehension failure in interpreting a nonliteral use of an utterance," Child Dev. **57**, 485–497.

Adachi, T. (**1996**). "Sarcasm in Japanese," Studies in Language **19**, 1–36.

Anolli, L., Ciceri, R., and Infantino, M. G. (**2002**). "From 'blame by praise' to 'praise by blame': Analysis of vocal patterns in ironic communication," Int. J. Psychol. **37**, 266–276.

Attardo, S., Eisterhold, J., Hay, J., and Poggi, I. (**2003**). "Multimodal markers of irony and sarcasm," Humor **16**, 243–260.

Auberge, V., and Cathiard, M. (**2003**). "Can we hear the prosody of smile?," Speech Commun. **40**, 87–97.

Banse, R., and Scherer, K. R. (**1996**). "Acoustic profiles in vocal emotion expression," J. Pers. Soc. Psychol. **70**, 614–636.

Bauer, R. S. (**1998**). "Hong Kong Cantonese tone contours," in *Studies in Cantonese Linguistics*, edited by S. Matthews (Linguistic Society of Hong Kong, Hong Kong), pp. 1–34.

Bauer, R. S., and Benedict, P. K. (**1997**). *Modern Cantonese Phonology* (Mouton de Gruyter, Berlin), Vol. **102**.

Berger, A. A. (**1987**). "Humor: An introduction," Am. Behav. Sci. **30**, 6–15.

Berlyne, D. E. (**1972**). "Humor and its kin," in *The Psychology of Humor*, edited by P. E. McGhee and J. H. Goldstein (Springer-Verlag, New York), pp. 43–60.

Boersma, P., and Weenink, D. (**2006**). PRAAT: Doing phonetics by computer, version. 4.5.16, http://www.praat.org/ (Last viewed 6/15/2006).

Bond, M. H. (**1991**). *Beyond the Chinese Face* (Oxford University Press, Oxford).

Brownell, H. H., Michel, D., Powelson, J., and Gardner, H. (**1983**). "Surprise but not coherence: Sensitivity to verbal humor in right-hemisphere patients," Brain Lang **18**, 20–27.

Bryant, G. A. (**2009**). "Prosodic contrasts in ironic speech," Discourse Processes (in press).

Bryant, G. A., and Fox Tree, J. E. (**2002**). "Recognizing verbal irony in spontaneous speech," Metaphor and Symbolic Activity **17**, 99–117.

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

Capelli, C. A., Nakagawa, N., and Madden, C. M. (**1990**). "How children understand sarcasm: The role of context and intonation," Child Dev. **61**, 1824–1841.

Chan, M. (**2001**). "Gender-related use of sentence-final particles in Cantonese," in *Gender Across Languages: The Linguistic Representation of Women and Men*, edited by M. Hellinger and H. Bussmann (John Benjamins, Amsterdam), pp. 57–72.

Cheang, H. S., and Pell, M. D. (**2008**). "The sound of sarcasm," Speech Commun. **50**, 366–381.

Cunningham, W. A., and Derks, P. (**2005**). "Humor appreciation and latency of comprehension," Humor **18**, 389–403.

Cutler, A. (**1974**). "On saying what you mean without meaning what you say," in papers from the *Tenth Regional Meeting of the Chicago Linguistic Society*, edited by M. W. LaGaly, R. A. Fox, and A. Bruck (Chicago Linguistic Society, Chicago, IL), pp. 117–127.

Cutler, A. (**1976**). Beyond Parsing and Lexical Look-UP: An Enriched Description of Auditory Sentence Comprehension, in *New Approaches to Language Mechanisms*, edited by R. J. Wales and E. Walker (North-Holland, New York), pp. 133–149.

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

Elfenbein, H. A., and Ambady, N. (**2003**). "Cultural similarity's consequences: A distance perspective on cross-cultural differences in emotion recognition," J. Cross Cult. Psychol. **34**, 92–110.

Fok, C. Y.-Y. (**1974**). *A Perceptual Study of Tones in Cantonese* (Centre of Asian Studies, University of Hong Kong, Hong Kong), Vol. **18**.

Gibbs, R. W., Jr. (**2000**). "Irony in talk among friends," Metaphor Symb. **15**, 5–27.

Grabe, E., Rosner, B. S., Garcia-Albea, J. E., and Zhou, X. (**2003**). "Perception of English intonation by English, Spanish, and Chinese listeners," Lang Speech **46**, 375–401.

Haiman, J. (**1998**). *Talk is Cheap: Sarcasm, Alienation, and the Evolution of Language* (Oxford University Press, Oxford).

Ho, D. Y. F. (**1986**). "Chinese patterns of socialization: A critical review," in *The Psychology of the Chinese People*, edited by M. H. Bond (Oxford University Press, New York), pp. 1–37.

Huang, P. P.-F., and Kok, G. P. (**1999**). *Speak Cantonese*, 3rd ed. (Far Eastern, Detroit, MI), Vol. **1**.

Khouw, E., and Ciocca, V. (**2007**). "Perceptual correlates of Cantonese tones," J. Phonetics **35**, 104–117.

Killingley, S.-Y. (**1986**). "Normal and deviant classifier usage in Cantonese," Anthropological Linguistics **28**, 321–336.

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

Kreuz, R. J., and Roberts, R. M. (**1995**). "Two cues for verbal irony: Hyperbole and the ironic tone of voice," Metaphor and Symbolic Activity **10**, 21–31.

Kwok, H. H., and Luke, K. K. (**1986**). "Intonation in Cantonese: A preliminary study," paper presented at the *19th International Conference on Sino-Tibetan Languages and Linguistics* (Ohio State University, Columbus, OH).

Laval, V., and Bert-Erboul, A. (**2005**). "French-speaking children's understanding of sarcasm: The role of intonation and context," J. Speech Lang. Hear. Res. **48**, 610–620.

Matthews, S., and Yip, V. (**1994**). *Cantonese: A Comprehensive Grammar* (Routledge, London).

McCluskey, K. W., Albas, D. C., Niemi, R. R., Cuevas, C., and Ferrer, C. A. (**1975**). "Cross-cultural differences in the perception of the emotional content of speech: A study of the developmental of sensitivity in Canadian and Mexican children," Dev. Psychol. **11**, 551–555.

McGhee, P. E. (**1976**). "Children's appreciation of humor: A test of the cognitive congruency principle," Child Dev. **47**, 420–426.

Muecke, D. C. (**1969**). *The Compass of Irony* (Methuen, London).

Muecke, D. C. (**1978**). "Irony markers," Poetics Today **7**, 363–375.

Myers-Roy, A. (**1976**). "Towards a definition of irony," in *Studies in Language Variation*, edited by R. W. Fasold and R. Shuy (Georgetown University Press, Washington, DC), pp. 171–183.

Norrick, N. R. (**2004**). "Non-verbal humor and joke performance," Humor **17**, 401–409.

Pell, M. D., Monetta, L., Paulmann, S., and Kotz, S. A. (**2009**). "Recognizing emotions in a foreign language," J. Nonverbal Behav. **33**, 107–120.

Rockwell, P. (**2000a**). "Lower, slower, louder: Vocal cues of sarcasm," J. Psycholinguist. Res. **29**, 483–495.

Rockwell, P. (**2000b**). "Actors', partners', and observers' perceptions of sarcasm," Percept. Mot. Skills **91**, 665–668.

Rockwell, P. (**2001**). "Facial expression and sarcasm," Percept. Mot. Skills **93**, 47–50.

Rockwell, P. (**2005**). "Sarcasm on television talk shows: Determining speaker intent through verbal and nonverbal cues," in *Psychology of Moods*, edited by A. Clark (Nova Science, New York), pp. 109–140.

Rockwell, P. (**2007**). "Vocal features of conversational sarcasm: A comparison of methods," J. Psycholinguist. Res. **36**, 361–369.

Ross, E. D., Edmondson, J. A., and Seibert, G. B. (**1986**). "The effect of affect on various acoustic measures of prosody in tone and non-tone languages: A comparison based on computer analysis of voice," J. Phonetics

14, 283–302.
Schaffer, R. (**1982**). "Are there consistent vocal clues for irony?," in *Parasession on Language and Behavior*, edited by C. S. Masek, R. A. Hendrick, and M. F. Miller (Chicago Linguistic Society, Chicago, IL), pp. 204–210.
Scherer, K. R. (**1986**). "Vocal affect expression," Psychol. Bull. **99**, 143–165.
Scherer, K. R., Ladd, D. R., and Silverman, K. E. (**1984**). "Vocal cues to speaker affect: Testing two models," J. Acoust. Soc. Am. **76**, 1346–1356.
Shultz, T. R. (**1972**). "The role of incongruity and resolution in children's appreciation of cartoon humor," J. Exp. Child Psychol. **13**, 456–477.
Schultz, T. R., and Horibe, F. (**1974**). "Development of the appreciation of verbal jokes," Dev. Psychol. **10**, 13–20.
Snow, D. (**2004**). *Cantonese as Written Language: The Growth of a Written Chinese Vernacular* (Hong Kong University Press, Hong Kong).
Suls, J. (**1983**). "Cognitive processes in humor appreciation," in *Handbook of Humor Research*, edited by P. E. McGhee and J. H. Goldstein (Springer-Verlag, New York), pp. 39–57.
Tartter, V. C., and Braun, D. (**1994**). "Hearing smiles and frowns in normal and whisper registers," J. Acoust. Soc. Am. **96**, 2101–2107.
Vance, T. J. (**1976**). "An experimental investigation of tone and intonation in Cantonese," Phonetica **33**, 368–392.
Vance, T. J. (**1977**). "Tonal distinctions in Cantonese," Phonetica **34**, 93–107.
Wicker, F. W., Barron, W. L. I., and Willis, A. C. (**1980**). "Disparagement humor: Dispositions and resolutions," J. Pers. Soc. Psychol. **39**, 701–709.
Wicker, F. W., Thorelli, I. M., Barron, W. L. I., and Ponder, M. R. (**1981**). "Relationships among affective and cognitive factors in humor," J. Res. Pers. **15**, 359–370.
Winner, E., and Leekman, S. (**1991**). "Distinguishing irony from deception: Understanding the speaker's second-order intention," Br. J. Dev. Psychol. **9**, 257–270.
Yumoto, E., Gould, W. J., and Baer, T. (**1982**). "Harmonics-to-noise ratio as an index of the degree of hoarseness," J. Acoust. Soc. Am. **71**, 1544–1550.