Available online at www.sciencedirect.com
ScienceDirect
Speech Communication 50 (2008) 366–381
SPEECH COMMUNICATION
www.elsevier.com/locate/specom

The sound of sarcasm

Henry S. Cheang *, Marc D. Pell

McGill University, School of Communication Sciences and Disorders, 1266 Pine Avenue West, Montreal, Quebec, Canada H3G 1A8

Received 4 April 2007; received in revised form 17 November 2007; accepted 19 November 2007

Abstract

The present study was conducted to identify possible acoustic cues of sarcasm. Native English speakers produced a variety of simple utterances to convey four different attitudes: sarcasm, humour, sincerity, and neutrality. Following validation by a separate naïve group of native English speakers, the recorded speech was subjected to acoustic analyses for the following features: mean fundamental frequency (*F*0), *F*0 standard deviation, *F*0 range, mean amplitude, amplitude range, speech rate, harmonics-to-noise ratio (HNR, to probe for voice quality changes), and one-third octave spectral values (to probe resonance changes). The results of analyses indicated that sarcasm was reliably characterized by a number of prosodic cues, although one acoustic feature appeared particularly robust in sarcastic utterances: overall reductions in mean *F*0 relative to all other target attitudes. Sarcasm was also reliably distinguished from sincerity by overall reductions in HNR and in *F*0 standard deviation. In certain linguistic contexts, sarcasm could be differentiated from sincerity and humour through changes in resonance and reductions in both speech rate and *F*0 range. Results also suggested a role of language used by speakers in conveying sarcasm and sincerity. It was concluded that sarcasm in speech can be characterized by a specific pattern of prosodic cues in addition to textual cues, and that these acoustic characteristics can be influenced by language used by the speaker.

Keywords: Verbal irony; Sarcasm; Prosody; Acoustic cues

1. Introduction

Communication and speech comprehension are heavily dependent on the use of implicit information (Sabbagh, 1999). Speakers convey implicit information to listeners by manipulating language and prosody (i.e., intonation and stress patterns), among other features, to express a particular message. The rules that govern how speakers produce language are well-documented (e.g., Grice, 1975). In contrast, how prosodic cues are used to express affective and attitudinal states has been studied much less. One context in which prosodic features appear to play a significant role is the communication of verbal irony, of which a key subtype is sarcastic irony, i.e., sarcasm (Kreuz and Roberts, 1993).

Verbal irony can be defined as expressions in which the intended meaning of the words is different from or the direct opposite of their usual sense; these expressions serve numerous functions in communication (see Gibbs, 2000; Haverkate, 1990, for a description of the forms and functions of verbal irony). Sarcasm is verbal irony that expresses negative and critical attitudes toward persons or events (Kreuz and Glucksberg, 1989). However, it bears noting that while researchers typically refer to or study "verbal irony" or "irony", they are generally referring to the negative attitude projected by ironic speakers. Hence, in many instances, the terms "verbal irony" and "sarcasm" have been conflated (Capelli et al., 1990). To be explicit, the focus of the present study is sarcasm because of its importance in communication. For example, sarcastic comments are quite pervasive in conversation, perhaps because listeners tend to find these remarks less threatening and more polite than overtly critical statements (Dews et al., 1995; Gerrig and Goldvarg, 2000; Jorgensen, 1996; Kumon-Nakamura et al., 1995). Viewed in another way, sarcastic

* Corresponding author. Tel.: +1 514 398 8496; fax: +1 514 398 8123.
E-mail addresses: henry.cheang@mail.mcgill.ca (H.S. Cheang), marc.pell@mcgill.ca (M.D. Pell).

0167-6393/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.specom.2007.11.003

comments can act to highlight and enhance the critical message intended by speakers (Colston, 1997). Overall, the degree to which sarcastic comments are considered polite or critical appears to vary as a function of the surface form of the message; sarcastic compliments are perceived as more mocking and less polite than direct compliments, whereas sarcastic insults are perceived as more mocking and more polite than direct insults (Pexman and Olineck, 2002). While not the focus of this study, a number of theories have been put forth to account for the contexts and linguistic mechanics under which speakers express the negative subtype of verbal irony, i.e., sarcasm (e.g., Clark and Gerrig, 1984; Grice, 1975; Sperber, 1984).

A point of general agreement for which there exists relatively little cohesive data is that speakers can signal their sarcastic intent to listeners using the prosodic content of speech. One explanation for the inconsistent amount of data is that different researchers have been studying distinct subtypes of verbal irony that may present with different prosodic patterns which may in turn lead to the conclusion that no true acoustic patterns of irony exist. Speakers may be using non-standard shifts in communicative contexts (i.e., changes in acoustic parameters, semantics, discourse, and facial expression, among others) to signal different subtypes of verbal irony (including sarcasm) by highlighting a deviation from an expected or typical message to listeners (Attardo et al., 2003; Bryant, in press; Bryant and Fox Tree, 2005; Kreuz and Roberts, 1995; Wilson and Wharton, 2006). An alternative perspective is that, in conjunction with situational context and vocabulary choice, specific acoustic cues known collectively as the "ironic tone of voice" help listeners to know when sarcasm is intended (Clark and Gerrig, 1984). However, given the different subtypes of verbal irony that exist (Gibbs, 2000), it may be more accurate to collectively refer to cues marking sarcastic speech as a sarcastic tone of voice (even though researchers typically do not make such a distinction). In principle, such a pattern of acoustic cues is similar to the predictable changes in acoustic cues that are associated with many affective and attitudinal states (Banse and Scherer, 1996). Note also that listeners can accurately recognize emotions (e.g., joy, anger) and certain attitudes (e.g., confidence, politeness) when listening to semantically-meaningless "pseudo-utterances" which communicate these meanings strictly through prosodic cues (Dara et al., in press; Monetta et al., in press; Pell, 2006). In an analogous manner, it is possible that speakers use a relatively consistent set of acoustic markers in conjunction with linguistic and contextual cues to signal sarcastic intent in speech.

In one study, Bryant and Fox Tree (2002) extracted sarcastic and non-sarcastic utterances from live radio conversations to determine whether listeners could discern sarcasm under different experimental conditions. In the absence of their original contexts, the texts of sarcastic and non-sarcastic target utterances were not found to be significantly non-biasing toward or away from a sarcastic interpretation when presented in a written format (i.e., in the absence of prosodic cues); however, when these utterances were heard by a separate group of participants, utterances extracted from sarcastic contexts were rated as significantly more sarcastic than those from non-sarcastic contexts. As well, additional participants could successfully match the spoken utterances to transcribed versions of their original contexts. One final separate group of participants rated the originally-sarcastic utterances as more sarcastic regardless of the presence of written versions of the original sarcastic contexts (Bryant and Fox Tree, 2002). These results show that the intent of sarcastic utterances can be recognized independent of their contexts, although one shortcoming cited by the authors is that the effects of verbal semantics and acoustic features for understanding sarcasm cannot be truly dissociated with these methods. Semantic cues (i.e., words or phrases) often signal sarcasm in conjunction with prosodic cues (Bolinger, 1989; Haiman, 1998) and there are some phrases or words which may be so closely tied to the sarcastic context that these expressions can independently signal sarcasm (i.e., these expressions become "enantiosemantic", Haiman, 1998, p. 39). Thus, the role of prosody for communicating sarcasm cannot be determined easily from this study.

In a follow-up study, Bryant and Fox Tree (2005) investigated how sarcasm is perceived from spontaneous utterances taken from radio shows when the stimuli were content-filtered (i.e., semantic cues were rendered unintelligible). They found that filtered utterances that conveyed "dripping sarcasm" (i.e., utterances in which the sarcastic intent was conveyed in a semantically and prosodically unambiguous fashion, p. 260) were correctly rated as more sarcastic by listeners than non-sarcastic utterances filtered in the same manner (Bryant and Fox Tree, 2005). Complementary to these findings, Rockwell (2000b) showed that naïve listeners can discern sarcasm from posed, although not spontaneous, utterances that were content-filtered prior to presentation. Rockwell's (2000b) analyses further revealed that listeners perceive sarcastic utterances as having lower pitch (i.e., the perceptual correlate of fundamental frequency ($F0$)), slower tempo (i.e., the perceptual correlate of speech rate) and greater loudness (i.e., the perceptual correlate of amplitude), although these conclusions were derived from analyses of perceptual attributes and hence do not supply direct evidence of acoustic changes which might correspond with sarcastic speech.

In the developmental literature, there are also compelling indications that contextual and prosodic cues for appreciating sarcastic intent in speech can be dissociated. Studies which have required children of different ages to recognize sarcasm on the basis of context alone, intonation alone, or combinations of the two sets of cues have shown a developmental progression in the differential use of sarcastic cues. In some reports, sensitivity to sarcastic situational contexts appears to develop first in young children (Ackerman, 1983, 1986; Winner and Leekman, 1991), whereas others claim that sensitivity to sarcastic intonation is acquired first (Capelli et al., 1990; Laval and Bert-

Erboul, 2005; Laval, 2004). Regardless of which milestone is attained first, these findings imply a functional dissociation in the cues which contribute to the expression of sarcasm in speech, underscoring the importance of prosody in this context and allowing for the possibility that a uniquely sarcastic prosody exists alongside contextual markers of this attitude.

To date, relatively little work focusing on sarcasm has been done that specifically examined the acoustic cues of sarcasm. Among the acoustic features most frequently cited in this literature are: heightened pitch/*F*0 variation; heightened loudness; alterations in speech timing (e.g., reduced speech rate or increased number of pauses); varied changes in voice quality; extra nasal resonance; and monotonic or lowered pitch/*F*0 (Cutler, 1974, 1976; Fonagy, 1971; Haiman, 1998; Mueke, 1969, 1978; Rockwell, 2000a; Schaffer, 1982). More recent work has shown exemplars of sarcasm to have very complex *F*0 manipulations across productions (Attardo et al., 2003) or to have high *F*0 levels (Laval and Bert-Erboul, 2005), or a combination of high *F*0 levels, high amplitude and a voice quality that could be characterized as "tight" (Anolli et al., 2002). Many of these findings highlight the importance of temporal and pitch (*F*0) measures, although note the differences in whether sarcasm is associated with a higher (Anolli et al., 2002; Attardo et al., 2003; Laval and Bert-Erboul, 2005) or lower (Rockwell, 2000a) pitch/*F*0 in the literature. Overall, it is difficult to infer what prosodic cues may be most critical for marking sarcasm, a situation that has perhaps arisen from the different methods used and/or differences in the languages studied across studies.

Another relevant consideration is that acoustic differences may characterize different subtypes of verbal irony. In addition to investigating the acoustic profile of sarcasm, Anolli et al. (2002) found significant differences in acoustic features between exemplars meant to express sarcasm and exemplars meant to represent humourous irony (i.e., conveying a positive and playful mood). Recall too, that in addition to conveying sarcasm and playful humour, verbal irony performs other functions as well (Colston and O'Brien, 2000a,b; Kreuz et al., 1991). Previous descriptions of acoustic markers of sarcasm may be confused with acoustic markers of other forms of verbal irony (should such cues exist) or other attitudes. We need to expand and clarify our knowledge regarding the cues that likely mark sarcasm.

1.1. The present study

In light of the evidence that prosody is important for understanding sarcasm in speech, the present study was designed to provide a comprehensive acoustic description of sarcastic utterances in English and to evaluate whether a consistent pattern of acoustic cues differentiates sarcasm from other "attitudes" as previously suggested (Clark and Gerrig, 1984; Haiman, 1998; Rockwell, 2000a). To achieve this, we recruited speakers who produced utterances to simulate sarcasm and three other "attitudes" (humour, sincerity, neutrality) and then undertook detailed acoustic analyses of those utterances which were reliably judged to communicate the intended attitude by a group of listeners. In this way, we sought to specify prosodic differences that might distinguish sarcasm from verbal irony intended to sound humourous as well as from sincere/neutral exemplars of the same utterance. We studied posed tokens of sarcasm to ensure greater experimental control over our stimuli and to permit a further manipulation of the type of utterance produced; our acoustic analyses were performed on utterances which included or excluded enantiosemantic terms (key phrases argued to be associated with sarcasm) to examine whether acoustic cues varied systematically due to the presence of this linguistic information.

To compare our findings with the existing literature, our analyses included previously-studied measures of *F*0 (mean, range, and standard deviation), amplitude (mean and standard deviation), and speech rate (e.g., Anolli et al., 2002; Bryant and Fox Tree, 2005; Rockwell, 2000b, 2005). As well, we included a measure of both voice quality and nasal resonance which have each been cited as important for expressing sarcasm, among other attitudes (e.g., Cutler, 1974; Haiman, 1998; Schaffer, 1982) but for which there are little empirical data. With respect to voice quality, the harmonics-to-noise ratio (HNR) was chosen to quantify potential differences between sarcastic utterances and other attitudes. The HNR can be defined as the averaged periodic (i.e., harmonic) component of a sound signal divided by the corresponding averaged noise component (Yumoto et al., 1982); this measure has been found to be a robust and reliable measure of voice quality changes in studies of vocal pathology and normal aging and correlates well with perceptual and other objective evaluations of voice quality in the wider literature (de Krom, 1995; Eskenazi et al., 1990; Pereira et al., 2002; Yumoto et al., 1982). In terms of nasality (i.e., resonance), one-third octave spectral analysis was employed as a means of gauging these differences and their relationship to expressions of sarcasm. This form of analysis, which is performed by applying a one-third octave filter to the important frequencies of a steady state portion of a vowel and measuring the resulting values, is known to correlate well with expert ratings of the presence and degree of clinical hypernasality regardless of etiology, patient age (pediatric or adult) or gender, and across languages (Kataoka et al., 1996, 2001a,b; Lee et al., 2003, 2004; Yoshida et al., 2000). These additional acoustic measures allow the present literature on sarcasm to be extended in a meaningful way.

Based on the literature reviewed, it was expected that exemplars of sarcasm would differ significantly from exemplars of humour and sincerity on measures of *F*0, amplitude, speech rate, voice quality, and nasality. While disparities in the literature prevent precise statements regarding the directionality of these cues, it was expected that sarcastic utterances would display a lower *F*0, greater

*F*0 variability, greater amplitude, and reduced speech rate when compared to utterances expressing the other attitudes (especially corresponding utterances which were intended to be sincere). The effects of communicating sarcasm or other attitudes on our measure of voice quality and nasality could not be predicted with any certainty from the literature. For example, while HNR is very sensitive to the presence of hoarseness (Yumoto et al., 1982), breathiness (de Krom, 1995), and general vocal changes brought about by pathology (Eskenazi et al., 1990; Pereira et al., 2002); or normal aging (Ferrand, 2002), this measure has not been applied to the investigation of sarcasm. The effects of sarcasm on our measure of nasality were similarly exploratory in nature.

2. Method

2.1. Stimuli production

2.1.1. Encoders and materials

Six native English speakers or "encoders" (three male and three female; mean age in years: 21.43, SD: 1.63; mean years of education: 15.33, SD: 1.63) were recruited to pose expressions of four distinct attitudes (sarcasm, humour, sincerity, and neutrality). As described below, each encoder was recorded producing a total of 96 utterances (24 representing each of the four attitudes). With the exception of neutrality, each recorded exemplar was generated in response to a biasing sentence to promote relatively naturalistic productions of a given attitude (i.e., as if the encoder were engaging in a scripted dialogue; see Table 1 for examples of all recording materials).

Target utterances were constructed to control the semantic content, syntactic content, number of syllables, and word frequency of the vocabulary (which was restricted to high-frequency words). As the primary purpose was to isolate possible acoustic cues of sarcasm, the text of each item was devised in a way that it could be spoken with each of the four attitudes without any changes. To facilitate productions of sarcasm, humour, and sincerity, sets of biasing sentences were created which were suitable to each attitude; for example, biasing sentences designed to promote the expression of sarcasm included very obvious harsh, unfairly critical, or insulting cues (e.g., "horrid woman" from Table 1). For humour, biasing sentences incorporated playful cues or explicit mention of a friendly relation (e.g., "your friend Shelley" from Table 1). To elicit sincerity, biasing sentences did not contain any information that suggested positive or negative associations; neither were there playful nor insulting cues. No biasing sentences were constructed to assist encoders to express neutrality.

A secondary objective was to evaluate the possible impact of different utterance forms on how speakers modulate the acoustic cues to express sarcasm. Therefore, within each group of tokens that expressed the target attitudes, the utterances took three different forms: utterances characterized by "keyphrases" cited by some authors as more likely to convey sarcasm than other comparable terms; sentences without keyphrases; and "combined sentences" composed of the keyphrases and the sentences combined in one utterance. Four sets of the three phrase types were constructed in total. The text of these exemplars ranged from 2 to 3 syllables for keyphrases and 7 to 8 syllables for sentences (review Table 1). When producing a given attitude, the same biasing sentence was used to elicit productions for each of the three associated phrase types. Hence, for every exemplar in the "healthy lady" set in Table 1, the biasing sentence (for sincerity) was "She runs 10 miles a day". To ensure the suitability of recording

Table 1
Biasing sentences and elicited sentences forming dialogues across four attitudes (sarcasm, humor, sincerity, and neutrality)

Phrase type	Text of elicited exemplar	Attitude	Biasing sentences associated with phrase type cluster
A:	I suppose; it's a respectful gesture.	Sarcasm	Don't you just love how your stupid mother-in-law always smirks and snorts loudly when you misspeak?
B:	It's a respectful gesture.	Humor	Not everyday that you see a priest give the finger, is it?
C:	I suppose.	Sincerity	It was nice of your supervisor to send flowers.
A:	Is that so; she is a healthy lady.	Sarcasm	That horrid woman smokes a pack a day.
B:	She is a healthy lady.	Humor	Your friend Shelley can't even do a single pushup.
C:	Is that so?	Sincerity	She runs 10 miles everyday.
A:	Oh boy; he is a superior chef.	Sarcasm	Our moronic boss gave us all food poisoning.
B:	He is a superior chef.	Humor	Your brother singed his eyebrows while making toast.
C:	Oh boy.	Sincerity	Butch just won another cooking contest; this is his twentieth win in the last 3 years.
A:	Yeah, right; what a spectacular result.	Sarcasm	The arrogant front-runner finished dead last.
B:	What a spectacular result.	Humor	Fascinating how she lost the eating contest to someone half her size huh?
C:	Yeah, right.	Sincerity	He broke three records in that race!

"A" items denote combined sentences, "B" items denote single sentences, and "C" items denote keyphrases. Note that no biasing sentences were provided to encoders to facilitate their production of neutrality.

materials, all biasing sentence/target utterance pairs were presented in written format to four native English speakers who judged the naturalness of the sentence pairs in a pilot study. The pilot participants were instructed to rate the pairs as being "natural", "somewhat natural", or "unnatural". Both biasing sentences and target utterances were revised on the basis of these ratings and submitted to different English speaking participants. Refinement continued until there was at least 75% agreement across raters that all sentence pairs reflected communicative interactions that were perceived as natural for the intended attitude.

2.1.2. Recording procedure

Each target utterance was recorded twice (non-sequentially) to express each attitude per encoder. During the recording session, the encoder was presented a series of cards which contained the biasing sentence (except for neutrality) and an associated utterance form (i.e., a keyphrase, a sentence, or a combined sentence). Encoders had to silently read the biasing sentence and then produce the target sentence to communicate the intended target attitude. Recording sessions always began with the elicitation of neutral utterances; in this condition encoders were asked to simply read the target sentence in a voice that did not convey any particular affect or attitude. Once the entire set of neutral utterances was recorded, utterances conveying the remaining attitudes were recorded in a fixed random sequence which was the same for all encoders. The total number of exemplars recorded in the experiment was 576 (i.e., 6 encoders $\times$ 4 attitudes $\times$ 4 items $\times$ 3 utterance forms $\times$ 2 repetitions).

At the onset of recording, encoders were provided with the definitions of each attitude and given brief, standardized descriptions of situations under which interlocutors are likely to express sarcasm, humour, or sincerity (e.g., "people use sarcastic utterances to respond to insulting comments directed at them"; "people use humourous statements to be playful with friends"). Given the primary aim of the current study (i.e., evaluate the potential acoustic features of sarcastic verbal irony against other forms, including positive verbal irony), it was important to highlight the fundamentally negative nature of sarcasm as the negativity represents the chief distinction across the target attitudes. The encoders were instructed to use the definitions, presented scenarios, and biasing sentences to guide their enactments of the attitudes (see Appendix A for text of all descriptive materials). To familiarize encoders with the recording procedure, each completed several practice trials before commencing with the recordings proper. The examiner did *not* provide any feedback or coach the encoder as to how a "proper" rendition of the required attitude should be produced. Encoders were allowed to repeat their recordings to their satisfaction, and when this occurred, the final utterance produced for each trial was considered for analysis. All recordings were captured onto digital audio tape in a sound-attenuated booth using a high-quality microphone (sampling rate of 44.1 kHz, 16 bit, mono; no precautions were taken against possible anti-aliasing).

2.1.3. Perceptual validation study

A perceptual validation study was undertaken to establish the reliability of the encoders' productions of specific attitudes prior to acoustic analysis. Sixteen native English speakers (eight males, eight females; mean age in years: 22.47, SD: 2.41; mean years of education: 16.38, SD: 1.71) were recruited from the same population as the encoders to judge the intended meaning of the recorded utterances (hereafter these participants will be referred to as "decoders"). Before testing, decoders were provided the same descriptions of the attitudes that had been given to the encoders. During the task, decoders were presented the full 576 utterances recorded from encoders in 19 blocks of approximately 30 utterances per block. Following presentation of each utterance, each decoder identified the conveyed attitude from among four alternatives (sarcasm, humour, sincerity, and neutrality) in a forced choice task. Utterances were retained as exemplars for acoustic analysis if the rating agreement was two times chance levels or greater (i.e., 50% or greater agreement across all 16 decoders that a particular exemplar reflects a specific attitude). For those stimuli categorized in a way that was not intended by the encoder, the original classification was dropped in favor of the decoders' collective perceptual ratings and that exemplar was recoded. This strategy was pursued to help meet the primary goal of evaluating what acoustic features signal sarcastic speech. There are possible mismatches between intended meaning and perceived meaning in exemplars of sarcastic speech (Rockwell, 2000a). Hence, it was important to ensure acoustic analyses were carried out on utterances that were robustly recognized by a range of decoders as expressing a particular attitude.

Subsequent to the perceptual validation study, we retained 489 utterances as "good" exemplars for acoustic analysis. As a whole, the utterances were representative of the target attitudes and were comparable across phrase types. Results of the validation study are summarized in Table 2.

2.2. Acoustic analyses

All validated utterances were subjected to acoustic analyses using Praat speech analysis software (Boersma and Weenink, 2007). As noted earlier, a number of acoustic parameters were selected for their potential importance for differentiating sarcasm from other attitudes based on previous research findings and trends; the measures obtained from each exemplar were:

(a) Mean *F*0 (in Hz): computed for the utterance as a whole to index the relative effects of pitch height or register on different attitudes.

Table 2
Total number of utterances retained as exemplars for various acoustic analyses subsequent to validation study

Acoustic parameter	Elicited phrase type	Attitude				Total per phrase type
		Sarcasm	Humor	Sincerity	Neutrality	
*F*0, amplitude, speech rate, HNR	Combined sentence	44 (70%)	23 (65%)	52 (68%)	42 (72%)	161
	Single sentence	11 (67%)	22 (64%)	106 (72%)	39 (68%)	178
	Keyphrase	56 (70%)	20 (68%)	41 (72%)	33 (65%)	150
	Total per attitude	111	65	199	114	
						Total per phrase type entered into separate analyses of variance
		Sarcasm	Humor	Sincerity	Neutrality	
1/3 octave Spectral value	Combined sentence	13 (68%)	22 (69%)	35 (68%)	33 (73%)	103
	Single sentence	10 (67%)	15 (67%)	60 (69%)	23 (68%)	108

Note that a grand total of 489 utterances were retained as exemplars for acoustic analysis. An utterance was retained as an exemplar if the rating agreement was 2× chance levels or greater (i.e., 50% or greater agreement across all 16 raters that a particular exemplar reflected a specific attitude). Mean agreement percentages for exemplars are in parentheses next to the total number of utterances retained per factorial combination of PHRASE TYPE and ATTITUDE. Following normalization to neutral exemplars, data from 368 exemplars were entered into statistical analyses for all measures of *F*0, amplitude, speech rate, and HNR.

(b) Standard deviation of *F*0 (in Hz): computed around the mean *F*0 for the whole utterance as an index of overall *F*0 variation.
(c) *F*0 range (in Hz): computed by subtracting the minimum from the maximum *F*0 value of each exemplar as a further index of *F*0 variation.
(d) Mean amplitude (in dB): computed for the utterance as a whole to index the intensity or loudness of utterances spoken with different attitudes.
(e) Amplitude range (in dB): computed by subtracting the minimum amplitude value from the maximum amplitude value of each exemplar to assess the degree of variation in speech intensity or loudness.
(f) Speech rate: computed as the number of syllables in each item divided by the total utterance duration (in milliseconds).
(g) Harmonics-to-noise ratio (HNR, in dB): voice quality measure computed from the 50-ms stable, central portion of vowels which were segmented from the stressed syllables of each utterance (most vowels in unstressed syllables were shorter than 50 ms; Yumoto et al., 1982). The HNR is the ratio of the averaged periodic component of a sound signal to the corresponding averaged noise component and is measured in decibels (dB) (Yumoto et al., 1982). Evaluating HNR serves as a probe for voice quality changes, rather than as a full test of voice shifts, as we only evaluate the HNR of stressed vowels and since the calculation of HNR is a measure of the total noise in the signal rather than a careful evaluation of the contributions of different acoustic perturbations (such as jitter or shimmer, Yumoto et al., 1982).
(h) One-third octave spectral values (in dB): measure of nasality performed on isolated 50-ms stable portions of stressed /i/ vowels by applying a one-third octave filter to 16 frequency regions that range from 125 to 6300 Hz (i.e., encapsulating the spectral region between the *F*0 past the third formant of /i/ vowels). Hence, values (measured in dB) from 16 one-third octave bands were taken to form a spectrum of values. The resulting spectrum was then normalized for statistical analysis; the process of normalization consisted of subtracting the value of each one-third octave band from the value of the one-third octave band containing the *F*0 of the vowel (Kataoka et al., 1996; Yoshida et al., 2000). For male encoders, *F*0 was located in the band centered between 160 and 200 Hz, whereas *F*0 for female encoders could be found in the band centered between 200 and 250 Hz. Successive bands were renumbered sequentially with the band containing the *F*0 relabeled as "0". Elevated nasal resonance in /i/ vowels has been characterized by amplitude increases between first and second formants (i.e., the spectral region defined by bands 6–8) and energy decreases following the second formants (i.e., the spectral region defined by bands 10–14) (Kataoka et al., 1996, 2001a; Lee et al., 2003; Yoshida et al., 2000).

The restricted choice of vowels for one-third octave analyses is justified as /i/ is the only vowel for which reliable indices of nasal resonance have been comprehensively determined using this method of analysis (Kataoka et al., 1996, 2001a; Lee et al., 2003; Yoshida et al., 2000). Other vowels will likely present with different (one-third octave) spectral profiles in the presence of nasality (Beddor, 1993; Maeda, 1993). Also, nasal coupling (and hence, nasal resonance) would be immediately detectable in /i/ over other vowels, given its narrow lingual constriction (Yoshida et al., 2000). As the vowel /i/ only occurred in the sentence portion of exemplars, data considered for analysis was collected from /i/ vowels extracted from the words "she", "he", and "superior (/i/ from /pir/)" of sentence and combined exemplars. It would have been preferable to have included language in keyphrases that contained /i/, but the extensive refinement employed to construct reliable and valid exemplars of target attitudes rendered this impossible. Consequently, we analyzed 108 /i/ vowels extracted

from sentence exemplars and 103 /i/ vowels segmented from combined sentence exemplars (the different number of /i/ vowels available for analysis was due to different numbers of sentence and combined sentence exemplars being retained subsequent to validation; see Table 2).

2.3. Statistical analyses

Prior to statistical analysis, acoustical measures were re-examined by the first author and a trained research assistant. Corrections were made to a small subset of exemplars for which the autocorrelation method produced obvious errors (e.g., when null values were returned by the Praat software due to automated script errors, mean *F*0 was taken using commands available in the graphical user interface instead – this occurred for less than 1% of all exemplars). With the exception of one-third octave spectral values (which were subjected to a previously established normalization procedure as part of calculation), measures obtained for exemplars of sarcasm, humour, and sincerity were each normalized in reference to the same measure obtained for exemplars of neutrality to allow comparisons across speakers and exemplar types. Normalization also corrected for unavoidable differences in microphone distance and instrument recording levels that would have varied somewhat across testing sessions. All individual values of *F*0, amplitude, and speech rate were normalized (per speaker) in reference to the set of neutral exemplars by dividing the difference between the averaged value of neutral exemplars for a given phrase type from an individual data point by the averaged value of neutral exemplars for a given phrase type (e.g. $[\text{mean } F0_{(\text{one keyphrase exemplar of humour})} - \text{mean } F0_{(\text{all keyphrase exemplars of neutrality})}]/\text{mean } F0_{(\text{all keyphrase exemplars of neutrality})}$). In the case of HNR, all HNR values taken from the stressed vowels of a particular exemplar were first averaged prior to normalization in reference to neutral utterances.

The normalized acoustic data (except for one-third octave spectral data) from the 368 remaining exemplars of sarcasm, humour, and sincerity were then subjected to a series of ANOVAs. Each acoustic measure was analyzed for the influences of PHRASE TYPE (keyphrase, sentence, "combined sentence") and ATTITUDE (sarcasm, humour, sincerity) in a repeated-measures design. For one-third octave spectral data, separate one-way ANOVAs involving the factor of ATTITUDE (sarcasm, humour, sincerity, neutrality) were conducted on each normalized octave band following Kataoka et al. (1996). Further differentiating the ANOVAs of the one-third octave spectral data from the statistical analyses of the other acoustic data is the fact that vowels taken from sentence exemplars ($n = 108$) were analyzed separately from combined sentence exemplars ($n = 103$). Data from exemplars of neutrality had to be included in the ANOVAs of one-third octave spectral values as normalization of these values did not eliminate neutrality as a condition. Significant effects and interactions were explored with Tukey's HSD method ($\alpha = 0.05$). Effects that were subsumed by higher-order interactions are not described.

3. Results

Table 3 furnishes the mean (un-normalized) measures of *F*0, amplitude, speech rate and HNR which characterized each attitude.

3.1. Fundamental frequency: mean, standard deviation, range

The impact of the three attitudes on parameters of *F*0 (mean, standard deviation, and range) is illustrated in Fig. 1. The ANOVA performed on values of mean *F*0 yielded a significant main effect for ATTITUDE ($F(2,359) = 16.57$, $p < 0.001$). Post hoc analyses revealed that sarcasm exemplars were significantly lower in mean *F*0 than humour and sincerity exemplars (see Fig. 1a).

The ANOVA on *F*0 standard deviation yielded main effects for PHRASE TYPE ($F(2,359) = 12.86$, $p < 0.001$) and ATTITUDE ($F(2,359) = 17.51$, $p < 0.001$). Elaboration of PHRASE TYPE indicated that keyphrase exemplars exhibited a significantly greater *F*0 standard deviation than both sentence exemplars and combined sentence exemplars. The main effect of ATTITUDE was explained by the fact that sincerity exemplars were produced with greater *F*0 standard deviations than humour and sarcasm exemplars (see Fig. 1b).

Analysis of *F*0 range yielded main effects for PHRASE TYPE ($F(2,359) = 12.71$, $p < 0.001$) and ATTITUDE ($F(2,359) = 6.35$, $p = 0.002$), and an interaction of these factors ($F(4,359) = 3.53$, $p = 0.008$). Post hoc inspection of the interaction showed that the three attitudes were only differentiated by *F*0 range for keyphrases, where sarcasm had a more restricted *F*0 range than sincerity. For sarcasm, *F*0 range did not differ as a function of the three phrase types, whereas for humour and sincerity, *F*0 range tended to be greater for keyphrases than for sentences (and combined sentences in the case of sincerity, see Fig. 1c).[1]

3.2. Amplitude: mean, range

The ANOVA on mean amplitude yielded a main effect of PHRASE TYPE ($F(2,359) = 5.58$, $p = 0.004$). Post hoc tests showed that sentence exemplars were produced with significantly lower amplitude than keyphrase exemplars and combined sentence exemplars. There was no significant

[1] To briefly estimate the extent to which some of our acoustic measures may have been interdependent, a series of three ANCOVAs (analyses of covariance) were conducted to verify the observed effects of PHRASE TYPE and ATTITUDE from the main analysis when potentially related measures were entered as a covariate. Mean *F*0 was reanalyzed with (1) *F*0 standard deviation and (2) *F*0 range as a covariate, and mean amplitude was reanalyzed with amplitude range as a covariate. All significant effects observed in the original analyses remained unchanged, suggesting that these cues operated in a relatively independent manner for communicating sarcasm.

Table 3
Means of non-normalized values of (a) mean fundamental frequency (*F*0), (b) *F*0 standard deviation (SD), (c) *F*0 range (i.e., *f*0maximum − *f*0minimum), (d) mean amplitude, (e) amplitude range (i.e., maximum amplitude − minimum amplitude), (f) speech rate, and (g) harmonics-to-noise ratio (HNR)

Utterance form	Attitude	(a) Mean *F*0		(b) *F*0 SD		(c) *F*0 range	
		Female	Male	Female	Male	Female	Male
Keyphrase	Humor	221.7 (49.9)	148.7 (19.6)	40.8 (33.3)	24.1 (6.0)	125.1 (90.3)	81.7 (21.6)
	Neutrality	203.5 (26.0)	130.1 (10.7)	14.6 (7.1)	11.8 (5.1)	51.7 (23.3)	44.7 (15.3)
	Sarcasm	193.8 (23.2)	136.4 (12.8)	21.2 (9.1)	20.3 (8.6)	76.2 (31.8)	69.3 (26.7)
	Sincerity	224.4 (39.2)	162.6 (28.2)	41.8 (23.2)	30.9 (11.0)	138.0 (78.0)	95.5 (34.2)
Sentence	Humor	209.2 (28.3)	167.5 (24.2)	30.8 (7.0)	24.6 (5.8)	112.5 (22.4)	93.2 (14.4)
	Neutrality	195.7 (20.7)	134.8 (17.0)	19 (15.7)	12.8 (6.9)	85.3 (79.0)	58.8 (30.6)
	Sarcasm	189.1 (15.1)	125.3 (13.1)	23.5 (7.4)	15.9 (5.1)	90.7 (22.6)	56.7 (19.0)
	Sincerity	220.8 (34.2)	145.5 (13.2)	33.9 (19.9)	25.6 (10.3)	117.2 (63.8)	94.6 (36.0)
Combined sentence	Humor	203.4 (12.9)	157.5 (21.2)	32.2 (13.1)	29.1 (7.9)	176.9 (123.2)	105.8 (25.8)
	Neutrality	197.9 (23.8)	136.9 (14.4)	17.0 (9.7)	19.4 (14.9)	93.5 (85.2)	92.2 (72.2)
	Sarcasm	193.0 (16.9)	137.5 (8.7)	28.2 (14.0)	26.3 (13.6)	145.2 (103.7)	129.9 (90.6)
	Sincerity	229.8 (30.1)	149.9 (16.2)	43.4 (17.5)	28.7 (10.9)	175.1 (68.8)	113.2 (48.4)

Utterance form	Attitude	(d) Mean amplitude		(e) Amplitude range		(f) Mean speech rate		(g) Mean HNR	
		Female	Male	Female	Male	Female	Male	Female	Male
Keyphrase	Humor	72.8 (2.6)	72.2 (1.7)	28.3 (7.9)	28.9 (5.6)	3.1 (0.6)	3.6 (0.6)	14.7 (6.1)	16.1 (5.1)
	Neutrality	67.7 (3.9)	67.5 (3.5)	32.7 (4.7)	32.3 (5.2)	3.9 (0.6)	3.4 (0.4)	16.0 (3.7)	14.5 (4.3)
	Sarcasm	70.0 (5.9)	69.8 (4.1)	31.7 (11.5)	31.2 (6.5)	2.8 (0.6)	2.6 (0.5)	16.2 (4.7)	13.7 (4.8)
	Sincerity	73.3 (4.1)	71.1 (3.4)	27.1 (5.1)	30.5 (6.0)	4.0 (0.8)	3.9 (0.7)	14.9 (3.8)	15.6 (3.9)
Sentence	Humor	71.1 (2.9)	72.1 (2.2)	35.2 (11.8)	35.7 (5.0)	5.1 (0.5)	6.0 (0.8)	14.4 (3.0)	17.0 (4.5)
	Neutrality	68.4 (3.0)	67.6 (3.1)	33.7 (4.7)	35.4 (6.9)	5.3 (0.8)	5.0 (0.7)	15.2 (4.5)	14.5 (4.5)
	Sarcasm	70.9 (2.1)	66.4 (3.3)	36.2 (8.0)	36.6 (6.3)	4.8 (0.8)	4.9 (0.7)	13.9 (2.2)	14.7 (2.0)
	Sincerity	69.8 (3.2)	67.8 (2.5)	34.6 (5.8)	34.8 (5.8)	5.5 (0.7)	5.3 (0.6)	15.7 (4.3)	13.1 (4.8)
Combined sentence	Humor	68.7 (3.6)	71.1 (1.5)	43.4 (4.1)	42.0 (4.5)	4.0 (0.8)	5.4 (0.4)	15.2 (4.1)	16.7 (1.9)
	Neutrality	65.6 (4.1)	65.9 (2.8)	41.4 (5.2)	42.3 (7.3)	4.4 (0.5)	4.2 (0.6)	18.0 (3.0)	15.7 (3.8)
	Sarcasm	66.9 (3.6)	67.0 (3.1)	42.2 (6.2)	43.6 (3.9)	4.1 (0.5)	4.2 (0.7)	14.2 (3.6)	13.3 (2.5)
	Sincerity	69.3 (3.7)	65.1 (2.4)	40.3 (5.5)	43.7 (5.0)	4.8 (0.7)	4.6 (0.4)	16.9 (2.9)	15.3 (2.7)

Overall standard deviation values are in parentheses.

main or interactive effect involving the ATTITUDE factor (both p's > 0.14).

The ANOVA performed on amplitude range yielded a main effect of PHRASE TYPE ($F(2,359) = 7.77$, $p < 0.001$). This effect could be explained by the observation that keyphrase exemplars were produced with a reduced amplitude range relative to exemplars of the other two phrase types. Again, there was no significant main effect or interaction involving ATTITUDE (both p's $> .20$).

3.3. Speech rate

The ANOVA on speech rate values yielded main effects for PHRASE TYPE ($F(2,359) = 30.46$, $p < 0.001$) and ATTITUDE ($F(2,359) = 26.19$, $p < 0.001$) and an interaction of these factors ($F(4,359) = 16.57$, $p < 0.001$). Tests performed on the interaction revealed that when producing keyphrase utterances only, sarcasm was expressed at a slower rate than both humour and sincerity, and humour was also produced more slowly than sincerity. When expressing both sarcasm and humour, speakers produced keyphrases significantly slower than combined sentences representing the same attitude (and sentences for sarcasm), although sincerity was expressed at a similar rate across phrase types.

3.4. Harmonics-to-noise ratio (HNR)

The ANOVA on HNR values revealed main effects of PHRASE TYPE ($F(2,359) = 4.16$, $p = 0.02$) and ATTITUDE ($F(2,359) = 3.09$, $p = 0.047$). In the former case, combined sentence exemplars were characterized by lower averaged HNR values relative to the two other phrase types. The main effect of ATTITUDE was explained by the fact that sarcastic messages exhibited a significantly lower averaged HNR value than sincerity.

3.5. One-third octave spectral values

One-third octave spectral data are presented in Fig. 2. One-way ANOVAs performed on normalized amplitude data from /i/ vowels in sentence exemplars revealed significant differences for one-third octave spectral band 2 ($F(3,104) = 3.06$, $p = 0.03$), band 3 ($F(3,104) = 2.93$, $p = 0.04$), band 10 ($F(3,104) = 4.05$, $p = 0.009$), band 11 ($F(3,104) = 3.65$, $p = 0.02$), band 13 ($F(3,104) = 2.80$, $p = 0.04$), and band 14 ($F(3,104) = 3.61$, $p = 0.02$). The significant difference found for analysis of band 2 in the sentence exemplars can be accounted for by the greater amplitude of /i/ vowels in sarcasm tokens relative to /i/ vowels in sincerity tokens. Differences found at band 3

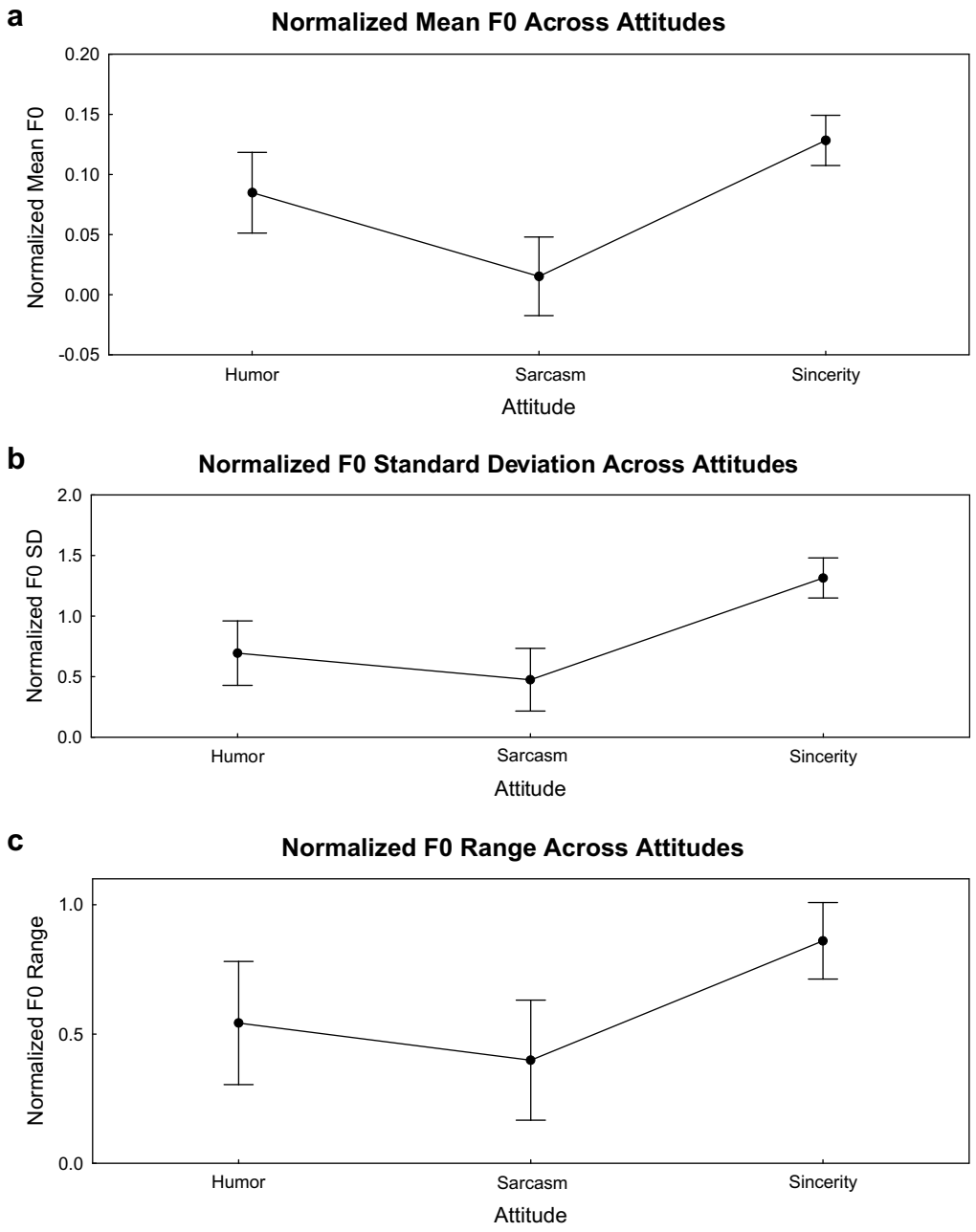

Fig. 1. Various speaker manipulations of fundamental frequency ($F0$) across exemplars of humor, sarcasm, and sincerity. (a) Normalized mean $F0$ values across attitudes; (b) normalized $F0$ standard deviation values across attitudes; (c) normalized $F0$ range values across attitudes.

are explained by /i/ vowels from humour tokens having greater amplitude than /i/ vowels in sincerity tokens. Greater amplitude of /i/ vowels in sarcasm tokens relative to sincerity and neutral tokens accounted for the significant differences found in analyses of bands 10, 11, and 14. Taken as a whole, sentence exemplars of sarcasm were characterized by a resonance profile different from the other attitudes.

For /i/ vowels extracted from combined sentence exemplars, statistical analysis of one-third octave spectral bands yielded significant differences in normalized amplitude for the majority of bands. Specifically, differences in attitude tokens were found in band 1 ($F(3,99) = 3.57$, $p = 0.02$), band 2 ($F(3,99) = 3.83$, $p = 0.01$), band 3 ($F(3,99) = 7.84$, $p < 0.001$), band 4 ($F(3,99) = 2.86$, $p = 0.04$), band 5 ($F(3,99) = 3.69$, $p = 0.01$), band 6 ($F(3,99) = 3.18$, $p = 0.03$), band 7 ($F(3,99) = 3.09$, $p = 0.03$), band 11 ($F(3,99) = 5.28$, $p = 0.002$), band 12 ($F(3,99) = 4.68$, $p = 0.004$), band 13 ($F(3,99) = 5.72$, $p = 0.001$), and band 14 ($F(3,99) = 3.87$, $p = 0.01$). A number of significant differences can be attributed to the greater amplitude of /i/ vowels in sarcasm tokens relative to sincerity exemplars (bands 1, 2, 4, 6, and 7). Analysis of band 3 showed two patterns: /i/ vowels in sarcasm tokens were greater in

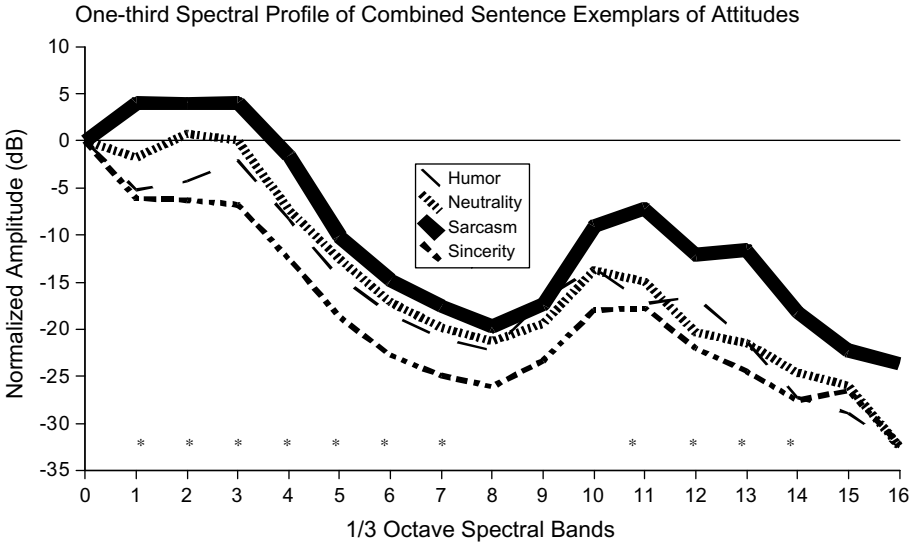

Fig. 2. Normalized one-third octave spectral values (in dB) for combined sentence exemplars grouped by attitude. Note the clear difference in one-third octave spectral profile of sarcasm relative to exemplars of the other attitudes. * denote bands for which significant differences in normalized amplitudes were found across attitude exemplars.

amplitude than for those in sincerity and neutrality tokens, and /i/ in neutrality tokens were higher in amplitude than /i/ in sincerity tokens. Similarly, in band 5, both sarcasm and neutrality tokens were produced with greater amplitude than /i/ vowels in sincerity tokens. Resonance differences most distinguished the attitudes in bands 11 and 13; sarcasm was greater in amplitude than the other three attitudes. Finally, analysis of bands 12 and 14 showed that /i/ vowels in sarcasm tokens were produced with greater amplitude than those in sincerity and neutrality tokens or sincerity and humour tokens, respectively. Overall, analysis of combined sentence tokens showed a more pronounced pattern of resonance differences across attitudes than sentence exemplars, with the profile for sarcasm being the most distinct. Fig. 2 summarizes acoustic and statistical findings for one-third octave spectral analyses for the combined sentence exemplars.[2]

[2] By normalizing our measures in reference to neutrality exemplars, our data do not permit insight into how neutral utterances compared to sarcasm, humour, and sincerity for our prosodic measures. To illuminate this briefly, all data were re-normalized in the following manner: all individual values of *F*0, amplitude, and speech rate, and the HNR data points derived for individual exemplars were standardized per speaker in reference to his or her entire set of exemplars 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 $F0_{(\text{one exemplar of humour})}$ − mean $F0_{(\text{all exemplars})}$]/$F0$ $SD_{(\text{all exemplars})}$). The ANOVA on each acoustic measure was then rerun with four levels of ATTITUDE (including neutrality). The newly-computed statistical results were consistent with the results of our main analyses and will not be reported. Rather, significant attitude-neutrality differences are summarized as follows: sarcasm was characterized by greater *F*0 SD, greater *F*0 range, greater amplitude and slower speech rate than neutrality; humour had greater mean *F*0, greater *F*0 SD, greater *F*0 range, greater amplitude, and slower speech rate than neutrality; sincerity had greater mean *F*0, greater *F*0 SD, greater *F*0 range, greater amplitude, and faster speech rate than neutrality.

4. Discussion

The present study sought to shed light on the prosodic markers of sarcasm and to dissociate these cues, if possible, from the effects of semantic cues by analyzing utterances which did or did not contain keyphrases associated with sarcasm. Acoustic changes that varied as a function of attitude regardless of phrase type, therefore, could be interpreted as prosodic cues which are most important to the expression of sarcasm (in reference to other attitudes in the present study). Based on this assumption, our results showed that overall reductions in mean *F*0, decreases in *F*0 variation (standard deviation), and changes in HNR (i.e., voice quality) were used most consistently by English speakers to differentiate sarcasm from other attitudes. These more central cues often coincided with further changes in resonance and reductions in speech rate which occurred in many, but not all, linguistic contexts as elaborated below.

4.1. The most prominent cues for signalling sarcasm

A reduction in mean *F*0 was the most consistently observed prosodic correlate of sarcasm in the present study; this cue distinguished sarcasm from both humour and sincerity regardless of phrase type (see Table 3). This result is unsurprising given that adopting a monotonic or lower *F*0 or pitch in speech are often cited acoustic cues of sarcasm (Attardo et al., 2003; Fonagy, 1971; Haiman, 1998; Rockwell, 2000a). A tendency to raise or lower voice pitch is known to be instrumental for signaling other attitudinal and affective states as well (Banse and Scherer, 1996; Bänziger and Scherer, 2005), emphasizing the signif-

icance of this cue for marking sarcasm and perhaps other interpersonal intentions during speech.

In addition to mean *F*0, sarcastic exemplars displayed smaller *F*0 standard deviation than exemplars of sincerity overall; moreover, *F*0 range was similarly reduced for sarcastic relative to sincere utterances although only for keyphrase exemplar types. Our findings for these (overlapping) measures are supported in part by descriptions of sarcasm as exhibiting periods of reduced *F*0 or pitch movements (Attardo et al., 2003; Fonagy, 1971, 1976). However, much of the literature suggests that sarcasm is marked by heightened *F*0/pitch variability as one of its major prosodic cues in contrast to our findings (Attardo et al., 2003; Cutler, 1974, 1976; Haiman, 1998). Pending further research, one can argue that changes in the extent of *F*0 variation produced by speakers is a relatively consistent feature of sarcastic speech, although the direction of these changes is not always uniform and/or this cue may interact more extensively with the nature of the language content (Bryant, in press).

Speakers also manipulated voice quality to convey sarcasm; specifically, we found that exemplars of sarcasm were differentiated from exemplars of sincerity by greater amounts of noise (as inferred from lower HNR values in sarcasm exemplars). As analysis of HNR in sarcastic speech is a novel approach, we need to draw from studies that have looked at the importance of voice quality manipulations on listener perceptions of affect or mood to interpret our findings (Gobl and Ni Chaisaide, 2003; Ladd et al., 1985; Banse and Scherer, 1996; Scherer, 1986; Scherer et al., 1984). First, it has been noted that differences in voice quality are mostly associated with how listeners perceive a speaker's mood and attitudes rather than their emotions, with the exception of anger (Gobl and Ni Chaisaide, 2003). In particular, "harsh" or "tense" voice quality tends to signal anger or negative states (e.g., "stressed", "hostile") to listeners (Gobl and Ni Chaisaide, 2003), and there is evidence that listeners classify utterances with a harsh or tense voice quality as conveying negative moods such as annoyance, irritation, and hostility (Ladd et al., 1985). These negative attributions may stem from a cluster of predictable physiological changes which occur in response to negative stimuli, such as heightened tension of the respiratory system and vocal apparatus as well as decreased salivation which are likely to produce a tense voice quality (Scherer, 1986).

As sarcasm is associated with critical (i.e., negative) attitudes on the part of the speaker, it is therefore noteworthy that shifts in voice quality in our sarcasm exemplars were characterized by increased noise (i.e., decreases in HNR). Not only does this finding confirm the importance of voice quality changes for expressing attitudes, it implies that the inherently negative nature of sarcasm may be encoded and expressed through similar acoustic changes as other negative messages, i.e., by increasing the level of noise in the voice. One can speculate that listeners become attuned to such shifts in voice quality as one of the major vocal cues for identifying sarcastic intent (Fonagy, 1971; Haiman, 1998; Mueke, 1969). However, further perceptual and acoustic analyses are necessary to confirm this possibility.

4.2. Effects of secondary cues and phrase type

There were other acoustic markers of sarcasm which tended to be less pervasive across linguistic contexts and may be more dependent on phrase type. Principally, we found that a reduced speech rate and different resonance patterns were associated with sarcasm, although these cues were prevalent for specific phrase types. In the case of speech rate, short keyphrase utterances such as "is that so" tended to be spoken more slowly than both "sentence" and "combined" phrase types overall, but interestingly, these keyphrases were produced significantly slower when the speaker intended to be sarcastic than either humourous or sincere. Similarly, exemplars of humour tended to be much slower when speakers produced keyphrases as opposed to combined sentences. These findings suggest that listeners are sensitive to (reduced) speech rate as a potential cue to the two different subtypes of verbal irony studied in the current investigation, albeit seemingly in an interactive fashion with phrase type: the smaller the utterance form, the more listeners appeared to respond to reductions in speech rate as an indicator of verbal irony (regardless of the presumed hostility or friendliness of sarcasm and humour, respectively). Speech rate was not a differentiating factor in how listeners identified attitudes for longer phrase types (e.g., combined sentences). Thus, our results extend previous data which have characterized sarcastic speech as having a reduced speech rate (Anolli et al., 2002; Bryant, in press; Cutler, 1974; Haiman, 1998; Mueke, 1978; Rockwell, 2000b) but further clarify that speech rate differences likely vary with language usage and may be more critical for signaling irony in short utterance types.

With respect to resonance, the one-third octave spectral data indicate that /i/ vowels extracted from both sentence and combined sentence exemplars of sarcasm were produced with significantly greater amplitude at nearly all crucial frequencies. Sarcasm differed most consistently from sincerity in sentence exemplars and presented a markedly different resonance profile from all other attitudes in combined sentence exemplars. Recall that previous investigations have discerned increases in amplitude in spectral regions near the first formant concurrently with decreases in amplitude in spectral regions near the second formant as a marker of heightened nasal emission (Kataoka et al., 1996, 2001a; Lee et al., 2003; Yoshida et al., 2000). It is important to note, however, the differing resonance patterns observed here are not consistent with patterns that denote heightened levels of nasality; the observed spectral profiles did not correspond with previously-established hypernasal profiles (i.e., energy increases in bands 6–8 and energy decreases in bands 10–14) (Kataoka et al., 1996, 2001a; Lee et al., 2003; Yoshida et al., 2000).

Rather, we found amplitude increases in nearly every band indicating that exemplars of sarcasm were produced with a different, but not nasal, resonance (and only for sentence and combined sentence phrase types).

One explanation for these results is that nasalization may have occurred within sarcasm exemplars in areas that were not analyzed (i.e., vowels other than /i/), but this appears unlikely. Given how /i/ is articulated (i.e., narrow lingual constriction), and the fact that only /i/ in stressed syllables was examined, any nasal emissions should have been immediately apparent in the present one-third octave spectral analyses. Previous claims that nasal resonance was associated with sarcastic expression were not based on objective measures (Cutler, 1974, 1976; Haiman, 1998) and many factors are known to impact on listeners' detection of nasality in speech (see Baken and Orlikoff, 2000, for an overview). It is therefore possible that previous assertions that sarcasm is associated with increased nasality stem from *perceived* differences in nasal resonance due to the patterns observed here.

Another possibility is that orofacial expressions of disgust, which are characterized by a facial sneer or palatal drop, occur in tandem with sarcastic speech and lead to these resonance changes. This idea suggests that speakers mimic disgust as a critical or supplementary means of projecting sarcasm to listeners (Cutler, 1974; Fonagy, 1971; Haiman, 1998), although comparisons between disgust and sarcasm are problematic since the vocal and orofacial gestures associated with disgust are known to be highly variable in speech and difficult to recognize by listeners (Pell et al., in review). A different way that facial expressions could affect resonance relates to the finding that sarcasm is signalled by exaggerated facial gestures, particularly affecting the mouth, or that speakers adopt a "blank face" to cue pending sarcasm (i.e., a purposefully inexpressive or motionless facial expression, Attardo et al., 2003; Cutler, 1974; Fonagy, 1976; Mueke, 1969; Rockwell, 2001, 2005). Observed changes in resonance may have been associated with such changes in facial expressions, especially mouth and lip movements (Tartter and Braun, 1994) or tongue contractions (Fonagy, 1971). Future refinements in methodology could include video recording and analysis of speakers' facial, lip, or tongue movements when producing sarcasm and other attitudes which could then be compared to spectral measures.

Finally, in contrast to some previous claims (Anolli et al., 2002; Bryant and Fox Tree, 2005; Rockwell, 2000a, 2005), no differences in amplitude or amplitude variability distinguished sarcasm from the other attitudes here. Since data on this measure are generally restricted to *perceptual* impressions of loudness (Rockwell, 2000a, 2005), further acoustic data will be needed to reconcile our findings since the relationship between objective acoustic features and their perceptual correlates is not precise (Sobin and Alpert, 1999; Williams and Stevens, 1972). The possibility that loudness as a prosodic feature of sarcasm varies across languages also cannot be discounted; our data on English speakers may differ from those of Anolli et al. (2002) who found that Italian speakers produced sarcasm with increased loudness. A cross-linguistic comparison would be useful in determining whether amplitude or other acoustic parameters associated with sarcasm vary significantly by language, which we are now undertaking (Cheang and Pell, in preparation).

4.3. On the "ironic tone of voice"

Researchers agree that prosody is instrumental for communicating sarcasm in speech, but there are different opinions regarding the acoustic cues associated with sarcastic intent and the manner in which linguistic context interacts with prosody to project a sarcastic message. Overall, our data show that sarcastic messages in English possess reliable acoustic features (characteristic changes in *F*0 and voice quality) which are produced by speakers irrespective of the linguistic context of these expressions. These findings are most consistent with the idea of an ironic tone of voice (Clark and Gerrig, 1984; Mueke, 1969), or more precisely, a sarcastic tone of voice (i.e, the existence of defining prosodic features which are used to communicate sarcasm in speech).

At the same time, the interaction of prosody with language and context is also emphasized by our data (Cutler, 1974; Haiman, 1998); here, speakers appeared to provide a set of "supplementary" acoustic cues when producing short, keyphrase utterances. One can speculate that speakers provide an enriched prosodic signal in this (or other) language contexts to ensure that shorter excerpts of speech are unambiguously treated as sarcastic by the listener when semantic features of an utterance are less indicative of this intent. In this way, expressing sarcasm may be similar to other attitudes in that certain prosodic cues may be salient *independent* of semantic information while other prosodic features work in conjunction with the language content (Scherer et al., 1984). If one accepts that prosodic cues operate both independently and interactively with other cues, the existence of a specific sarcastic prosody can be reconciled with claims that verbal irony is marked by a variety of communicative behaviours (Attardo et al., 2003; Bryant, in press; Bryant and Fox Tree, 2005).

Our data also furnish clues about the acoustic properties of sincerity and humour. Prominent acoustic differences were most apparent between sarcasm and sincerity rather than humour; measures of mean *F*0, *F*0 standard deviation, and HNR clearly distinguished these two attitudes regardless of linguistic context (sincere utterances displayed a higher mean *F*0, greater *F*0 standard deviation, and less noise in the signal). This result should come as little surprise given the opposing speaker intentions underlying the articulation of sarcasm (indirect, semantically-inconsistent criticism) and sincerity (direct literal appraisal) which are shown here to be distinguished by multiple acoustic cues in speech. In the case of humour, we found no evidence of a highly contrastive prosodic pattern from the

other attitudes, in line with claims that humourous intent is normally communicated through a combination of factors or cues. For example, some explicit play cue must alert the listener to preclude seriousness (Berlyne, 1972; Suls, 1983), and once a play context is established, the meaning of humourous utterances can be determined when expectations generated by the speaker are violated in a variety of ways (Suls, 1983, 1972; Brownell et al., 1983). In light of these factors, it may have been difficult to capture how prosody is used to convey humour with our approach, and it is also likely that speakers had greater difficulty to convey humourous intent in single, posed utterances as suggested by the relatively low number of perceptually-valid humour exemplars retained from the validation study (review Table 2). Nonetheless, exemplars of sarcasm and humour were found to be distinct on certain acoustic measures and it would be prudent of future researchers to evaluate different subtypes of verbal irony separately rather than as a uniform class (Anolli et al., 2002; Bryant, in press).

Finally, our inclusion of "enantiosemantic" phrases in the text of our tokens merits some commentary. Based on results of our validation study, it is readily apparent that there was an uneven distribution in the number of exemplars perceived as conveying each attitude according to phrase type; the presence of keyphrases tended to facilitate perception of sarcasm while their absence facilitated perception of sincerity (review Table 2). This pattern of findings is consistent with previous (empirically untested) claims that certain phrases become associated with sarcastic speaker intent (Haiman, 1998), again underscoring the significant role of lexical-semantic information in the expression of sarcasm. Nonetheless, it bears re-emphasizing that attitudes other than sarcasm could be conveyed and recognized from each keyphrase in our validation study; also, for those utterances which were retained, speakers executed changes in *F*0 and voice quality to mark sarcasm irrespective of whether particular keyphrases were present or absent, highlighting the potential independence of these prosodic features as a central property of sarcastic speech.

4.4. Limitations and future directions

The present study analyzed non-spontaneous ("posed") speech, and while great care was taken to ensure that adequate exemplars of each attitude were subjected to analysis, the literature would benefit from future studies which investigate the production and comprehension of sarcasm in both controlled and spontaneous situations (cf. Bryant and Fox Tree, 2005; Rockwell, 2000b). Nonetheless, we do not believe that differences in stimulus acquisition would affect our main conclusions about the acoustic form of sarcasm since previous perception studies have shown that listeners can recognize posed or spontaneous sarcasm (Bryant and Fox Tree, 2005; Rockwell, 2000b). More likely, our data represent instances in which speakers successfully simulated coded prosodic features for expressing sarcasm which resemble those used in spontaneous contexts and which were decoded by listeners in our validation study.

One way to expand current knowledge regarding the role of prosody for communicating sarcasm and other attitudes is to study the effects of brain damage on these crucial abilities. Patients with right hemisphere damage often display impairments in producing and comprehending prosody in affective and attitudinal contexts (e.g. Heilman et al., 1984; Tompkins, 1991). Recently, we have shown comparable difficulties to communicate emotions and 'prosodic attitudes' such as confidence or politeness in patients with Parkinson's disease (Cheang and Pell, 2007; Pell and Leonard, 2003; Pell et al., 2006). Deriving a better understanding of fundamental acoustic markers of attitudes in speech and their biasing linguistic and contextual factors would be germane to improve diagnosis and treatment of many "pragmatic language disorders" in these patients.

To conclude, the present analyses strongly suggest that there is a distinct pattern of acoustic cues associated with sarcasm in speech, one that diverges most clearly from expressions of sincerity. One cluster of cues, reduced *F*0 and HNR and decreased *F*0 standard deviation, robustly marked sarcastic utterances, whereas changes in resonance and reductions in speech rate appeared to be supplementary cues for marking sarcasm in certain linguistic contrasts. While this pattern of acoustic cues does not precisely match those observed for other attitudes or affective states (Banse and Scherer, 1996; Mullenix, 2005), more research is needed to clarify the extent to which the "sound of sarcasm" is uniquely specified in speech.

Acknowledgements

This research was conducted as part of the first author's doctoral research and was supported in part by a CIHR-K.M. Hunter Doctoral Training Award. Further funding was provided by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada to the second author. The authors would also like to thank Paul Boersma, Valter Ciocca, and Alice Lee for help in devising PRAAT scripts to conduct one-third octave analyses, and Marie Desmarteau and Andrea MacLeod for their assistance in the preparation of the manuscript.

Appendix A

A.1. Summary of attitude descriptions given to all participants

A.1.1. Sarcasm

Definition: A sharp, bitter, or cutting expression or remark; a bitter gibe or taunt.

Description: There are two particular circumstances that we would like to draw your attention to with respect to sarcastic utterances. We want you to note that sarcasm is often

used by people as a retort to an insult. Sarcasm is also used by people to criticize situations or other people who they find unpleasant. The common theme that runs in these two situations is that there should be an element of malice. People may use sarcasm under other conditions, but the situations just described are what we would like you to keep in mind.

A.1.2. Humor

Definition: Something (action, speech, or writing) that is ludicrous or absurdly incongruous or designed to be comical or amusing.

Description: We would like you to keep two situations in mind regarding humorous, playful utterances. One circumstance is that people attempt to make humorous comments about unexpected but amusing or unthreatening events. Humorous banter might also occur when people who have friendly relations find themselves in happy and non-serious interactions. The key element with such utterances is that they are made under happy or potentially silly situations. There are other circumstances under which people may make humorous utterances, but these are the ones that we would like you to keep in mind.

A.1.3. Sincerity

Definition: Free from pretence or deceit; genuine, honest, frank.

Description: With respect to sincere utterances, we are referring to situations where people are expressing what they are honestly feeling or thinking on a subject for which they have no personal connection or stake. That is, speakers mean exactly what they are saying; they are not trying to project any hidden messages or intentions. Sincere utterances are different from neutral utterances.

A.1.4. Neutrality

Definition: Displaying or containing no overt emotion; dispassionate, detached.

Description: Neutral utterances do not convey any emotion or mood. That is, neither attitude nor intention is expressed. Sincere utterances and neutral utterances are two wholly different things.

A.2. Instructions specific to encoders

> "We are trying to understand how people communicate their moods verbally. We would like you to act out a number of sentences with certain moods so that we may record and analyze them.
>
> You will be shown pairs of sentences. Please read both sentences silently first. Imagine that the first sentence is directed at you from another speaker and that the second sentence in the pair is your response. I want you to act out the second sentence. Importantly, I want you to say the sentence in a way that you feel is as natural sounding as possible; do not be theatrical. Essentially, say it the way you would if you were in that situation.
>
> There are very specific moods that we would like you to enact depending on the sentence. We will tell you which mood we would like you to enact before we present you with the sentences. The first sentence in the sentence pair will always be designed to help evoke the mood that we would like you to act out.
>
> [Then the participant was provided with a description of an attitude prior to the recording of that particular attitude; see above]."

A.3. Instructions specific to decoders

> "You will be hearing sets of about 30 sentences in a row. Each set of sentences will be followed by a short break. For every sentence, the speaker is trying to express either a sarcastic, humorous, or sincere attitude. Alternately, the speaker is attempting to express no attitude at all, that is, the sentence sounds neutral. Please note that you will be hearing many different voices and that not all speakers will sound the same.
>
> [Then the participant was provided with descriptions of all attitudes; see above].
>
> While you listen to each sentence, you will also see the names of these attitudes on the computer screen. Your task is to judge which attitude the speaker is trying to express. It is important that you try to listen to how each speaker sounds rather than paying specific attention to the words being spoken. After you have heard the whole sentence, choose the attitude that is closest to what you hear by clicking the label on the computer screen with your mouse. Please do not overly rush an answer; the computer will not move on to the next sentence if you try to answer before a sentence has finished playing. We will begin with a short practice session. Ready?"

References

Ackerman, B., 1983. Form and function in children's understanding of ironic utterances. Journal of Experimental Child Psychology 35, 487–508.

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

Anolli, L., Ciceri, R., Infantino, M.G., 2002. From "blame by praise" to "praise by blame": analysis of vocal patterns in ironic communication. International Journal of Psychology 37 (5), 266–276.

Attardo, S., Eisterhold, J., Hay, J., Poggi, I., 2003. Multimodal markers of irony and sarcasm. Humor: International Journal of Humor Research 16 (2), 243–260.

Baken, R.J., Orlikoff, R.F., 2000. Clinical Measurement of Speech & Voice, second ed. Singular Publishing Group Inc., New York.

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

Bänziger, T., Scherer, K.R., 2005. The role of intonation in emotional expressions. Speech Communication 46, 252–267.

Beddor, P.S., 1993. The perception of nasal vowels. In: Huffman, M., Krakow, R. (Eds.), Nasals, Nasalization, and the Velum. Academic Press, San Diego, pp. 171–196.

Berlyne, D.E., 1972. Humor and its kin. In: McGhee, P.E., Goldstein, J.H. (Eds.), The Psychology of Humor. Springer-Verlag, New York, pp. 43–60.

Boersma, P., Weenink, D., 2007. Praat: doing phonetics by computer (Version 5.0.02) [Computer program]. Retrieved 27 December 2007. Available from: <http://www.praat.org/>.

Bolinger, D., 1989. Intonation and Its Uses: Melody in Grammar and Discourse. Stanford University Press, Stanford, CA.

Brownell, H.H., Michel, D., Powelson, J., Gardner, H., 1983. Surprise but not coherence: sensitivity to verbal humor in right-hemisphere patients. Brain and Language 18, 20–27.

Bryant, G.A., Fox Tree, J.E., 2002. Recognizing verbal irony in spontaneous speech. Metaphor and Symbolic Activity 17 (2), 99–117.

Bryant, G.A., Fox Tree, J.E., 2005. Is there an ironic tone of voice? Language and Speech 48 (3), 257–277.

Bryant, G.A., in press. Prosodic contrasts in ironic speech. Discourse Processes.

Capelli, C.A., Nakagawa, N., Madden, C.M., 1990. How children understand sarcasm: the role of context and intonation. Child Development 61, 1824–1841.

Cheang, H.S., Pell, M.D., 2007. An acoustic investigation of Parkinsonian speech in linguistic and emotional contexts. Journal of Neurolinguistics 20 (3), 221–241.

Cheang, H.S., Pell, M.D., in preparation. The acoustics of sarcasm in Cantonese and English.

Clark, H.H., Gerrig, R.J., 1984. On the pretense theory of irony. Journal of Experimental Psychology: General 113 (1), 121–126.

Colston, H.L., 1997. I've never heard anything like it: overstatement, understatement, and irony. Metaphor and Symbol 12 (1), 43–58.

Colston, H.L., O'Brien, J., 2000a. Contrast and pragmatics in figurative language: anything understatement can do, irony can do better. Journal of Pragmatics 32, 1557–1583.

Colston, H.L., O'Brien, J., 2000b. Contrast of kind versus contrast of magnitude: the pragmatic accomplishments of irony and hyperbole. Discourse Processes 30 (2), 179–199.

Cutler, A., 1974. On saying what you mean without meaning what you say. In: LaGaly, M.W., Fox, R.A., Bruck, A. (Eds.). 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: Wales, R.J., Walker, E. (Eds.). Dik, S.C., Kooij, J.G. (Eds.), North-Holland Linguistic Series, Vol. 30. North-Holland Publishing Company, Amsterdam, pp. 133–149.

Dara, C., Monetta, L., Pell, M.D., in press. Vocal emotion processing in Parkinson's disease: reduced sensitivity to negative emotions. Neuropsychology.

de Krom, G., 1995. Some spectral correlates of pathological breathy and rough voice quality for different types of vowel fragments. Journal of Speech and Hearing Research 38 (4), 794–811.

Dews, S., Kaplan, J., Winner, E., 1995. Why not say it directly? The social functions of irony. Discourse Processes 19 (3), 347–367.

Eskenazi, L., Childers, D.G., Hicks, D.M., 1990. Acoustic correlates of vocal quality. Journal of Speech and Hearing Research 33 (2), 298–306.

Ferrand, C.T., 2002. Harmonics-to-noise ratio: an index of vocal aging. Journal of Voice 16 (4), 480–487.

Fonagy, I., 1971. Synthèse de l'ironie. Phonetica 23, 42–51.

Fonagy, I., 1976. Spoken language: dynamic and changing. Journal de Psychologie Normale et Pathologique 73 (3–4), 273–304.

Gerrig, R.J., Goldvarg, Y., 2000. Additive effects in the perception of sarcasm: situational disparity and echoic mention. Metaphor and Symbol 15 (4), 197–208.

Gibbs Jr., R.W., 2000. Irony in talk among friends. Metaphor and Symbol 15 (1–2), 5–27.

Gobl, C., Ni Chaisaide, A., 2003. The role of voice quality in communicating emotion, mood and attitude. Speech Communication 40, 189–212.

Grice, H.P., 1975. Logic and conversation. In: Cole, P., Morgan, J.L. (Eds.), Speech Acts, Vol. (3). Syntax and Semantics, pp. 41–58.

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

Haverkate, H., 1990. A speech act analysis of irony. Journal of Pragmatics 14, 77–109.

Heilman, K.M., Bowers, D., Speedie, L., Branch Coslett, H., 1984. Comprehension of affective and nonaffective prosody. Neurology 34, 917.

Jorgensen, J., 1996. The functions of sarcastic irony in speech. Journal of Pragmatics 26, 613–634.

Kataoka, R., Michi, K., Okabe, K., Miura, T., Yoshida, H., 1996. Spectral properties and quantitative evaluation of hypernasality in vowels. Cleft Palate: Craniofacial Journal 33, 43–50.

Kataoka, R., Warren, D.W., Zajac, D.J., Mayo, R., Lutz, R.W., 2001a. The relationship between spectral characteristics and perceived hypernasality in children. Journal of the Acoustical Society of America 109, 2181–2189.

Kataoka, R., Zajac, D.J., Mayo, R., Lutz, R.W., Warren, D.W., 2001b. The influence of acoustic and perceptual factors on perceived hypernasality in the vowel. Folia Phoniatrica et Logopaedica 53, 198–212.

Kreuz, R.J., Glucksberg, S., 1989. How to be sarcastic: the echoic reminder theory of verbal irony. Journal of Experimental Psychology: General 118 (4), 374–386.

Kreuz, R.J., Roberts, R.M., 1993. On satire and prosody: the importance of being ironic. Metaphor and Symbolic Activity 8 (2), 97–109.

Kreuz, R.J., Roberts, R.M., 1995. Two cues for verbal irony: hyperbole and the ironic tone of voice. Metaphor and Symbolic Activity 10 (1), 21–31.

Kreuz, R.J., Long, D.L., Church, M.B., 1991. On being ironic: pragmatic and mnemonic implications. Metaphor and Symbolic Activity 6, 149–162.

Kumon-Nakamura, S., Glucksberg, S., Brown, M., 1995. How about another piece of pie: the allusional pretense theory of discourse irony. Journal of Experimental Psychology: General 124 (1), 3–21.

Ladd, D.R., Silverman, K.E., Tolkmitt, F.J., Bergmann, G., Scherer, K.R., 1985. Evidence for the independent function of intonation contour type, voice quality, and *F*0 range in signaling speaker affect. Journal of the Acoustical Society of America 78 (2), 435–444.

Laval, V., 2004. Pragmatique et langage non littéral: la compréhension des demandes sarcastiques par les enfants. Psychologie Française 49, 177–192.

Laval, V., 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, A.S.-Y., Ciocca, V., Whitehill, T.L., 2003. Acoustic correlates of hypernasality. Clinical Linguistics and Phonetics 17 (4–5), 259–264.

Lee, S.Y.A., Ciocca, V., Whitehill, T., 2004. Spectral analysis of hypernasality. Journal of Medical Speech-Language Pathology 12, 173–177.

Maeda, S., 1993. Acoustics of vowel nasalization and articulatory shifts in French nasal vowels. In: Huffman, M., Krakow, R. (Eds.), Nasals, Nasalization, and the Velum. Academic Press, San Diego, pp. 147–167.

Monetta, L., Cheang, H.S., Pell, M.D., in press. Understanding speaker attitudes from prosody by adults with Parkinson's disease. Journal of Neuropsychology.

Mueke, D.C., 1969. The Compass of Irony. Methuen, London.

Mueke, D.C., 1978. Irony markers. Poetics 7, 363–375.

Mullenix, J., 2005. The perceptual processing of vocal emotion. In: Clark, A. (Ed.), Psychology of Moods. Nova Science Publishers, Hauppage, NY, pp. 123–139.

Pell, M.D., 2006. Cerebral mechanisms for understanding emotional prosody in speech. Brain and Language 96 (2), 221–234.

Pell, M.D., Leonard, C.L., 2003. Processing emotional tone from speech in Parkinson's disease: a role for the basal ganglia. Cognitive, Affective and Behavioral Neuroscience 3, 275–288.

Pell, M.D., Cheang, H.S., Leonard, C.L., 2006. The impact of Parkinson's disease on vocal-prosodic communication from the perspective of listeners. Brain Language 97, 123–134.

Pell, M.D., Alasseri, A., Paulmann, S., Kotz, S.A., in review. Factors in the recognition of vocally expressed emotions: a comparison of three languages.

Pereira, J.G., Cervantes, O., Abrahao, M., Parente Settanni, F.A., Carrara de, A.E., 2002. Noise-to-harmonics ratio as an acoustic measure of voice disorders in boys. Journal of Voice 16, 28–31.

Pexman, P.M., Olineck, K.M., 2002. Understanding irony: how do stereotypes cue speaker intent? Journal of Language and Social Psychology 21, 245–274.

Rockwell, P., 2000a. Actors, partners, and observers perceptions of sarcasm. Perceptual and Motor Skills 91, 665–668.

Rockwell, P., 2000b. Lower, slower, louder: vocal cues of sarcasm. Journal of Psycholinguistic Research 29, 483–495.

Rockwell, P., 2001. Facial expression and sarcasm. Perceptual and Motor Skills 93, 47–50.

Rockwell, P., 2005. Sarcasm on television talk shows: determining speaker intent through verbal and nonverbal cues. In: Clark, A. (Ed.), Psychology of Moods. Nova Science Publishers Inc., New York, pp. 109–140.

Sabbagh, M.A., 1999. Communicative intentions and language: evidence from right-hemisphere damage and autism. Brain Language 70, 29–69.

Schaffer, R., 1982. Are there consistent vocal clues for irony? In: Masek, C.S., Hendrick, R.A., Miller, M.F. (Eds.), Parasession on Language and Behavior. Chicago Linguistic Society, Chicago, IL, pp. 204–210.

Scherer, K.R., 1986. Vocal affect expression: a review and a model for future research. Psychological Bulletien 99, 143–165.

Scherer, K.R., Ladd, D.R., Silverman, K.E., 1984. Vocal cues to speaker affect: testing two models. Journal of the Acoustical Society of America 76 (5), 1346–1356.

Sobin, C., Alpert, M., 1999. Emotion in speech: The acoustic attributes of fear, anger, sadness, and joy. Journal of Psycholinguistic Research 28 (4), 1999.

Sperber, D., 1984. Verbal irony: Pretense or echoic mention? Journal of Experimental Psychology: General 113 (1), 130.

Suls, J., 1972. A two-stage model for the appreciation of jokes and cartoons: An information-processing analysis. In: Goldstein, J.H., McGhee, P.E. (Eds.), The Psychology of Humor. Academic Press, New York, pp. 81–100.

Suls, J., 1983. Cognitive processes in humor appreciation. In: McGhee, P.E., Goldstein, J.H. (Eds.), Handbook of Humor Research. Springer-Verlag, New York, pp. 39–57.

Tartter, V.C., Braun, D., 1994. Hearing smiles and frowns in normal and whisper registers. Journal of the Acoustical Society of America 96 (4), 2101–2107.

Tompkins, C.A., 1991. Automatic and effortful processing of emotional intonation after right or left hemisphere brain damage. Journal of Speech and Hearing Research 34 (4), 820–830.

Williams, C.L., Stevens, K.N., 1972. Emotions and speech: Some acoustical correlates. Journal of the Acoustical Society of America 52 (4), 1238–1250.

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

Winner, E., Leekman, S., 1991. Distinguishing irony from deception: Understanding the speaker_s second-order intention. British Journal of Developmental Psychology 9, 257–270.

Yoshida, H., Furuya, Y., Shimodaira, K., Kanazawa, T., Kataoka, R., Takahashi, K., 2000. Spectral characteristics of hypernasality in maxillectomy patients. Journal of Oral Rehabilitation 27, 723–730.

Yumoto, E., Gould, W.J., Baer, T., 1982. Harmonics-to-noise ratio as an index of the degree of hoarseness. Journal of the Acoustical Society of America 71 (6), 1544–1550.